



ISLS Annual Meeting 2024 June 10-14, 2024  
Learning as a Cornerstone of Healing, Resilience, and  
Community

17<sup>th</sup> International Conference on Computer-  
Supported Collaborative Learning (CSCL)

CSCL Proceedings

Edited by: Jody Clarke-Midura, Ingo Kollar, Xiaoqing Gu & Cynthia  
D'Angelo



**International Society of  
the Learning Sciences**



University at Buffalo  
**UB** Graduate School of Education



ISLS Annual Meeting 2024  
*Learning as a Cornerstone of Healing, Resilience, and  
Community*  
Buffalo, USA, June 10-14  
Workshops: June 8-9  
University at Buffalo

**17<sup>th</sup> International Conference on  
Computer-Supported Collaborative  
Learning (CSCL) 2024**

**- CSCL Proceedings -**

4<sup>th</sup> Annual Meeting  
of the  
International Society of the Learning Sciences (ISLS)

Editors:  
Jody Clarke-Midura, Ingo Kollar, Xiaoqing Gu & Cynthia D'Angelo



ISLS Annual Meeting 2024

*Learning as a cornerstone of healing, resilience, and community*

Buffalo, USA, June 10-14

Workshops: June 8-9

University at Buffalo

***17<sup>th</sup> International Conference on Computer-Supported Collaborative Learning  
(CSCL) 2024***

© 2024 International Society of the Learning Sciences, Inc. [ISLS]. Rights reserved.



[www.isls.org](http://www.isls.org)

ISBN: 979-8-9906980-1-7 (ICLS Proceedings, PDF Version)

ISSN: 2543-0157

Cite as: Clarke-Midura, J., Kollar, I., Gu, X., & D'Angelo, C. (Eds.). (2024). *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning - CSCL 2024*. Buffalo, USA: International Society of the Learning Sciences.

All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the copyright notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior written permission of the International Society of the Learning Sciences. The International Society of the Learning Sciences is not responsible for the use which might be made of the information contained in this book.

Conference logo design by Amber M. Winters, MBA, Assistant Dean for Marketing and Director of Communications and Events, Graduate School of Education, University at Buffalo.

# Conference Organizers and Committees

## Conference Co-Chairs

Christopher Hoadley, *University at Buffalo, USA*  
X. Christine Wang, *University at Buffalo, USA*

## Local Conference Committee

Christopher Hoadley, *University at Buffalo, USA*  
X. Christine Wang, *University at Buffalo, USA*  
Jennifer A. Meka, *University at Buffalo, USA*  
Adrienne Decker, *University at Buffalo, USA*

## Hybrid Track Co-Chairs

D. Teo Keifert, *University of North Texas (USA)*  
Cynthia D'Angelo, *University of Illinois at Urbana-Champaign (USA)*  
Enrique (Henry) Suárez, *University of Massachusetts Amherst (USA)*

## Community Engagement Day Committee Co-Chairs

Wil Green, *University at Buffalo Graduate School of Education (USA)*  
Noemi Waight, *University at Buffalo Graduate School of Education (USA)*

## Social Media Communications Co-Chairs

Justice Walker, *University of Texas at El Paso, USA*  
Jennifer Olsen, *University of San Diego, USA*  
Naomi Thompson, *University at Buffalo, USA*  
Stacy Scheuneman, *University at Buffalo, USA*

## Accommodations Committee Co-Chairs

David Jackson, *University at Buffalo (USA)*  
Rachel Bonnette, *University at Buffalo (USA)*  
Joe Curnow, *University of Manitoba (Canada)*

## Volunteer Committee Co-Chairs

Adrienne Decker, *University at Buffalo (USA)*  
Sameer Honwad, *University at Buffalo (USA)*  
Grace (Yaxin) Xing, *University at Buffalo (USA)*  
Finn Goehrig, *University at Buffalo (USA)*

## Essential Support Team

Nicole Hoover, *ISLS Conference Coordinator*  
Caroline Hurley, *Event Coordinator*  
Meagan O'Hara, *ISLS 2024 Webmaster*  
Alicja Żenczykowska, *Conference System Architect*  
Paulo Blikstein, *Conference System Chair*  
Thomas Hillman, *EasyChair Guru*  
Varun Bhatt, *Accommodations & Technology Support*

## Copy Editors

Shiri Kashi, *University of Haifa, Israel*  
Liat Rahmian, *University of Haifa, Israel*

## Senior Reviewers

Aditi Kothiyal  
Amy Stornaiuolo  
Annelies Raes  
Armin Weinberger  
Bodong Chen  
Bram De Wever  
Brian Belland  
Catherine Dornfeld Tissenbaum  
Chandan Dasgupta  
Chen-Chung Liu  
Chris Teplovs  
Christopher Proctor  
Claudio Alvarez  
Crina Damsa  
Daniel Bodemer  
Daniel Spikol  
Davinia Hernández-Leo  
Dongsik Kim  
Engin Bumbacher  
Frank Fischer  
Gerry Stahl  
H. Ulrich Hoppe  
Hanni Muukkonen  
Hillary Swanson  
Ingo Kollar  
Iris Tabak  
Jody Clarke-Midura  
Justice Walker  
Kristine Lund  
Marc Jansen  
Michelle Perry  
Murat Cakir  
Mutlu Cukurova  
Omid Noroozi  
Oskar Lindwall  
Pierre Tchounikine  
R Ben Shapiro  
Ritsuko Oshima  
Roberto Martinez-Maldonado  
Rolf Steier  
Sébastien George  
Seng Chee Tan  
Stefan Trausan-Matu  
Sten Ludvigsen  
Susan Bridges  
Toshio Mochizuki  
Wenli Chen  
Xiaoqing Gu  
Yannis Dimitriadis  
Yoav Bergner  
Yu Xia

## Reviewers

Alan Muller	Ishari Amarasinghe	Qizhi Xu
Alejandra Martínez-Monés	James Planey	Ravi Sinha
Allison Hall	Jason Ji	Rawad Chaker
Aloysius Ong	Jeffrey Bush	Rogers Hall
Alwyn Vwen Yen Lee	Jessica Roberts	Rogers Kaliisa
Alyssa Wise	Jessica Vandenberg	Rujun Gao
Amanda Barany	Jina Kang	Saad Shehab
Andrew Estrada Phuong	João Adriano Freitas	Safinah Ali
Andy Stoiber	Joe Curnow	Sai Gattupalli
Anna Yinqi Zhang	Joshua Quick	Sarah Jaewon Lee
Anne Fensie	Julia Eberle	Sarita Singh
Ashwin Tudur Sadashiva	Jumana Almahmoud	Sean Kao
Ayano Ohsaki	Kamakshi Rajagopal	Sebastian Simon
Bahar Shahrokhan	Katherine Moore	Sebastian Strauß
Ben R Shapiro	Kevin Lenton	Shawna Thomas
Bertrand Schneider	Kim MacKinnon	Shotaro Naganuma
Calixto Gutierrez Braojos	Krishna Kathala	Sotheara Veng
Carolyn P. Rosé	Laura Brandl	Tenzin Doleck
Ceren Ocak	Lauren Bagdy	Tong Li
Chai Shaoming	Lauren Penney	Valentina Nachtigall
Changzhao Wang	Li-Jen Wang	Vanessa Dennen
Chao Yang	Liru Hu	Vasiliki Laina
Chengliang Wang	Lisa Hardy	Vivek Sabanwar
Chew Lee Teo	Lisa Paneth	Wenhao Chen
Chrissa Manoli	Lishan Zheng	Xiaoqing Gu
Christian Hartmann	Loris Jeitziner	Xinran Zhu
Christoph Richter	Lu Lawrence	Xintian Tu
Conrad Borchers	Luis Morales-Navarro	Xueqi Feng
Corey Schimpf	Lydia Y. Cao	Xuesong Cang
Dacun Hong	Mahir Akgun	Yawen Yu
Diler Oner	Masataka Kaneko	Yiannis Georgiou
Edith Bouton	Matin Yarmand	Yilang Zhao
Elizabeth Charles	Megan Humburg	Ying Hu
Elizabeth Koh	Mei-Hwa F. Chen	Yipu Zheng
Emma Anderson	Mengting Yu	Young Hoan Cho
Feng Lin	Michael Rook	Yuanru Tan
Francisco Enrique Vicente	Michael Schneider	Yurdagül Bogar
Castro	Michael Tscholl	Yuxin Chen
Freydis Vogel	Mohammadreza Farrokhnia	Zhichun Liu
Gamze Türkmen	Nathaniel Cradit	Zuokun Zuokun
Gaoxia Zhu	Nga Hoang	
Gautam Biswas	Nicolae Nistor	
Guangji Yuan	Nicole Hutchins	
Guo Su	Nikki Lobczowski	
Ha Nguyen	Nisha Devasia	
Haesol Bae	Norman Bier	
Haiyang Xin	Olga Chernikova	
Hakeoung Hannah Lee	Osnat Atias	
Heyjin Park	Pauliina Rikala	
Hironori Egi	Paulina Haduong	
Huiying Cai	Pei-Yi Lin	
I-Chien Chen	Pengjin Wang	
Indrani Dey	Preeti Raman	
Injila Rasul	Qiao Jin	

## Preface

This volume contains the proceedings of the long papers, short papers, posters, and symposia that were presented at the 17th International Conference on Computer-Supported Collaborative Learning, which was part of the Fourth Annual Meeting of the International Society for the Learning Sciences (ISLS2024), held in Buffalo, New York. Our community convened at the University at Buffalo to share new research and advance our field in light of this year's conference theme, "Learning as a cornerstone of healing, resilience, and community." Despite ongoing global crises, longstanding members and newcomers gathered in beautiful Buffalo to share new research, engage in lively discussions, expand their networks, and strengthen their practices for fostering human flourishing.

Although this is the 17th annual gathering of our community, it marks only the fourth occasion that the International Conference on Computer-Supported Collaborative Learning (CSCL) and the International Conference of the Learning Sciences (ICLS) have been held together during the Annual Meeting of the International Society of the Learning Sciences (ISLS). ISLS2024 also marks the inaugural introduction of Hybrid Symposia sessions.

The CSCL Proceedings feature long papers, short papers, posters and symposia, all subject to rigorous double-blind peer review. We received 128 submissions from 28 countries across Europe, Asia-Pacific and America. The submissions covered a broad range of CSCL research, design and technological innovation. In total, 33% (21 out of 63) long paper submissions, 36% (14 out of 39) short paper submissions, and 65% (15 out of 23) posters were accepted in the category in which they were submitted. Three symposia (2 CSCL and 1 hybrid) were accepted. In addition, a number of submissions were accepted in another category (short papers or posters). As a result, the CSCL Proceedings features 21 long papers, 36 short papers, 54 posters, two symposia, and one hybrid symposium.

The CSCL program included research on collaborative learning in groups, organizations, communities, and networks and the design of technologies for collaboration in the contexts of education, business, and society. To advance CSCL theories, technologies, and designs, the CSCL program brought together scholarship from education, computer science (including artificial intelligence (AI), learning analytics and data mining), psychology, linguistics, anthropology, sociology, and business. The program featured contributions that focused on various topics, among them:

**Human-AI Collaboration and Educational Technology.** The advancement of AI technologies has paved the way for innovative approaches to collaboration and learning. Papers under this theme explore how AI tools, such as GPT-4 and generative AI, facilitate human-AI collaboration in various educational contexts. For example, one study investigates the transparency and trustworthiness of AI-integrated qualitative coding, while another examines the effectiveness of AI tutors and virtual peer learners in supporting collaborative learning tasks. These papers highlight the potential of AI to enhance learning experiences and research in CSCL.

**Collaborative Learning Strategies and Scaffolding.** Papers in this theme evaluate different approaches to fostering collaboration, including online training programs, re-annotation behaviors in immersive environments, and the use of fixed and adaptive scaffolds in game-based learning. These studies provide insight on the cognitive and emotional regulation patterns in collaborative settings, the impact of VR video-based learning, and the influence of goals on collaboration scripts. In doing so, they contribute to knowledge on the design of collaborative learning environments.

**Data Analytics, Assessment, and Participation Equity in Learning.** Data analytics and assessment play a critical role in understanding collaborative learning processes. Papers in these sessions studied public engagement in science communication on Twitter, student perspectives on learning data ethics through game design, and behavior patterns in collaborative problem-solving tasks. Studies contributed to knowledge on participation equity, proposing metrics to evaluate fairness in collaborative learning, and investigating teacher stress levels in orchestrating CSCL activities. These papers highlight the importance of data-driven approaches to enhance participation equity and assess collaborative learning outcomes.

The program chairs extend their gratitude to the numerous individuals who dedicated countless hours ensuring the program was of high quality. We thank the Hybrid track co-chairs, D. Teo Keifert, Cynthia D'Angelo, and Enrique (Henry) Suárez, for their support reviewing the CSCL hybrid symposia. We thank the 144 reviewers and



50 meta-reviewers who conducted over 378 reviews and meta-reviews for CSCL. The thorough and constructive reviews were essential to the program's success. We thank the leadership of ISLS and the ISLS committees for their guidance and support. We thank Alicja Żenczykowska and Paulo Blikstein, who rock. And last, but certainly not least, we thank the amazing ISLS2024 local organizing team.

The 2024 conference embodied the value of community, collaboration, of sharing knowledge, expertise and experiences in working within the CSCL field, across disciplinary areas and across continents.

Jody Clarke-Midura, Ingo Kollar, Xiaoqing Gu,  
Cynthia D'Angelo (Hybrid Symposium Track)  
Program Co-chairs, CSCL 2024

## Acknowledgments

ISLS Annual Meeting 2024 conference organizers would like to give special thanks to the following committees and people for their support:

- International Society of the Learning Sciences (ISLS)
- ISLS Board of Directors and Executive Committee
- University at Buffalo, USA
- Wallace Foundation
- National Science Foundation (NSF)
- EasyChair Support Team
- Whova Support Team
- Conference Organizing Committee including
  - LS and CSCL Program Committees
  - Hybrid Track Co-chairs
  - Community Engagement Day Committee Co-Chairs
  - Social Media Communications Co-Chairs
  - Accommodations Committee Co-Chairs
  - Volunteer Committee Co-Chairs
  - Doctoral Consortium Co-Chairs
  - Early Career Workshop Co-Chairs
  - Mid-Career Workshop Co-Chairs
  - Interactive Tools and Demos Co-Chairs
  - Pre-Conference Workshops and Tutorials Co-Chairs
  - ISLS 2024 Essential Support Team
- Reviewers and Sr. Reviewers
- Best Paper Award Committees
- Conference Proceeding Copy Editors

# Table of Contents

## CSSL Proceedings

### Long Papers

Making Human-AI Contributions Transparent in Qualitative Coding .....	3
<i>Sariah Lopez-Fierro, Ha Nguyen</i>	
Learning with a Scripted AI Tutor or a Virtual AI Learner: Experimental Investigation of Collaborative Learning Using CoCot.....	11
<i>Yugo Hayashi, Shigen Shimojo, Tatsuyuki Kawamura</i>	
Twitter Talk: Informal Learning in Public Responses to Science Communication.....	19
<i>Ashley Boone, Jessica Roberts</i>	
Exploring the Efficacy of an Online Collaborative Problem Solving Training Program .....	27
<i>Jessica Andrews-Todd, Emily Kerzabi, Carol M. Forsyth</i>	
I Know Where You are Looking! Broadening the Notion of Annotation in a Collaborative Immersive Simulation Environment .....	35
<i>Yiqiu Zhou, Jina Kang</i>	
Examining Coordinated Computer-Based Fixed and Adaptive Scaffolds in Collaborative Problem-Solving Game Environments .....	43
<i>Daeun Hong, Chen Feng, Xiaotian Zou, Cindy E. Hmelo-Silver, Krista Glazewski, Tianshu Wang, Bradford Mott, James Lester, Krista Glazewski, Tianshu Wang, Bradford Mott, James Lester</i>	
What Are They Regulating? Research on Cognitive, Task and Emotional Regulation Patterns in CSCL.....	51
<i>Shurui Gao, Weijia Chen, Zihan Guo, Fati Wu</i>	
Student Perspectives on Learning and Teaching Data Ethics Through Speculative Game Design .....	59
<i>Nisha Devasia, Andrea Figueroa, Sarah Evans, Bernease Herman, Cecilia Aragon</i>	
Use of Generative AI for Boundary Crossing in Interdisciplinary Collaborative Research .....	67
<i>Shotaro Naganuma, Tsubasa Minematsu, Kana Matsueda, Jun Oshima</i>	
The Role of Individual Preparation on Coordination in Computer-Supported Collaborative Learning: A Neuroscience Perspective on Learners' Inter-Brain Synchronization .....	75
<i>Wenli Chen, Qianru Lyu, Mei-Yee Mavis Ho, Jessica Tan, Wei-Peng Teo, Siew Cheng Aileen Chai, Guo Su</i>	
Identification of Different Patterns in Solving Collaborative Jigsaw Puzzle Tasks Using Hidden Markov Models .....	83
<i>Qizhi Xu, Zongting Ge, Mengxiao Zhu, Ziwei Dou, Jing Wang</i>	
Interaction Forms of Collaborative VR Video Learning: An Exploratory Study .....	91
<i>Qiao Jin, Ishan Sodhi, Angie Chen, Svetlana Yarosh</i>	
Do Different Goals Affect the Configuration of University Students' Internal Collaboration Scripts? Results of an Epistemic Network Analysis Study .....	99
<i>Tugce Özbek, Martin Greisel, Christina Wekerle, Ingo Kollar</i>	
Investigating Collaborative Problem Solving Behaviors during STEM+C Learning in Groups with Different Prior Knowledge Distributions .....	107
<i>Caitlin Snyder, Cai-Ting Wen, Nicole M. Hutchins, Caleb Vatrul, Chen-Chung Liu, Gautam Biswas</i>	
Improving Hybrid Brainstorming Outcomes with Scripting and Group Awareness Support.....	115
<i>Mohammadreza Farrokhnia, Omid Noroozi, Harm Biemans, Yvette Baggen</i>	

Towards a Framework to Link Them All. A Provocation in the Ongoing Debate on Collaborative Learning.....	123
<i>Sebastian Simon, Iza Marfisi-Schottman, Sébastien George</i>	
A Meta-Space for “Super Notes”: How Two Teachers Supported Cross-Classroom Collaboration for Knowledge Building.....	131
<i>Yanqing Sun, Jianwei Zhang</i>	
Shall We Play a Game? Distributed Games with a Generative AI Player.....	139
<i>Jenna Matthews, Ha Nguyen, Hillary Swanson</i>	
Developing Participation Equity Criteria: A Fresh Perspective via Turn-Taking Simulation.....	147
<i>Liru Hu, Jiajun Wu</i>	
A Knowledge Building-Modeling Approach to Scientific Inquiry.....	155
<i>Dina Soliman, Xueqi Feng</i>	
Exploring Teacher Orchestration Load in Scripted CSCL: The Role of Heart Rate Variability.....	163
<i>Lubna Hakami, Batuhan Sayis, Ishari Amarasinghe, Davinia Hernández-Leo</i>	
<b>Short Papers</b>	
Enhancing Argumentative Essay Structure and Quality through a Collaborative Argumentation Environment in University EFL Students.....	173
<i>Li-Jen Wang, Ying-Tien Wu, Teng-Yao Cheng</i>	
Using Pose Data to Explore Changes in Students’ Self-Efficacy and Community Belonging in Makerspaces.....	177
<i>Bertrand Schneider, Edwin Chng</i>	
Toward an Assessment of Students’ (Social) Computational Literacy.....	181
<i>Michael Foster, Matthew Dunham, Christian Cammarota, Mike Verostek, Ben Zwickl, Tony E. Wong</i>	
Examine Student Resource Uses in a Game-based CSCL.....	185
<i>Xiaotian Zou, Daeun Hong, Feng Chen, Cindy E. Hmelo-Silver, Krista Glazewski, Tianshu Wang, Bradford Mott, James Lester</i>	
Measuring Help-seeking in Online Course Discussion Forums with Privacy-preserving Large Language Models.....	189
<i>Nigel Bosch, Destiny Williams-Dobosz, Michelle Perry</i>	
Examining Collaborative Clinical Reasoning within Synchronous Computer-Supported Collaborative Learning: A Cross-Cultural Comparison of Dutch and Finnish Medicine Students.....	193
<i>Chrissa Manoli, Hannie Gijlers, Chandan Dasgupta, Bas de Leng</i>	
Convergent Group Understanding: Indicator of Knowledge Integration?.....	197
<i>Xuesong Cang, Litong Zeng, Mike Tissenbaum, Dana Gnesdilow, Sadhana Puntambekar</i>	
Inferential Detectors to Automate Feedback on Social Skills in the Wild: The Case of the Advice Hour Graphic Novel.....	201
<i>Seth Corrigan</i>	
Behavior Sequence Patterns in Collaborative Problem Solving: A Multiple-perspective Analysis.....	205
<i>Mengxiao Zhu, Xin Wang, Qizhi Xu, Jessica Andrews-Todd</i>	
The Relationships between Students’ Epistemological Beliefs and Social Interactions and Epistemic Actions in Collaborative Inquiry Practice.....	209
<i>Xiaojie Zhang, Shihui Feng, Carol K. K. Chan, Jianhua Zhao</i>	
Supporting Pre-Service Teachers’ Evidence-Informed Reasoning Through Peer-Feedback: Effects of Feedback Provision and Feedback Integration Scaffolds.....	213
<i>Julia Hornstein, Martin Greisel, Ingo Kollar</i>	
Unveiling Collaborative Learning Dynamics: Insights from Remote Jigsaw Instruction on Knowledge Creation Perspectives.....	217
<i>Ayano Ohsaki, Toshinobu Kasai</i>	

Who Wants to Hear Bad News? How the Epistemic Perspective Determines the Perception of Peer Feedback.....	221
<i>Martin Greisel, Julia Hornstein, Anna Weidenbacher, Johanna Ott, Ingo Kollar</i>	
Learning Under Stress: Enhancing Team-Based Simulation Training with Multimodal Data.....	225
<i>Daniel Sánchez</i>	
Evaluating Metaspace as an Infrastructure for Teacher Cross-Community Collaboration in Knowledge Building.....	229
<i>Guangji Yuan, Bodong Chen, Chew Lee Teo, Carol K. K. Chan</i>	
Co-Design a Logic Model for Inclusive AI-Powered Learning Application with Primary School Teachers.....	233
<i>Ariel Han, Seth Corrigan, Shenshen Han, Kylie Pepler</i>	
Source Memory and Collaborative Learning: The Role of Group Composition and Conflicting Information.....	237
<i>Oktay Ülker, Daniel Bodemer</i>	
Detecting Patterns of Idea Novelty and Complexity in Student Knowledge Building Discourses.....	241
<i>Guangji Yuan, Chew Lee Teo, Zheng Chu, Jianwei Zhang, Mei-Hwa F. Chen, Tianlong Zhong, Chun Yen Chang-Sundin, Alwyn Vwen Yen Lee, Ong Kian Keong Aloysius</i>	
Adaptive Instructional Support in Computer-Supported Collaborative Learning: A Meta-analysis.....	245
<i>Lixiang Gao, Frank Fischer, Matthias Stadler, Olga Chernikova</i>	
Navigating Academic Pathways with Collaborative Data: An Analysis of Enrollment Decision-Making.....	249
<i>Nathaniel W. Cradit, Meghan Krawczyk</i>	
An Investigation of Dashboard in Collaborative Inquiry: The Dynamic Interplay between Technology and Pedagogy in Classroom Orchestration.....	253
<i>Chen Feng, Haesol Bae, Daeun Hong, Krista Glazewski, Cindy E. Hmelo-Silver, Xiaotian Zou, Tianshu Wang, Thomas Brush, Seung Lee, Bradford Mott, James Lester</i>	
Proposing an Evaluation Framework on Educational Technology that Supports Synchronous Online Collaborative Learning.....	257
<i>Yu Xia, Shulong Yan</i>	
Exploring the Dynamics and Trends of Knowledge Exchange: A Structured Topic Modeling Approach of the CSCL Conference Proceedings.....	261
<i>Rogers Kaliisa, Sonsoles López-Pernas, Kamila Misiejuk, Mohammed Saqr, Crina Damsa</i>	
Scripting Undergraduates' Interdisciplinary Collaborative Learning to Enhance Their Interdisciplinary Competence.....	265
<i>Guo Su, Wenli Chen, Gaoxia Zhu, Chencheng Le, Lishan Zheng</i>	
The Influence of Interaction Patterns on Performance in Online Collaborative Problem Solving.....	269
<i>Somi Park, Bookyung Shin, Seonyoung Keum, Youngsun Choo, Mi Song Kim, Young Hoan Cho</i>	
"We did it we did it!": Exploring Middle Schoolers' Collaborative Learning in an Online Machine Learning Curriculum.....	273
<i>Changzhao Wang, Jennifer Kahn</i>	
Co-Designing an AI Tool to Support Discourse Based Math Instruction in a High-Dosage Tutoring Context.....	277
<i>Chelsea Brown, Jeffrey B. Bush</i>	
Exploring Families' Science Talk Using a Mobile Augmented Reality App: An Epistemic Network Analysis Method.....	281
<i>Leqi Li, Susan M. Land, Heather Toomey Zimmerman, Yanping Xu</i>	
Audio Analysis of Group Discussion Patterns in Noisy Classrooms Before, During, and After Teacher-Group Interactions.....	285
<i>Chris Palaguachi, Cynthia D'Angelo, Elizabeth Dyer, Nessrine Machaka</i>	

A Framework for Collaborative Learning in Immersive Virtual Reality.....	289
<i>Paul Hatch, Joseph E. Michaelis</i>	
Parsing the Use of Computational Concepts with Scratch Projects.....	293
<i>Joey Huang</i>	
MOSAIC Protocol: Analyzing Small Group Work to Gain Insights into Collaboration Support for Middle School STEM Classrooms.....	297
<i>Nga Hoang, Jeffrey B. Bush, Indrani Dey, Emily Watts, Charis Clevenger, William R. Penuel</i>	
Speech-based Automatic Classroom Activity Detection.....	301
<i>Bahar Shahrokhian, David Zhou, Kurt Vanlehn</i>	
Orchestrating Change: A Narrative Analysis of Teacher Practice in a Post-Pandemic, Technology-Enhanced Learning Environment.....	305
<i>Yuxin Chen</i>	
Toward Productive Multivocality in AI Development: Excavating Ethics Concerns among an Interdisciplinary Team.....	309
<i>Grace Xing, Sanaz Ahmadzadeh Siyahrood, X. Christine Wang, Jinjun Xiong</i>	
<b>Symposia</b>	
Perspectives from the Field: Scaffolding Peer Critique and Feedback.....	315
<i>Christian Schunn, Omid Noroozi, Wenli Chen, Qianru Lyu, Siew Cheng Aileen Chai, Xinyi Li, Justice Walker, Sayed Mohsin Reza, Alex Acquah, Andi Scarola, Amanda Barany, J. Bryan Henderson, Annabel Stoler, Eve Manz, Jinzhi Zhou, Cindy E. Hmelo-Silver, Danielle Murphy, Zachary Ryan, Christina Stiso, Qiuyu Lin, Joshua Danish, Ravit Golan Duncan, Clark A. Chinn, Leslie Rupert Herrenkohl</i>	
Stepping Outside the Ivory Tower: How Can We Implement Multimodal Learning Analytics in Ecological Settings, and Turn Complex Temporal Data Sources into Actionable Insights?.....	323
<i>Bertrand Schneider, Richard Davis, Roberto Martinez-Maldonado, Gautam Biswas, Marcelo Worsley, Nikol Rummel</i>	
Bridging Social Annotation Practice with Perspectives from the Learning Sciences and CSCL.....	331
<i>Xinran Zhu, Yeonji Jung, Bodong Chen, Daniel Hickey, Grant Chartrand, Remi Kalir, Justin Hodgson, Chris Andrews, Alyssa Wise, Hong Shui, Pingting Chen, Rukmini Manasa Avadhanam</i>	
<b>Posters</b>	
Examining Scratch Programming: Exploring its Capacity to Cultivate Computational Thinking-aligned Learning Opportunities.....	341
<i>Faiaz Rahman, Kathy L. Schuh</i>	
Not the ‘Usual Suspects’: Caregiver Networking and Imagination in a Family Engagement Program.....	343
<i>Kira Baker-Doyle, Sarai Coba-Rodriguez, Andrea Vaughan, Evelyn Pollins</i>	
CocoNote Supported Project-Based Learning Environment: Perspectives of Construction and Collaboration.....	345
<i>Haiyang Xin, Qiannan Niu, Linling Lan, Zhuoxin Xiao, Fan Wu</i>	
Real-time Continuous Perceived Difficulty Annotation during Video Learning Using a More Natural Interface.....	347
<i>Xiaobo Liu, Yu Zhang</i>	
Scaffolding Students’ Adoption of Norms for Peer Critique.....	349
<i>Jinzhi Zhou, Lauren Albert, Cindy E. Hmelo-Silver, Joshua Danish, Zachary Ryan, Christina Stiso, Danielle Murphy, Ravit Golan Duncan, Clark A. Chinn, Qiuyu Lin</i>	
Inadequately Gifted: A Thematic Analysis of Memes as Embodied by gifted Experiences on Reddit.....	351
<i>Andria Gentry</i>	
Teachers' Discourse on WhatsApp In a Virtual Professional Learning Community.....	353
<i>Iris Saban, Linor L. Hadar</i>	

A Structural Equation Modeling Investigation of the Theory of Immersive Collaborative Learning (TICOL).....	355
<i>Valdemar Stenberdt, Noëlle Bender, Adéla Plechatá, Guido Makransky</i>	
Well-Being in CSCL Environments: Investigating Gender Variations .....	357
<i>Khadija El Aadmi-Laamech, Patricia Santos, Davinia Hernández-Leo</i>	
Multimodal Observations of Pair Programming Processes in a Shared Visual Workspace .....	359
<i>Siem Buseyne, Fien Depaepe, Annelies Raes</i>	
Task Complexity and Split Attention in Individual and Collaborative Learning .....	361
<i>John Guzmán, Jimmy Zambrano R.</i>	
Investigating the Knowledge Building Process of Collaborative Learning between Student and Virtual Tutor Supported by ChatGPT: A Discourse Analysis .....	363
<i>Xiaoyan Chu, Jiaqi Xu, Xuesong Zhai</i>	
Exploring the Link Between Personality Traits and Verbal Interactions in Computer-Supported Collaborative Problem Solving.....	365
<i>Siem Buseyne, Amelie Vrijdags, Sameh Said-Metwaly, Thierry Danquigny, Jean Heutte, Fien Depaepe, Annelies Raes</i>	
Collaborative Mayhem: Supporting Elementary Students Social-emotional Learning in Engineering Design.....	367
<i>Corey Schimpf, Askari Mehdi, Linda Smith, Amanda Thompson</i>	
Effect of Providing Knowledge Construction as A Preparatory Activity on Learner’s Critical Thinking Achievement in Online Discussion.....	369
<i>Lintang Matahari Hasani, Kasiyah Junus, Lia Sadita, Tsukasa Hirashima, Yusuke Hayashi</i>	
Supporting Joint Attention for Hybrid Groupwork With Multi-Image and Cursor Sharing in University Classes .....	371
<i>Hiroki Oura, Ryo Yoshikawa, Fumihiko Mitsunaga, Takaki Kondo, Wakako Fushikida, Hiroshi Kato</i>	
Examining Intrinsic and Extrinsic Game Mechanics in Jigsaw CSCL Scripts.....	373
<i>René Lobo-Quintero, Davinia Hernández-Leo</i>	
Designing for Large-Scale Social Annotation Experience: Real-Time Interactions on the Margins of Educational Material.....	375
<i>Jumana Almahmoud, Michele Igo, Marc Facciotti, Kamali Sripathi, David Karger</i>	
Linking Disagreement Reasons and Integrative Conclusions: An Epistemic Script to Promote Integration of Multiple Conflicting Information Based on Disagreement Analysis .....	377
<i>Toshio Mochizuki, Eowyn P. O’Dwyer, Clark A. Chinn, Myat Min Swe, Htay Min Khaung, Seiji Sekine</i>	
Tracing the Synchronized Physiological Arousal during GA, Supported Collaborative Learning Process.....	379
<i>Xiaoran Li</i>	
Are Your Top Social Annotation Friends Also Your Offline Group Members?.....	381
<i>Gaoxia Zhu, Tianlong Zhong, Hanxiang Du, Juan Zheng, Xiaoshan Huang, Shan Li, Chenyu Hou</i>	
The Multimodal Analysis of Learners’ Mathematical Proof Activity in a CSCL Environment.....	383
<i>Masataka Kaneko, Hironori Egi, Takuya Kitamoto, Takeo Noda</i>	
Epistemic Agency for STEM Learning with Design-Based Scaffolds.....	385
<i>Jialiang LIU, Gaowei Chen, Carol K. K. Chan, Jianhua Zhao</i>	
Co-Designing Synchronous Asymmetric Multiplayer Learning Game Scenarios .....	387
<i>Gaëlle Guigon, Mathieu Vermeulen, Mathieu Muratet, Thibault Carron</i>	
Training an Avatar-Based Conversational AI Chatbot and Investigating the Implementation of Communicative Action Principles .....	389
<i>Erum Afzal, Julia Fecke, Lenuta Alboaie, Edith Braun</i>	



Character Based Models and Computational and Embodied Action Tweaking for Sensemaking .....	391
<i>Joshua Danish, Nitasha Mathayas, Mengxi Zhou, Selena Steinberg, Morgan Vickery</i>	
Explore Preservice Teacher’s Interaction Pattern Mediated by Teacher Dashboard in the Video-Based Professional Learning Communities.....	393
<i>Huiying Cai, Linmeng Lu, Xiaoqing Gu</i>	
Opening the Black Box of the Effects of Instructional Support on Collaborative Diagnostic Processes with Process Mining .....	395
<i>Katharina M. Bach, Andreas Wildner , Martin R. Fischer, Frank Fischer, Olga Chernikova, Constanze Richters</i>	
Can ChatGPT be Always Powerful: Evidence from a Postgraduate Writing Class .....	397
<i>Chunlin Lei, Dan Zhu, Junjie Zhang</i>	
Understanding Knowledge Building Processes using Epistemic Network Analysis and Sequential Analysis .....	399
<i>Xiaojie Zhang, Carol K. K. Chan, Shihui Feng, Jianhua Zhao</i>	
Toward a Learning Agreement for Generative AI Use: First-Year University Student Perspectives .....	401
<i>Marc Beardesley, Ishari Amarasinghe, Davinia Hernández-Leo</i>	
Collaborative Diagnostic Reasoning in a CSCL environment: A Multi Study Structural Equation Model.....	403
<i>Laura Brandl, Matthias Stadler, Constanze Richters, Anika Radkowsch, Ralf Schmidmaier, Martin R. Fischer, Frank Fischer</i>	
Design, Development, and Evaluation of an Immersive Virtual Reality Learning Environment (IVRLE) for Motor Learning in Fencing.....	405
<i>Jonathan J. Foo, Wenli Chen, Wei-Peng Teo</i>	
Evaluating Video Conferencing and Desktop Virtual Platforms for Supporting Remote Classroom Activities	407
<i>Matin Yarmand, Ru Wang, Haowei Li, Nadir Weibel</i>	
Youth Perspectives on the Roles and Risks of AI in Their Classrooms .....	409
<i>Megan Humburg, Dalila Draganić-Cindrić, Cindy E. Hmelo-Silver, Krista Glazewski, James Lester</i>	
Reflecting on Prior Collaborations to Prepare for Group Work .....	411
<i>Joshua Katz, Molly Goldstein, Emma Mercier</i>	
Evaluating a New Knowledge Building Analytics Tool (KBAT) .....	413
<i>Calixto Gutiérrez-Braojos, Carlos Rodríguez-Domínguez, Daniela Linda, Paula Rodríguez-Chirino</i>	
Cultivating Epistemic Doubt: a Key Competence for Productive Participation in the Era of Artificial Intelligence .....	415
<i>Sini Davies, Kati Sormunen, Kaiju Kangas</i>	
Reimagining “The Lantern” as a Web-Based Tool for Enhanced Supervised Collaborative Problem Solving.....	417
<i>Héctor Florido, Davinia Hernández-Leo</i>	
Is a Collaboratively Created Visualization Worth 100,000 Words? A Case Study on Collaborative Learning Processes of Expert Groups in a Jigsaw Classroom .....	419
<i>Julia Eberle, Martin Bordewieck</i>	
Teaching Multimedia Design through Online Collaboration during COVID-19: Laurillard’s Conversational Framework.....	421
<i>Blessed A. Mhungu, Ralitsa D. Debra, Dick Ngambi</i>	
Use of AI to Foster Greater Awareness and Reflexivity Around Harmful Content in Online Discussion Tools.....	423
<i>Kim MacKinnon, Jingyu Liu, Cresencia Fong, Jai Pandey</i>	
Iterative or Linear Design? Insights into Teacher Professional Development on Dialogic Teaching .....	425
<i>Ying Zhang, Yu Liu, Zixiang Yu, Yu Li, Gaowei Chen</i>	





Navigating Boundary Crossing in a Global Engineering Course: A Case Study .....	427
<i>Ndubuisi Anuli, Asare Philip, Ajaz Mesha, James D. Slotta</i>	
Supporting dialogic reflection through visualising classroom talk in teachers' and students' learning communities .....	429
<i>Pengjin Wang, Fan Chen, Deliang Wang, Gaowei Chen</i>	
Spatial Manipulative Note-Taking Tool for Small Group Face-to-Face Collaboration in Science Class .....	431
<i>Litong Zeng, Xuesong Cang, Afroza Sultana, Stacy Cernova, Megan Wang, Dana Gnesdilow, Sadhana Puntambekar, Ali Mazalek, Mike Tissenbaum</i>	
Tensions in Designing Technological Interactives in a Justice-Centered Middle School Biology Curricular Unit.....	433
<i>Sugat Dabholkar, Ravit Golan Duncan, Frieda Reichsman, Rishi Krishnamoorthy, Edna Tan, Burrell Smithen, Jaya Joshi</i>	
Interaction Patterns and Structures in the R/Personalfinance Weekend Help and Victory Thread .....	435
<i>Jose Sandoval-Llanos, Priya Sharma</i>	
Initial Examination of Identity Development When Starting in an Online, Global, Collaborative, Learning Environment.....	437
<i>Danielle P. Espino, Nastassia Jones, Kristina Lux, Lu Liang, Leah-Maria Gaskin, Eric Hamilton, Seung Lee</i>	
Making Space for Critical Action: Re-visioning Computational Thinking .....	439
<i>Preeti Raman, Rubaina Khan, Meera Sharma, Renato Carvalho, Chandan Dasgupta, Aakriti Mahajan, James D. Slotta</i>	
Fine-Tuning Large Language Models for Data Augmentation to Detect At-Risk Students in Online Learning Communities .....	441
<i>Hongming Li, Anthony F. Botelho</i>	

# Long Papers



# Making Human-AI Contributions Transparent in Qualitative Coding

Sariah Lopez-Fierro, Utah State University, sariah.lopezferro@usu.edu  
Ha Nguyen, Utah State University, ha.nguyen@usu.edu

**Abstract:** Generative artificial intelligence (AI) tools offer promise in facilitating code discovery, refinement, and large-scale analyses in qualitative coding. While prior work has focused on the feasibility of AI tools, limited research has examined the process of human-AI collaboration in qualitative coding tasks. In this paper, we present a self-study that details how we use GPT-4, a generative AI tool, to assist with the early stages of inductive qualitative coding. We conduct two analyses that highlight human-AI collaboration: (1) how AI integration augments the coding workflow, as researchers iterate between reading data, open coding, refining, and organizing codes, and (2) how AI's output contributes to the development and refinement of the codebook. These visualizations aim to improve the transparency and trustworthiness of AI-integrated qualitative coding. They illustrate how to apply insights from collaborative learning research to explore human-AI collaboration.

## Introduction

Artificial Intelligence (AI) tools offer promises in augmenting the qualitative research workflow. Specifically, these tools can facilitate code discovery and refinement, idea synthesis, and establishment of code agreement (Gao et al., 2023; Lennon et al., 2021; Xiao et al., 2023). Recent development in generative AI tools (such as ChatGPT) enables multi-turn, conversational exchanges between researchers and the tools. This interaction style provides opportunities for iterative idea elaboration and refinement for human coders and AI (Zambrano et al., 2023). While prior research has illuminated the usability and effectiveness of AI in qualitative coding (Hong et al., 2022; Gao et al., 2023; Rastogi et al., 2023), limited work has attended to the *process* of how human researchers integrate AI into the process. Illuminating the decisions within the analytic process allows the work to be transparent and transferable to other contexts (Aguinis et al., 2018; Lincoln & Guba, 1985). We study human-AI coding as *knowledge construction*, defined as the process in which learners (coders) develop understanding through interaction and co-construction of a knowledge artifact (the codebook) (Bereiter, 2005). We attend to two layers of analysis to examine human-AI inductive coding. First, we explore the human researchers' talk moves, such as reading data excerpts, developing open codes, comparing codes, and organizing codes, when working with a generative AI tool (GPT-4). Second, we leverage insights from collaborative learning and learning analytics research to visualize human's and AI's contributions to idea development (Chen & Zhang, 2016). We ask:

RQ1: How can we visualize knowledge construction in human-human and human-AI inductive coding?

RQ2: How can we make transparent human-AI contributions to the development of qualitative codes?

## Literature review

### Qualitative coding process

Qualitative researchers rely on both inductive and deductive processes to explore meaning in data. Inductive coding processes, with approaches like thematic analysis and grounded theory, encompass generating codes from the data and iteratively organizing the codes into themes or categories (Charmaz, 2014). Meanwhile, deductive coding relies on theories and pre-defined coding schemes in a top-down manner (Linneberg & Korsgaard, 2019). The focus of the current work is the inductive coding process, where researchers are exploring emergent concepts in unstructured data. This focus is driven by emergent work that calls for understanding of how AI can support code exploration in qualitative coding (Jiang et al., 2021).

Researchers take specific steps in inductive coding. They read the data, develop initial codes, refine, organize codes to develop a preliminary codebook, test the codebook, conduct final coding, and finalize themes (Richards & Hemphill, 2017). In collaborative data analysis, where multiple coders jointly analyze a dataset, researchers bring different perspectives that may enrich code development and reduce bias. However, this may introduce challenges such as disagreement. Scholars have thus called for establishing transparency by providing readers access to information about the steps and decisions that researchers make in the coding process (Aguinis et al., 2018). Similarly, Lincoln and Guba (1985) emphasize trustworthiness to establish validity in qualitative work. Researchers attend to *credibility*, to detail the extent to which the context and data are articulated. They also communicate the consistency and lack of bias in the research process to establish *confirmability*. Further,

researchers need to provide enough information about data analyses for *dependability*. Finally, researchers provide thick descriptions of the data and findings as evidence for *transferability* to other contexts. Like human-human qualitative coding process, establishing transparency and trustworthiness through outlining the analytic steps is critical in AI-integrated qualitative analysis (Nguyen et al., 2021).

## AI tools in inductive coding: A knowledge construction process

AI tools offer various affordances for inductive qualitative coding. They assist human researchers in generating, defining, and organizing codes, thus reducing the cognitive burdens on researchers (Gebreegziabher et al., 2023). These tools can support collaborating researchers to reach agreement on contentious codes (Gao et al., 2023; Zambrano et al., 2023). Additionally, AI's suggestions offer alternative data interpretations or assist researchers in understanding key data patterns. Such interpretations may promote uncertainty and idea discovery for follow-up discussion and code refinement (Jiang et al., 2021). For example, PaTAT, an AI-enabled tool for thematic analysis, focuses on presenting emerging data patterns from user-annotated codes (Gebreegziabher et al., 2023). These patterns allow users to identify interesting phenomena and revise and merge codes. Importantly, the AI-integrated coding process showcases principles of *knowledge construction*, where collaborating partners engage in mutual idea refinement and explanations (Bereiter, 2005). In inductive coding, researchers jointly refine and expand upon a shared knowledge artifact (e.g., the codebook). Similarly, an AI tool that contributes to the coding process can be seen as “a collaborator”, as one of the researchers on our team playfully pointed out.

Throughout human-AI collaboration, the rationale behind AI's decisions should remain transparent (Rietz & Maedche, 2021; Wu et al., 2022). This transparency allows researchers to observe AI's decision-making and gauge trustworthiness (von Eschenbach, 2021). Researchers can document iterations of the coding rules for AI (Gebreegziabher et al., 2023; Rietz & Maedche, 2021), or prompt AI to provide explanations for its generated codes and examples (Zambrano et al., 2023). These procedures are possible, thanks to the increasing natural language capacity of generative AI tools to provide explanations in conversational ways. We build on these efforts to make transparent the AI-integrated coding process. Our focus on transparency serves two purposes: support researchers to document the analyses as part of knowledge construction and make coding decisions explicit within the research team (Ganji et al., 2022), and provide research credibility and confirmability (Nguyen et al., 2021).

## Visualizing the collaborative workflow

Research on data visualization has illustrated the importance of visuals in collaborative discussion, to build common understanding, surface possible discrepancies, and support knowledge integration among group members (Mengis & Eppler, 2006). Visualizing the AI's contributions to the process can illustrate how the tool can be leveraged, and the specific decisions and interpretations made to build upon human's and AI's insights. These visualizations enable the identification of relationships and patterns that may not always be evident in collaborative analyses, such as when researchers (and AI tools) have different interpretations or inconsistencies when coding the same data (Drouhard et al., 2017; Ganji et al., 2022).

We employ two visualizations to examine human-AI collaborative coding processes: *process mining* and *idea threads*. To answer RQ1, we use process mining, a method to analyze and visualize processes based on event data logs (Janssenswillen et al., 2019), to consider how human and AI create and refine ideas. Process mining showcases how collaborators iterate between moves and differentiate between more and less productive collaboration (Song et al., 2022). For RQ2, we visualize the human-AI collaboration as idea threads and specify which ideas are contributed by AI. Idea threads provide temporal visualizations of group discourse and show links between ideas that are generated, improved, and revisited over time. They reveal the focus and development of knowledge construction (Zhang et al., 2015). The visuals' attention to process reveals how human and AI coders engage in knowledge construction to construct, revisit, and refine shared knowledge (Chen & Zhang, 2016).

## Methodology

### Study context and data sources

We presented a self-study narrative of how to incorporate generative AI in the early stages of the qualitative coding process. As a self-study, we incorporated our reflection as researchers on the process and highlight moments in which the AI tool challenged or augmented our workflow.

The data corpus ( $n = 100$  responses) for our analysis came from a larger research project that examined how to use LLMs to communicate about climate change sentiments (Nguyen et al., 2024). To create the corpus, we prompted OpenAI's GPT-4 (Generative Pretrained Transformer 4) to generate responses that covered different science communication strategies and identity (e.g., race, gender, age, ability status). We generated GPT-4's responses off the shelf with the template: Write a {communication medium} about a/an {age} {gender} {race}

with {ability} talking about climate justice {issues} in {location}. A response might read “I hope this newsletter finds you well. My name is Noah, and I am a student at our school. I wear two badges, a multilingual learner, and a student with disability [...] allow me to see the world with a unique lens to discuss climate justice”.

We started the analysis with an open coding approach. This process involved the authors (a PhD student and a faculty in Education), both of whom had prior knowledge of the research context and data. After two coding iterations, we generated an initial codebook with seven main codes, 14 subcodes, and 8 sub-subcodes. We leveraged GPT-4 to assist with the coding process in OpenAI’s playground (temperature=1, max length=4096 tokens, top p=1). Drawing from prior research on AI prompting (Xiao et al., 2023; Zambrano et al., 2023), we used exploratory approaches to ask GPT-4 to generate new codes, provide data excerpts to ground the AI’s output, and ask for AI’s critique on the codebook (Table 1). We iteratively compared the AI’s insights with our codebook to add, refine, and merge codes in two more iterations. In the final codebook (4<sup>th</sup> iteration), 32.14% of the codes remained the same from previous iterations, 28.57% directly took AI’s suggestions as new codes, and 39.28% incorporated AI ideas (e.g., merging codes, adding to code descriptions, changing levels of codes and subcodes).

**Table 1**  
*Example Prompts and AI Responses*

	Example 1	Example 2	Example 3
Prompts	Generate all codes, subcodes, sub-subcodes, and definitions from the following response: {data excerpt}	Can you create code and subcode with respect to culture?	We developed a code for identity, with subcodes for race, ability status, mental health, and culture. Find examples in the responses for these subcodes: {excerpt 1}, {excerpt 2}
AI responses	Main Code: Role of the school community in Climate Justice Subcode: Understanding the concept of Climate Justice Definition: It entails grasping how our actions have severe consequences such as warmer summers ...	Certainly. One main code could be centered on "Cultural Influence on Climate Activism." Subcode: Indigenous Values and Climate Action Definition: Understanding the importance of indigenous values, knowledge, practices.	Here are examples in the responses that fit into these subcodes under code "Identity": Race. From the first response: "Everyone who knows me well, knows that I carry two badges, one of being part of a beautiful multiracial family."

We analyzed several data sources: (1) the discussion transcript of the researchers while integrating AI to understand human-human and human-AI coding (75 min; audio-recorded, transcribed via Whisper API), (2) the prompts used for GPT-4, (3) the AI’s responses, and (4) versions of the codebook (as screenshots of sticky notes of codes and subcodes on Google Jamboard throughout the coding process). We did not provide GPT-4 with the researchers’ coding discussion transcript; just excerpts from the data corpus to generate AI’s codes and examples.

## Analytical approaches

**Table 2**  
*Coding Move Definitions*

Code	Definition	Example from human-human transcript
read	Read the data excerpts	He [the excerpt] is saying “efforts in community workshop”.
open_code	Come up with first impression codes	Culture, Ancestry knowledge, Mother Earth, Location – are words that come to mind.
define_code	Ask questions and add description to refine or elaborate on codes	Researcher 1: What we have for Location that's not overlapping? Researcher 2: I think the location is going to be linked to all of the other things? We need to have location specific actions."
compare	Find relationships between two or more codes and subcodes	It has location on the list as well. It says “accurate place-based details”.
organize_code	Group codes/subcodes	I am thinking individual and collective actions can be subcodes?

find_examples	Find examples from data excerpts to support created codes	Subcode: Accurate Details Example from first response: "Growing up in Santa Barbara, we have seen how our culture and practices shaped understanding ..."
process	Discuss approaches to the coding process	What's going to happen is we will prompt the AI, based on the initial coding we just developed.
prompt	Find approaches to prompt AI	It would be interesting to see if it can generate codes, subcodes, and sub-subcodes.
AI	Interact with AI output	I like this part, "not all communities are equal".

**Table 3**  
*Sample Data Structure for the Process Mining Analysis*

resource_id	speakers	text	event	event_id	timestamp
human_human	R1	We are going to talk through the open code process [...] What do you think?	process	1	00:00:00
human_human	R2	There is code community	open_code	2	00:00:31
human_human	R2	coming from a community of surfers	read	3	00:00:37

### RQ1: How can we visualize knowledge construction in human-human and human-AI coding?

We developed a codebook to capture the different moves throughout the coding processes (Table 2). The coded moves align with initial inductive coding processes, including reading the data, developing codes, refining codes, and organizing codes (Richards & Hemphill, 2017). We added codes to describe AI integration, namely prompt (finding approaches to prompt the AI) and interactions with the AI (for example, reading AI's responses). The codebook was developed through three iterations. The authors separately coded 30% of the code and reached substantial inter-rater agreement (average Cohen's  $\kappa = .93$ ). The first author coded the rest of the data.

We then used process mining to illustrate the frequencies and connections between the moves. Moves were organized as sequences of events and visualized with the R package "BupaR" as a directly-follows graph (DFG; Janssenswillen et al., 2019). A DFG includes nodes (representing activities) and edges (target activity immediately follows the source activity). Table 3 illustrates how the data was structured for the analyses. The "resource\_id" column denoted the two processes (human-human and human-AI). Consecutive utterances under the same moves were grouped within the same rows, and moves were noted under the "event" column. The process visualizations (Figures 1, 2) highlight how frequently a move occurred, and how frequently one move preceded or followed another (e.g., "defining code" followed "open coding" 50% of the time).

### RQ2: How can we make transparent human's and AI's contributions to code refinement?

To visualize the contributions of human and AI to code refinements, we followed examples from computer-aided qualitative analysis (O'Kane et al., 2021). We wrote memos of the code iterations and the influence of GPT-4 on the output (e.g., how a code or subcode changes in each iteration, and whether the change is attributed to AI, the discussion, or a combination). We manually created a visualization in Figma to show how the codes changed, how the data were linked to these changes, and how insights from the researchers and AI contributed to the process.

## Findings

### Visualizing AI's impact on the coding workflow

We present the coding moves between the human researchers (Figure 1), and when incorporating GPT-4 (Figure 2). The percentages in the figure denote the frequencies of each move, out of all coded utterances. For example, reading code occurred in 35.94% of the utterances. Once the researchers started reading the data, they transitioned to open coding 78.26% of the time, defining code 17.39% of the time, and referring to the process 4.35% of the time. During human-human coding process (Figure 1), reading data and open coding were the most frequent moves. Examining the timestamps revealed that researchers spent the most time in these moves (64.06% of time spent in the human-human session). In comparison, during human-AI coding that followed human-human process (Figure 2), iterations between examining the AI's responses and comparing codes were dominant. Researchers spent the most time in those moves (54.86% of time spent in the human-AI session). The AI responses provided perspectives to add to and refine the codes, leading to more cycles of defining, comparing, and organizing codes. These processes reflect how researchers reacted to the additional AI-generated insights.

Consider the following vignette to illustrate how AI-generated insights augmented the workflow. Early in the human-human coding process, researchers came up with the following open codes: "human-nature relationship", "personal changes", and "action". These codes were based on data segments that highlighted climate

action, such as “We can make climate-friendly choices in our daily lives, like recycling more”. The researchers organized the code “action” into subcodes for “individual” and “collective”, upon finding more data examples.

During human-AI coding, the researchers prompted GPT-4 to generate codes and subcodes, based on the same data excerpts they had examined. The AI generated multiple subcodes, including the following: Significance of cultural and heritage learning for climate justice; practical ways to promote climate justice; individual uniqueness and contribution to climate justice, and collaborative efforts for climate justice. These subcodes prompted the researchers to engage in code comparisons, for example:

Researcher 1: It has both the individual and the collective, which I thought was kind of cool.

Researcher 2: Like, ways of how individuals can promote climate justice, and there are ways as protest ... how collaborative efforts for climate justice.

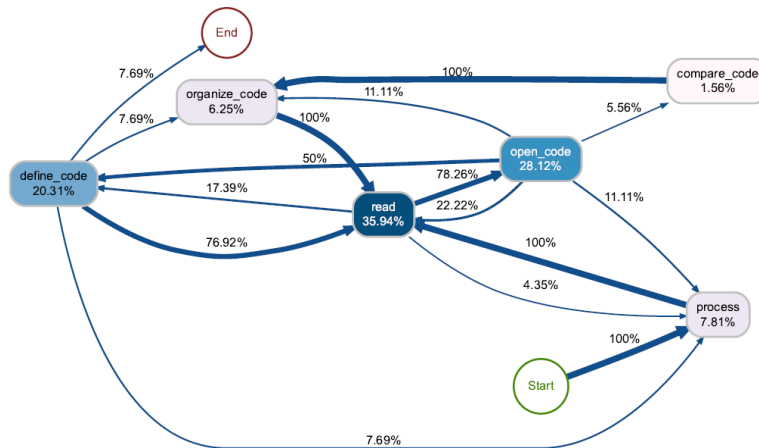
Researcher 1: This is different, though. This is more on the identity aspect ...

Researcher 2: You [your identity] as a contribution to climate justice

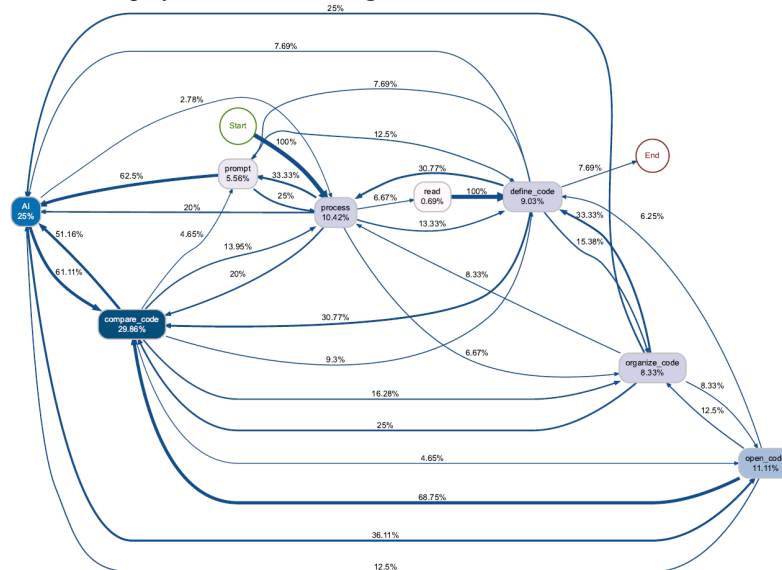
Researcher 1: Yeah, so like unique identity contribution

Here, the AI’s output primed the researchers to refine their code definitions, “action as linked to identity”. Integrating AI insights supported nuanced interpretations of the data and subsequent code comparisons, and refinement. Rather than directly adopting the AI’s codes, the new insights prompted further discussion.

**Figure 1**  
*Process Map of Human-Human Coding*



**Figure 2**  
*Process Map of Human-AI Coding*

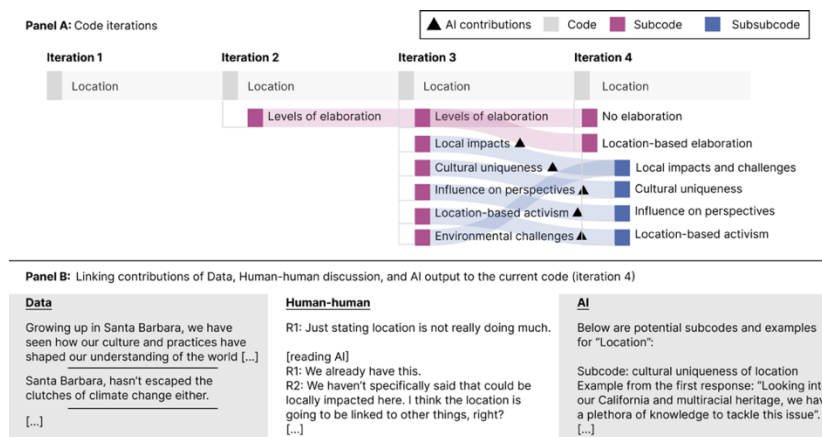




## Visualizing AI's impact on the code iterations

The AI's output not only augmented the coding workflow; it also influenced the researchers' analysis. We observed that the output allowed researchers to iteratively refine the preliminary codes and examine data excerpts through new perspectives. Like prior work (Nguyen et al., 2021; O'Kane et al., 2021), we note that the analytic pipeline should clearly indicate the contribution of computer-based analyses to shaping data interpretations. We provide an illustrative figure that maps out the iterations of one code, *location* (panel A, Figure 3), as well as how the iterations are driven by the data excerpts, researchers' assumptions, and AI's output (panel B, Figure 3). The figure was manually created, based on the research team's analytic memos of AI's and human's contributions to the codebook in each coding iteration.

**Figure 3**  
*Iterations of the "Location" Code, Linking Revisions to Data, Researchers' Discussions, and AI's Output*



Iterations one and two outlined code elaboration between the researchers (adding a subcode), and in iteration three, we incorporated the AI's responses. We asked GPT-4 to generate subcodes under the main code "location", based on three provided data excerpts. The AI's responses generated several new subcodes (e.g., local impacts, cultural uniqueness, influence on perspectives; AI's contributions were denoted by the triangles in Figure 3). The researchers then engaged in a follow-up discussion to investigate more data, compare the codes, and organize the human's and AI's generated codes (iteration four). Here, the notation for AI contributions suggests human-AI sensemaking process, where AI's ideas were further refined by human coders. In fact, a substantial proportion of the final codebook (39.28%; see Methods) reflected codes combining AI's and human's insights.

Documenting where AI influenced the code development supports transparency. It articulates the assumptions and contexts of the codebook development and allows readers to examine whether the researchers' interpretations (based on iterative conversations, data excerpts, and AI's output) are possible. This analysis contributes to the credibility of qualitative research (Lincoln and Guba, 1985). Establishing an audit trail to outline the analytic steps also contributes to the research's confirmability, so that other researchers can replicate the research procedures (Lincoln and Guba, 1985).

## Discussion

Recent research that incorporates generative AI into the inductive research process has focused on investigating feasibility and user perceptions (Gao et al., 2023; Rastogi et al., 2023). These research strands highlight the role of human researchers in staying close to the data and exercising caution in interpreting AI's output (Morgan, 2023). Our work examines the process of human-AI collaboration and provides practical insights for qualitative research. We explore how AI incorporation influences the research team's workflow, as indicated by comparing the human-human and human-AI process maps (Figures 1, 2). This temporal analysis may help researchers to answer additional questions, such as how AI can lead to reflexive moments marked by sequences of asking follow-up questions, revisiting data excerpts, and comparing codes to clarify the codebook. Researchers may also investigate whether the prompts for AI and subsequent workflow are aligned with the analytic intent. For example, if researchers intend to use the tool for open coding, we might expect to see more moves for reading data and open coding. Meanwhile, if researchers intend to streamline a preliminary codebook (such as in our case), there might be more code comparisons and refinement.

We provide a way to visualize the audit trail—tracing code iterations to instances in the data, insights from researchers' sensemaking, and AI's output (Figure 3). Computer-aided qualitative data analysis software supports text retrieval, to relate qualitative codes to coded text segments for review (O'Kane et al., 2021). We suggest an additional visual that illuminates insights from AI's output. The analysis provides a template for researchers to include some of these insights in writing up the methods of qualitative data analyses. For instance, researchers could detail (1) which parts of the coding process AI tools contribute to, (2) the prompts used for these tools and illustrative AI output, and (3) how the outputs are incorporated into the codebook development. These insights enhance the transparency of the coding process for both the research team and external communication to other scholars. On one hand, the insights help researchers to examine their data and coding processes. On the other hand, communicating these analytic steps improves the credibility and confirmability of the research.

Importantly, our findings have implications for computer-supported collaborative learning (CSCL) research. First, findings shed lights on idea evolution in human-AI collaboration as part of knowledge construction efforts (Bereiter, 2005). Knowledge building through idea refinement is a core CSCL activity (Wise et al., 2021). As AI tools become increasingly integrated into learning environments, scholars can use the visualizations to identify and investigate moments in the collaboration that are particularly productive for idea development. Second, the visualizations are not only useful to researchers but can be represented back to users (i.e., human qualitative researchers). In turn, users can modify their activities and AI integration to align the collaborative workflows with their desired uses (e.g., open coding, code refinement, or at-scale analyses).

The limitations of the current work can guide future research. We only documented the early stages of the inductive coding process, where researchers developed and refined a preliminary codebook. Additionally, the presented work relied on manual analyses. Future research could build on our initial efforts to experiment with AI uses throughout the qualitative coding process. Our visualizations illustrate possible additions to qualitative coding interfaces. We encourage future work to automate these visualizations to document human-AI collaboration, and study how users engage with the visualizations to refine data analysis and interpretation.

## Conclusion

We position generative AI as another research team member to engage in inductive coding. Attending to the process of human-AI coding helps us gain a more nuanced understanding of the tool's impact on the coding workflow and idea refinement. We present two visualizations, process mining and idea threads, as templates for other researchers. The visualizations make transparent the research process and decision-making that are often opaque. They allow researchers and readers to evaluate the credibility and confirmability of data interpretations stemming from human discussion, data excerpts, and AI contributions.

## References

- Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals*, 12(1), 83-110.
- Bereiter, C. (2005). *Education and mind in the knowledge age*. Routledge.
- Charmaz, K. (2014). *Constructing grounded theory*. Sage.
- Chen, B., & Zhang, J. (2016). Analytics for knowledge creation: Towards epistemic agency and design-mode thinking. *Journal of Learning Analytics*, 3(2), 139-163.
- Drouhard, M., Chen, N. C., Suh, J., Kocielnik, R., Pena-Araya, V., Cen, K., ... & Aragon, C. R. (2017, April). Aeonium: Visual analytics to support collaborative qualitative coding. In *2017 IEEE Pacific Visualization Symposium (PacificVis)* (pp. 220-229). IEEE.
- Ganji, A., Orand, M., & McDonald, D. W. (2018). Ease on Down the Code: Complex Collaborative Qualitative Coding Simplified with 'Code Wizard'. *Proceedings of the ACM on human-computer interaction*, 2(CSCW), 1-24.
- Gao, J., Guo, Y., Lim, G., Zhan, T., Zhang, Z., Li, T. J. J., & Perrault, S. T. (2023). CollabCoder: A GPT-Powered Workflow for Collaborative Qualitative Analysis. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 354-357.
- Gebreegziabher, S. A., Zhang, Z., Tang, X., Meng, Y., Glassman, E. L., & Li, T. J. J. (2023, April). Patat: Human-ai collaborative qualitative coding with explainable interactive rule synthesis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).
- Hong, M. H., Marsh, L. A., Feuston, J. L., Ruppert, J., Brubaker, J. R., & Szafir, D. A. (2022, October). Scholastic: Graphical Human-AI Collaboration for Inductive and Interpretive Text Analysis. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-12).

- Janssenswillen, G., Depaire, B., Swennen, M., Jans, M., & Vanhoof, K. (2019). bupaR: Enabling reproducible business process analysis. *Knowledge-Based Systems*, 163, 927-930.
- Jiang, J. A., Wade, K., Fiesler, C., & Brubaker, J. R. (2021). Supporting serendipity: Opportunities and challenges for Human-AI Collaboration in qualitative analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCWI), 1-23.
- Lennon, R. P., Fraleigh, R., Van Scoy, L. J., Keshaviah, A., Hu, X. C., Snyder, B. L., ... & Griffin, C. (2021). Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. *Family Medicine and Community Health*, 9(Suppl 1).
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. sage.
- Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative research journal*, 19(3), 259-270.
- Mengis, J., & Eppler, M. J. (2006). Seeing versus arguing the moderating role of collaborative visualization in team knowledge integration. *Journal of Universal Knowledge Management*, 1(3), 151-162.
- Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. *International Journal of Qualitative Methods*, 22, 16094069231211248.
- Nguyen, H., Ahn, J., Belgrave, A., Lee, J., Cawelti, L., Kim, H. E., ... & Villavicencio, A. (2021). Establishing trustworthiness through algorithmic approaches to qualitative research. In *Advances in Quantitative Ethnography: Second International Conference, ICQE 2020, Malibu, CA, USA, February 1-3, 2021, Proceedings 2* (pp. 47-61). Springer International Publishing.
- Nguyen, V., Nguyen, H., Ludovise, S., & Santagata, R. (2024, June). AI for Climate Justice: Assessing Large Language Models from an Intersectional Lens In *4<sup>th</sup> Annual Meeting of the International Society of the Learning Sciences*. International Society of the Learning Sciences.
- O’Kane, P., Smith, A., & Lerman, M. P. (2021). Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software. *Organizational Research Methods*, 24(1), 104-139.
- Rastogi, C., Tulio Ribeiro, M., King, N., Nori, H., & Amershi, S. (2023, August). Supporting human-ai collaboration in auditing llms with llms. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 913-926).
- Richards, K. A. R., & Hemphill, M. A. (2018). A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical education*, 37(2), 225-231.
- Rietz, T., & Maedche, A. (2021, May). Cody: An AI-based system to semi-automate coding for qualitative research. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).
- Song, Y., Cao, J., Yang, Y., & Looi, C. K. (2022). Mapping primary students’ mobile collaborative inquiry-based learning behaviours in science collaborative problem solving via learning analytics. *International Journal of Educational Research*, 114, 101992.
- von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607-1622.
- Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. *International handbook of computer-supported collaborative learning*, 425-443.
- Wu, T., Terry, M., & Cai, C. J. (2022, April). Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-22).
- Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. Y. (2023, March). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 75-78).
- Zambrano, A. F., Liu, X., Barany, A., Baker, R. S., Kim, J., & Nasiar, N. (2023, October). From nCoder to ChatGPT: From Automated Coding to Refining Human Coding. In *International Conference on Quantitative Ethnography* (pp. 470-485). Cham: Springer Nature Switzerland.
- Zhang, J., Chen, M., Tao, D., Sun, Y., Lee, J., & Judson, D. (2015). Fostering Sustained Knowledge Building through Metadiscourse Aided by the Idea Thread Mapper In Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. Ludvigsen, S. (Eds.) (2015). *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 1*. Gothenburg, Sweden: The International Society of the Learning Sciences.

## Acknowledgements

This research is supported by a pilot grant from Utah State University’s College of Education & Human Services.

# Learning with a Scripted AI Tutor or a Virtual AI Learner: Experimental Investigation of Collaborative Learning Using CoCot

Yugo Hayashi, Shigen Shimojo, Tatsuyuki Kawamura  
y-hayashi@acm.org, cp0013kr@ed.ritsumei.ac.jp, tatsu@fc.ritsumei.ac.jp  
Ritsumeikan University, Osaka, Japan

**Abstract:** This study investigated effective methods for using conversational agents for learner-learner support in a collaborative explanation task using concept maps. Scripted facilitation methods that use conversational agents were compared with the proposed method that integrates the reciprocal teaching strategy. This method uses an ACT-R cognitive architecture-based virtual peer learner agent to identify, reason, and construct a concept map from dyads in real time. An experiment was conducted under two conditions: 1) a tutor as a mediating agent that prescribes metacognitive scripts, and 2) an adaptive virtual peer agent that prescribes concept maps. The results demonstrated that both methods improved the learning performance compared with the control condition. Notably, learners working with the peer learner agent made better use of procedural knowledge during tasks. This study suggests that a “horizontal” learning approach with the agent as a peer is as effective as the methods with facilitated instruction using scripts.

## Introduction

Studies in learning science emphasize the importance of interactive learning and that knowledge can be acquired through dialogue and engagement with the assistance of others (Vygotsky, 1980). In classrooms, teachers are adapting to these findings by implementing active learning strategies, in which constructive interaction in collaborative activities facilitates the learning process (Miyake, 1986; Chi, 2009). However, collaborative learning is not without its challenges, especially when novice learners struggle in group settings, resulting in gaps in intended transmission of knowledge. This problem is compounded in online learning environments because the lack of verbal/nonverbal conversational cues hinders effective communication, such as establishing common ground for language comprehension (Clark, 1996). To address these issues, there is an increasing emphasis on the use of information technology and AI to support collaboration. Both conventional methods and AI-based tutoring, including scripted prompts and adaptive facilitation, have been investigated for their potential to enhance collaborative learning (Rummel & Spada, 2007). However, direct teacher intervention may impede the initiative and independence of the students, which raises concerns.

This study presents an alternative approach, advocating for AI-assisted collaborative learning strategies that focus on a more learner-centered model and using methods of learning-by-teaching (Hatano & Inagaki, 1991). Our focus is on the framework of learner-learner reciprocal interaction using conversational agents that play the role of peer learners, and foster a sense of independence by facilitating reflection, leading to deeper understanding. This paper presents a comparative analysis between agent-based scripted facilitation strategies and the method described to determine whether it is equivalent or if the AI-integrated approach offers greater potential effectiveness and benefits.

## The scripted pedagogical approach: Facilitation from tutors

Pedagogical scripting has been studied as a useful strategy to support scaffolding for novice learners. A typical example is a knowledge-based teaching approach that uses collaboration scripts to provide learners with the necessary advice for collaborative learning (Rummel & Spada, 2007; Vogel et al., 2017). For example, Rummel & Spada (2007) used scripts on how to facilitate conversations to examine the usefulness of different types of interventions and facilitate improvement in learner-pair interaction and learning performance. In addition, Tegos & Demetriadis (2017) demonstrated the effectiveness of prompt presentations based on the academically productive talk (APT) framework, which is a theory for constructive in-class conversations. Such studies reveal that learning activities are facilitated when prompts encourage students to think deeply, clarify their reasoning, and listen purposefully.

In contrast, in the field of intelligent tutoring systems (ITS), there have been attempts to support learning activities by presenting system-based prompts that enable scripted facilitators automatically and adaptively (Anderson et al., 1995; Graesser et al., 2005; Azevedo et al., 2017). For example, the Cognitive Tutor (Anderson et al., 1995) is a learning support system that uses the cognitive architecture and the ACT-R implemented system in the algebra domain. The system monitors each step of the learner's problem-solving process in detail, and

provides adaptive support tailored to the learner's cognitive state. AutoTutor (Graesser et al., 2005) supports learning activities by providing prompts that present goals and metacognitions in response to interactions with educational conversational agents through natural language dialogs. These systems are also believed to present prompts to learners by reading their emotions from their gaze and facial expressions (D'Mello et al., 2008). Furthermore, Meta Tutor (Azevedo et al., 2017) aims to teach and train learners' metacognitive activities. The system focuses on identifying the learner's state, determining the state based on the model, and stimulating metacognition through facilitation.

However, the majority of these studies are systems for one-on-one tutoring with learners rather than systems for learner-learner collaborative learning. Recent studies have used a pedagogical conversational agent (PCA) in collaborative learning to facilitate cognitive activities and promote interaction among learners. In this series of studies, a conversational agent intervened as a third party while dyads engaged in an explanatory activity and discovered what type of facilitation methods were useful. Studies have shown that affective feedback from PCA includes positive emotions (Hayashi, 2012), using multiple PCAs to split different roles during facilitation (Hayashi, 2018), and adding real-time gaze feedback from the collaborative partner using eye trackers (Hayashi, 2020). These studies have shown that the use of such methods is effective for knowledge acquisition. However, there are limitations in the learning process, such as lack of proactivity. One problem arises from the intervention, which can be interpreted as a top-down approach to presenting information. Learners' natural conversation may be interrupted by the PCA intervention, resulting in learning activity between learners that is not self-directed. Moreover, some learners using AI-based tutoring rely too much on the system's advice, resulting in a state of passivity that hinders the development of substantive dialogue (Koedinger & Aleven, 2007). As noted in the field of learning science, it is important to promote active and interactive processes from a passive state (Chi, 2009), but it is difficult to do so only by presenting scripted prompts using PCA.

Based on the above discussion, this study proposes alternative facilitation methods that use PCAs in collaborative learning. In the next section, a method that emphasizes reciprocal teaching techniques, including learning-by-teaching and the use of AI-based technology for learner-centered tutoring, is described in the following section.

### **The reciprocal teaching approach: Learning with a virtual peer learner**

Using worked-out examples is a well-known method for novice learners to reduce cognitive load and facilitate self-explanation (Sweller et al, 1998). Profoundly applicable in disciplines including computer programming, mathematics, and physics, worked-out examples may comprise a solution, a problem formula, or the solution itself. The effectiveness of learning from such an example also depends on how learners can explain the solution method and problem-solving steps. It is important for learners to explain the logical basis of the solution presented through self-explanation (Chi, 2009). Moreover, examples presented by other collaborative learners can be used as important sources of reflection and provide opportunities to explain the activities (Miyake, 1998). This type of learning is known as reciprocal teaching, and one form of mutual learning between learners is known as learning by teaching (Palincsar & Brown, 1984). The basic idea of learning-by-teaching is that the learner assumes the role of the teacher, and learners in the role of the teacher can introduce new topics to other learners, lead discussions, and help each other solve learning tasks. There are also two basic forms of learning by teaching. In the first form, the roles are reversed, with the learner taking the role of the teacher in special activities but otherwise taking the role of learner. This form is also known as "horizontal interaction" (Hatano & Inagaki, 1991), in contrast to the top-down, facilitation-based tutoring method explained in the previous section.

Research has been conducted by ITS and AIED regarding the implementation of learning methods through AI-based tutoring (Biswas et al., 2005; Matsuda et al., 2015). For example, a study by Biswas et al.(2005), developed a system that allows learners to teach what they have learned to a knowledge-based agent known as Betty, in the form of a concept map that mimics a semantic network. If a learner has better knowledge, Betty solves the problem by reasoning based on a concept map created by the learner. If Betty does not solve the problem correctly, the learner's concept map is shown to be incorrect, and the learner must reconstruct the concept map. By interacting with Betty, the learner deepens their understanding of the concept by recognizing their errors and actively correcting their ideas. This study presents an example of a reciprocal teaching approach in which learners interact in a horizontal manner, which is different from script-based methods (horizontal interaction methods). Such methods can be used in learner-learner centered collaborative learning.

Based on the literature review on reciprocal teaching approaches and learning-by-teaching, this study centers on the application of this method within the collaborative learning setting proposed, i.e., collaborative learning by explanation activities and concept mapping. In this study, a tutoring system was developed based on the theory of learning by teaching and designed a virtual peer agent that can recognize, learn, and construct a

concept map using a cognitive architecture. In particular, it was investigated how the horizontal interaction methods can improve learning outcomes comparable to those with scripted based vertical interactions.

### Goal and hypothesis of this study

In this study, a collaborative learning scenario was developed in which learner pairs engaged in conceptual explanations using concept maps. The usefulness of the reciprocal teaching method by a virtual AI student (horizontal interaction), in addition to the metacognitive scripted facilitation by the AI tutor (vertical interaction) was examined experimentally. For this investigation, a comparative review was conducted of the following two types of agents:

- (1) Scripted metacognitive facilitation by the AI agent (vertical type).
- (2) Concept map examples from a virtual peer learner agent (horizontal type).

To investigate (2), Collaborative Concept-mapping Tutor ver. 1 (CoCot) was used. This system incorporates an AI agent implemented with a cognitive architecture, ACT-R (Anderson et al., 1995) that can learn from learner-constructed concept maps and generate examples of concept maps that learners can observe in real time and reason about based on the example. The hypotheses were as follows:

- *H1*: Learners who interact with scripted metacognitive agents perform better than learners who do not use such tools.
- *H2*: Learners who interact with virtual peer learners perform better than learners who do not use such tools.
- *H3*: Learners who interact with virtual peer learners perform as well as or better than those who interact with scripted metacognitive agents.

## Methods

### Experimental design

In this study, a collaborative tutoring system known as CoCot was used and it investigated the effects of using two components: (1) a scripted AI tutor that prompts metacognitive suggestions, and (2) a virtual AI learner that learns and generates a concept map based on the learner's output. Three experimental conditions were set for this investigation: (1) a "scripted agent condition" in which facilitation is provided, (2) a "peer agent condition" in which the AI agent generated a concept map, and (3) a "control condition" in which these features were not present. The experimental design consisted of between-subject testing.

### Participants

A total of 100 undergraduate students (mean age 19.07 years, SD 2.304; female, 60; male, 40) participated in the experiment. Participants (hereafter referred to as "learners") took part in the experiment in exchange for class credit. This study was approved by the Ethics Committee of Ritsumeikan University (Kinugasa-Hito-2020-7). Learners were randomly assigned to each condition. Four learners who did not follow the instructions of the experimenter or responded inappropriately during the task were excluded from the analysis, leaving 32 participants that were assigned to each condition.

### Task

Learners took part in an explanatory activity task, in which they had to reason and explain the mental model of a particular person based on a psychological concept. This task is an extension of that of Hayashi & Shimojo (2022). Specifically, learners were instructed to read an episode about a fictitious student who is apprehensive about the start of the new academic year, and explain this person's mental model in the context of "attribution theory." As learning materials, learners were given a text about attribution theory that explained the fundamental aspects of causal attribution. Causal attribution theory is concerned with three tendencies observed when people make causal attributions: (1) internal-external, (2) stable-unstable, and (3) controllable-uncontrollable. In this context, (1) internal refers to the attribution of a particular event to a cause within the actor, whereas external refers to attributing a certain event to a cause outside the actor. The stability in (2) indicates temporal stability, and instability refers to temporal instability. In other words, they were either stable or unstable over time. In (3), controllable means that the factors of an event can be controlled by oneself or by others, whereas uncontrollable means that the factors of an event cannot be controlled by oneself or others.

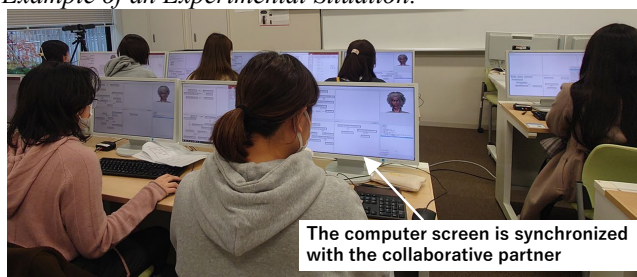
The episode used in this task was adapted from Weinberger & Fischer (2006). The text of the episode is as follows: "My father told me that the Peter family always had anxiety and inferiority complexes about toward their math teacher. I barely passed my math exam last year, but I was worried about the new school year." Learners

read this episodic text and applied attribution theory to explain why Peter felt anxious. Learners were instructed to use text-based chats and concept maps for the explanation activities.

### Procedure

The experiment was conducted in a room with 60 PCs connected via a local area network. Learners sat freely in front of a PC terminal in the room and were randomly assigned to each condition (See Figure 1.) They were then given instructions on the content of the experiment and began the experiment. In the experiment, learners first practiced using a concept map. Then, participants read the episode and learned about the attribution theory. They were instructed to provide an attribution theory-based explanation of the episode using a concept map and chat system. After this instruction, a pretest was administered to check how well the learners had understood the content of the texts. They were asked to write freely about what they knew about attribution theory.

**Figure 1**  
*Example of an Experimental Situation.*



Next, learners were asked to work independently on a task that required them to use their own concept maps to make connections between events that were related to what they had learned. This task was referred to as the “single phase.” The concept maps created in this phase were recorded on the server, reloaded from the server, and displayed in the subsequent “collaborative phase,” in which they worked with a partner. In the collaborative phase, the dyads communicated with each other through the system by creating a concept map while conversing with their partners through chat. This phase lasted 20 min. Finally, a post-test was conducted that was identical to the pre-test, to observe their growth in understanding of the concepts. *After the experiment, a debriefing was conducted and the contents of the experiment was explained, after which the experiment was terminated.*

### System

CoCot ver. 1, which was developed for this study consists of six modules: (1) a client that operates on the user side; (2) a module that draws concept maps; (3) a server that manages data and collects logs; (4) a module that performs knowledge inference; (5) an agent that handles representations of physical actions; and (6) a module for generating facilitation and presentation. For (4), a cognitive architecture, ACT-R, was implemented in the system to recognize the dyads’ concept map knowledge (nodes and links) and the reason for the new knowledge based on a rule-based model, and to generate a concept map that is different from the learner’s concept map.

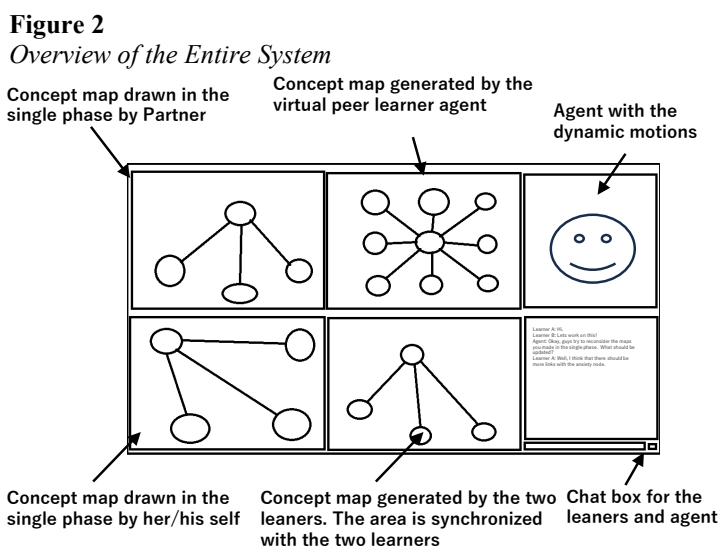


Figure 2 shows the overview of the system. The client interface displayed the following: (1) the face images of the agents, (2) the chat, (3) the concept maps created by the agents, (4) the concept maps created jointly by the dyads, and (5) the concept maps created by each participant in a single phase. The representative modules used in the current version of the system in this experiment are described in detail below.

In this study, a concept map system was implemented to enable learners to perform explanatory activities. The concept map was created using a tool developed in C#, which is a modification of Hayashi & Shimojo (2022). The functions of the tool include creating concept map elements (nodes, links, and link labels) on the screen, deleting these elements, and synchronizing one's screen with others. When creating nodes and links, participants selected from a prepared list (e.g., "internal," "external," and "anxiety" as nodes) rather than filling in their own labels. This enabled the facilitation and analysis described below to follow specific procedures.

The facilitation presentation strategy used in CoCot was adopted from Explanation Support by Pedagogical Agent (ESPA), which was developed in a series of previous studies by Hayashi & Shimojo (2022). In this study, a mechanism was implemented for presenting metacognitive facilitation prompts. The agent sent a prompt to facilitate metacognition. The facilitation prompts included (1) clarifying the learner's goals, (2) reflections on why they think so, and (3) how they might achieve effective communication (Hayashi, 2020).

To implement an AI system that generates knowledge comparable to that of a human learner, a cognitive architecture was used that is considered useful for simulating human knowledge acquisition and generation. Specifically, we used ACT-R, a representative model of cognitive architecture (Anderson et al., 1995). For this study, a model was developed to automatically generate concept maps using a computer simulation model based on Hayashi & Shimojo (2022). ACT-R uses the information of the nodes and links from the concept map generated by the learners to generate a new concept map. In doing so, the model does not simply generate the same concept map as the learner, but also uses the learner's typical knowledge to retrieve prior knowledge in the model obtained in previous studies (Hayashi & Shimojo, 2022). Specifically, ACT-R uses the knowledge of the dyads as a cue to select the node or link to generate and retrieve knowledge from its own pre-defined knowledge. In addition, the ACT-R model learns about the new nodes and links input by the learners, and the activation level for knowledge retrieval changes according to the frequency of knowledge use. With this method, as learners browse the concept maps generated by the virtual peer agent, they see concepts that are comparable to their own, but slightly "different," which is expected to facilitate reflection and allow them to rethink their own activities. This incorporates the idea of learning by teaching mentioned at the beginning of this paper (Biswas et al., 2005; Matsuda et al., 2015).

## Dependent variable

### Learning performance

The pre- and post-tests administered before and after the task were used as the dependent variable for learning performance, which measured the learner's understanding of the concept. The difference between the pre- and post-test scores was obtained for the dependent variable. The text of the pre- and post-tests was evaluated by two coders on a scale of 1 to 5 based on coding criteria from a previous study (Hayashi & Shimojo, 2022). As a result, Krippendorff's  $\alpha$  coefficient was 0.83, and disagreements were discussed and judged.

### Learning process

The coding scheme proposed by Rummel & Spada (2007) was implemented in this research. The aforementioned elements comprise the following: (1) mutual understanding, (2) dialog management, (3) information pooling, (4) consensus building, (5) task division, (6) time management, (7) technical coordination, and (8) reciprocal interaction. Each category was rated on a 5-point scale (1 = not very applicable; 5 = very applicable). Again, a second rater was assigned and the inter-coder agreement was 0.81. Differences were discussed and judged.

### Text analysis

This paper uses lexical network analysis, which is one of the methods used to qualitatively examine the content of learners' discussions. In this case, R was used to (1) draw lexical networks and (2) calculate the density, average shortest path length, and global efficiency of the networks. In conversation analysis, each learner's utterance was considered as a unit and the relationships between the words that occurred in the utterances were examined. Through morphological analysis, only nouns, verbs, and adjectives were extracted, and lexical network analysis was performed. The next step was to construct an adjacency matrix from the extracted speech data. An adjacency matrix was then constructed from the obtained speech data and calculated the (1) density, (2) average shortest path length, and (3) global efficiency.

## Results



## Learning performance

A one-way between-subject ANOVA was conducted on learning performance. The results were statistically significant under all conditions ( $F(2, 93) = 5.0130, p = .0085, \eta_p^2 = 0.0973$ ). Multiple comparisons revealed significant differences between the scripted agent and control conditions, and between the peer agent and control conditions ( $p = .032, p = .003$ ). This finding indicated that the methods in both experimental conditions were more useful in promoting learning performance than the methods in the control condition. This supports hypotheses H1 and H2. The following sections examine in detail how these conditions affect the learning process.

## Learning process

An investigation was undertaken to ascertain whether there were differences in the learning process between the two conditions. A mixed-way ANOVA was conducted on the learning process. The results show that the interaction between the factors was significant ( $F(16, 744) = 2.5565, p = .0007, \eta_p^2 = 0.0521$ ). A simple main effect test revealed a difference in the "Technical coordination;"  $F(2, 93) = 3.2309, p = .0440, \eta_p^2 = 0.0650$ ). Multiple comparisons were made, and the results showed that learners in the peer agent condition scored higher than those in the scripted agent and control conditions ( $p = .0234; p = .0406$ ).

## Lexical analysis

The analysis showed that the peer agent condition had a higher density than the control and peer agent conditions (See Table 1.) The average shortest path length was used to evaluate the speed and efficiency of information propagation in a network. In this case, the average shortest path length was shortest in the scripted agent condition, indicating that the relationships between the extracted words were propagated efficiently. This indicates that important words were used consecutively in the learners' utterances. In addition, the scripted agent condition has high global efficiency, which indicates that the nodes (words) in the network were efficiently linked to other words.

**Table 1**

*The Indices Obtained from the Lexical Network Analysis (individual task orientation high vs. low group). (Comparison of high and low individual task orientation groups)*

	Density	Average Shortest Path Length	Global Efficiency
Control condition	0.0092	50	0.0754
Scripted agent condition	0.0084	22	0.1276
Peer agent condition	0.0128	39	0.1014

The results of the network analysis of the above three indices are consistent with the results of learning performance and support the hypothesis that the two pedagogical methods are more effective than the control. In the scripted-agent and peer-agent conditions, the learners participated in an in-depth discussion regarding the task's content (topic), which was pertinent to the task's content and therefore benefited the learners' learning performance

## Discussion

### Scripted AI tutor vs. virtual AI learner

In this study, an experiment was conducted to investigate two facilitation methods based on scripted prompts and a horizontal interaction method using a virtual peer learner (Hatano & Inagaki, 1991). In the experiment, three conditions were compared: the scripted agent condition, in which the tutoring agent provided metacognitive suggestions; the peer agent condition, in which the agent generated maps based on the learner's concept map generation; and a control condition, in which no support was provided. The results of the experiment showed that the difference in learners' pre-and post-performance was larger in both conditions than in the control condition, which is consistent with the results of previous studies (Hayashi, 2012; Hayashi, 2018; Hayashi, 2020). This indicates that metacognitive facilitation had a beneficial effect on the learners' understanding of the theory they were learning in the task. In contrast, learning performance in the peer agent condition, which was challenging in this study, was equivalent to the method of presenting metacognitive facilitation in the scripted agent condition.

In the peer agent condition, the dyad concept maps were monitored in real-time, and the cognitive architecture used by ACT-R generated a concept map on the learner's screen as an instance of the AI peer learning from the learners' concept maps. Here, learners can observe the drawing of a different concept map generated by the AI peer agent, which includes the knowledge they were using. Through such observations, learners can reflect on their own concept maps and observe how the agent interprets their knowledge. The concept map may contain errors and allow learners to reflect on discrepancies, which stimulates metacognition.

Furthermore, in the peer agent condition proposed in this study, learners may have acquired metacognitive and procedural knowledge by observing of the concept map generated by the AI peer. This can be confirmed by the fact that the peer agent condition was rated higher than the control condition in terms of learning processes such as "technical coordination." Again, ACT-R does not simply redraw the learner's existing knowledge; rather, it incorporates prior knowledge as it is implemented based on (Hayashi and Shimojo, 2022) and draws while using that knowledge. Thus, learners can see maps similar to their own as well as new nodes and links generated by the agent. In the early stages of the learning process, when learners have not yet created their own maps, it is possible that the generation of examples by the AI agents may have been useful for acquiring procedural knowledge on how to draw collaboratively, which may have facilitated conversations about the technical coordination evaluated in the learning process.

### Limitations and future directions

The current model implemented by the cognitive tutor was based on the simulation model of the authors' previous study (Hayashi & Shimojo, 2022b). The model was generated based on (1) base-level activity and (2) activation-diffusion parameters (Anderson, 1995). This was a representative model that can successfully generate an average concept map, with correct and incorrect nodes and links. However, other types of maps can be modeled as ACT-Rs that can produce correct or incorrect concept maps.

One of our future tasks will be to investigate how a learner's performance changes when presented with erroneous examples and error models. To do this, the quality of the concept map presented to the learner is manipulated by changing the quality of knowledge and the thresholds for knowledge activation in the model. One question to investigate is how learner performance changes when they are presented with erroneous examples and errors. Further investigation with different types of models may lead to erroneous examples that can trigger divergent thinking. Moreover, low-performing novice learners, presenting correctly worked-out examples can promote deeper understanding.

In addition, the experiment was designed to examine the differences between two types of facilitation methods (scripted prompts vs. generative examples). Further studies combining vertical and horizontal support methods may improve the learning performance. To examine this possibility, one of our future tasks is to add an experimental condition that scrutinizes the synergistic effects of advice presentation and case generation conditions.

### Conclusion

This study investigated the effects of two facilitation methods on the collaborative learning of dyads engaging in an explanation task using concept maps. An experimental investigation was conducted on a learning support method based on metacognitive facilitation using a (1) scripted tutors agent, which has been studied in previous research, and (2) a method that incorporates a reciprocal teaching approach, such as learning by teaching and learning by observing the example generated by a virtual peer learner agent. These two approaches were incorporated into the collaborative tutoring system, CoCot ver. 1. This system was developed based on ACT-R, a cognitive architecture, and implements an agent that uses the learner's knowledge and draws conclusions when generating concept maps. To investigate the effects of the two facilitation methods, experiments were conducted under the following two conditions: (1) a scripted agent condition, in which an AI agent facilitates metacognition to the learner, and (2) a peer agent condition, in which the AI agent learns and exemplifies the learner's concept map. The experiment was conducted in a computer room with learners seated at PC terminals connected to a local area network, and a concept map creation task was performed using a server-client chat system in a computer room. The results from the experiment showed that both scripted and peer agent conditions facilitated learning performance compared to the control condition, in which no support was provided. Furthermore, in the peer agent condition, the learning process was observed with respect to the use of procedural and technical knowledge during collaborative learning. These results indicate that not only the directed presentation of scripted facilitation by the agent to the learner but also the horizontal, reciprocal teaching method by the agent acting as a virtual peer learner can be an effective method for collaborative learning situations.

### References

- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207.
- Azevedo, R., Taub, M., & Mudrick, N. (2017). Handbook of self-regulation of learning and performance, chapter Understanding and reasoning about real-time cognitive, affective, metacognitive processes to foster self-regulation with advanced learning technologies. Routledge, New York, NY.
- Biswas, G., Leelawong, K., Schwartz, D., & Vye, N. (2005). Learning by teaching: a new paradigm for educational software. *Applied Artificial Intelligence*, 19(3), 363–392.
- Chi, M. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105.
- Clark, H. H. (1996). Using language. Cambridge University Press.
- D’Mello, S., Craig, S., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner affect from conversational cues. *User Modeling and User Adapted Interaction*, 18(1), 45–80.
- Graesser, A., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618.
- Hatano, G. & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In Resnick, L. B., Levine, J., and Teasley, S., editors, *Perspectives on socially shared cognition*, pages 331–348. American Psychological Association, Washington, DC.
- Hayashi, Y. (2012). On pedagogical effects of learner-support agents in collaborative interaction. *Proceedings of the 11th international conference on intelligent tutoring systems (ITS 2012)* (pp. 22–32).
- Hayashi, Y. (2019). Multiple pedagogical conversational agents to support learner-learner collaborative learning: Effects of splitting suggestion types. *Cognitive Systems Research*, 54, 246–257.
- Hayashi, Y. (2020). Gaze awareness and metacognitive suggestions by a pedagogical conversational agent: an experimental investigation on interventions to support collaborative learning process and performance. *International Journal of Computer-Supported Collaborative Learning*, 15, 469–498.
- Hayashi, Y. & Shimojo, S. (2022a). Modeling perspective taking and knowledge use in collaborative explanation: Investigation by laboratory experiment and computer simulation using ACT-Rr. 647–652.
- Koedinger, K. R. & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264.
- Matsuda, N., Cohen, W., & Koedinger, K. (2015). Teaching the teacher: Tutoring simstudent leads to more effective cognitive tutor authoring. *International Journal of Artificial Intelligence in Education*, 25(1), 1–34.
- Miyake, N. (1986). Constructive interaction and the interactive process of understanding. *Cognitive Science*, 10(2), 151–177.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175.
- Rummel, N. & Spada, H. (2007). Can people learn computer-mediated collaboration by following a script? In *Scripting computer-supported collaborative learning*, pages 39–55. Springer US.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Tegos, S. & Demetriadis, S. (2017). Conversational agents improve peer learning through building on prior knowledge. *Journal of Educational Technology & Society*, 20(1), 99–111.
- Vogel, F., Wecker, C., Kollar, I., Fischer, F., & Schrader, J. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477–511.
- Vygotsky, L. S. (1980). *The development of higher psychological processes*. Harvard University Press.
- Weinberger, A. & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education*, 46(1), 71–95.

## Acknowledgments

This study was supported by the JSPS KAKENHI (grant numbers JP20H04299 and 23H03510).

# Twitter Talk: Informal Learning in Public Responses to Science Communication

Ashley Boone, Georgia Institute of Technology, [aboone34@gatech.edu](mailto:aboone34@gatech.edu)  
Jessica Roberts, Georgia Institute of Technology, [jessica.roberts@cc.gatech.edu](mailto:jessica.roberts@cc.gatech.edu)

**Abstract:** Scientists are increasingly turning to Twitter as a method for public science communication, creating opportunities for informal learning on social media. However, it is not clear to what extent non-scientists are engaging in meaningful learning interactions around these messages. In an analysis of 1,297 user profiles and 1,401 Twitter replies, we find that non-scientist users of Twitter are substantively engaging with educational content posted by popular scientists through retweets and replies. Our analysis shows that public responses include social talk, learning talk, and conflict talk. Scientists who regularly tweeted educational content at least five times or more per week generated more learning talk per post, suggesting that regularly tweeting educational content is one way to increase learning talk on social media. Overall, our analysis highlights opportunities for public participation in science and informal learning through expert communication on social media.

## Introduction

Digital networks, including social media platforms, have enabled connectivity that creates new modes of interaction between scientists and members of the public (Brüggemann et al., 2020; Couldry, 2016). Considering the enormous potential of Twitter (now X) to disseminate educational content, it is no surprise that many academics and subject matter experts share their expertise on Twitter (Hedman & Djerf-Pierre, 2013; Van Noorden, 2014; Walter et al., 2019), and in an era that has been dubbed “post-normal” science (Funtowicz & Ravetz, 1993; Rainey et al., 2021; Turnpenny et al., 2011), this direct communication from scientists has become a primary source of information to the public. It is critical, therefore, to understand these contributions, as well as the contributions of the public responding directly to them.

Past work has shown that content produced by scientists on Twitter is reaching audiences outside the scientific community (Côté & Darling, 2018), but we do not yet understand whether non-scientist users of these social media networks are engaging in productive “learning talk” (Allen, 2003; Roberts & Lyons, 2017) around educational content shared on Twitter. As an informal, free-choice environment, Twitter holds many of the affordances of other informal spaces for supporting self-directed learning, including engagement in joint construction of meaning through dialogue. To understand the extent to which Twitter can support productive learning talk about scientific topics, our research explores three research questions:

- RQ1: To what extent are replies to educational tweets authored by members of the general public?
- RQ2: To what extent do replies from members of the general public include substantive learning talk?
- RQ3: What factors are correlated with increased substantive learning talk on Twitter?

This paper contributes empirical evidence of substantive learning talk from non-scientist members of the general public in response to educational content authored by professional scientists using social media as a platform for public science communication. Building on previous findings that popular scientists on Twitter are followed by diverse audiences (Côté & Darling, 2018), we show that the majority of replies to educational tweets come from members of the general public and that these replies include substantive learning talk, along with social and conflict talk. Lastly, we find that regularly tweeting educational content increases the number of responses that include substantive learning talk.

## Background

### Informal learning on social media

With affordances for networked connectivity, social media has emerged as a tool for informal and collaborative learning. In addition to its use as a collaborative tool to augment formal classroom learning (Ansari, 2020; Ebner et al., 2010; Junco et al., 2013), Twitter has also facilitated informal learning about social movements (Gleason, 2013) and the exchange of socially positive information and resources (Kumar & Gruzd, 2019).

While there is evidence that learning is happening in informal online environments, the primary purpose of these online spaces may be to socialize, not to learn. Marsh suggests that social media is better served to provide social support, emotional meaning-making, and identity than to support meaning-making around scientific content

(Marsh, 2018). Social conflict is also present in these sites though Marsh observes that conflict, specifically conflict emerging polarized views on science, generates interest in posts about controversial topics (Marsh, 2018). The work presented here considers how social support, social conflict, and learning conversations around scientific content coexist in informal learning online.

## Post-normal science communication

The participation of professional scientists in science communication on digital networks reflects emerging norms of post-normal science communication (Brüggemann et al., 2020). Post-normal science describes situations that challenge traditional scientific practices, such as climate change, pandemic response, genetic engineering, prenatal diagnostics, artificial intelligence, and big data. Post-normal science is characterized by high levels of uncertainty, the development of policies that require value judgments, and urgent political decision making (Funtowicz & Ravetz, 1993; Rainey et al., 2021; Turnpenny et al., 2011).

Brüggemann et al. (2020) argue that post-normal science challenges, along with the introduction of digital networks and increasingly polarized discourse, have given rise to new roles, practices, and norms of both scientists and journalists. For scientists, communication work has transitioned away from objectivity, organized skepticism, tightly guarded peer reviews, and restricted communication with the public. These shifting norms are described as *post-normal science communication* in which the role of the scientist transitions from that of a *pure scientist* or *science arbiter* to become *advocates* and *brokers* of knowledge, ideas, and dialog. This shift requires new norms of transparency, interpretation, advocacy, and participation (Brüggemann et al., 2020). Social media can help facilitate these norms, particularly the norm of participation, by opening conversations around educational science content.

## Expert communication on Twitter

Twitter has emerged as a tool for science communication. Benefits for scientists participating in Twitter include staying updated on literature and funding opportunities, crowdsourcing ideas, promoting published work, building a support network, and participating virtually in conference conversations (Knight & Kaye, 2016; O’Keeffe, 2019; Parra et al., 2016). Some scientists turn to Twitter specifically because of its ability to directly reach non-scientific audiences (Choo et al., 2015). Fundamental Twitter features like retweeting and non-reciprocal following leverage the power of weak ties to bridge the spread of information across social, geographic, and cultural groups (Zhao et al., 2010). Despite the potential for scientists on Twitter to reach a non-scientist audience, the extent to which scientists are actually reaching beyond the scientific community has been a concern. Côté and Darling (Côté & Darling, 2018) analyzed the accounts of followers of over 100 faculty members in ecology and evolutionary biology to determine whether Twitter was being used as a platform to engage other scientists in conversation (*inreach*) or to reach new audiences (*outreach*). The authors find that while academics with fewer than 1000 followers are primarily followed by scientists and academics; academics with more than 1000 followers reach a more diverse audience. Ultimately, these results indicate that scientists have the potential to broadly disseminate scientific information after initial efforts to gain followers (Côté & Darling, 2018).

Despite the body of work that has been done in understanding how academics use Twitter, there has been little work analyzing the way that users respond to these attempts. The success of tweets is usually measured by engagement metrics that are limited to the number of people who like, retweet, and reply to content (Lundgren et al., 2022). This study seeks to better understand science communication practices by analyzing the public responses of users on Twitter through a qualitative analysis of replies.

## Study design

This research aims to characterize learning talk around scientific topics led by expert scientists on Twitter. Our goal is not to quantify and report on the full spectrum of educational content on Twitter, but to characterize a broad variety of educational engagement. To this end, we gathered recent educational tweets from scientists and qualitatively analyzed the responses to the tweets and the users who responded to them, as described next.

## The dataset

We performed a systematic review of the top five results from Google searches with key words “popular scientists on Twitter” and “science Twitter accounts” to develop a list of 159 Twitter accounts lead by professional scientists: individuals who had obtained a doctoral degree in a scientific field and had published at least once in an academic setting. We narrowed this set of 159 scientists to focus on the fifty scientists with the most followers; the authors agreed on this number to create a manageable dataset likely to have a high volume of engagement. The fifty scientists included in this dataset each had 96,248 followers or more. For each of the 50 most popular

scientists, we used the Twitter API to request up to 50 most recent tweets between September 8th and September 15th 2022. This resulted in a collection of 742 tweets from 50 different scientists posted in a seven day period.

We created exclusion and inclusion criteria to identify tweets from the scientists that included educational content. First, we excluded retweets to focus on how people responded to tweets authored rather than shared by popular scientists. Second, we excluded tweets that were not in English, as the research team was not able to effectively analyze non-English content. For each of the remaining 318 tweets, we determined if the text presented information beyond common knowledge within the scientists' domain of expertise, for example a tweet by the National Institute of Health director stating “Older adults who experience #homelessness later in life are at higher risk of premature death. #NIH-supported research at @UCSF examined how factors like housing stability, drug use, and chronic conditions, like diabetes, affected their mortality risk.”

Reviewing tweets in reverse chronological order, we identified a maximum of five most recent educational tweets from each scientist. This resulted in a collection of 69 educational tweets from 23 different scientists. 27 scientists from our initial set of 50 did not post any educational content in the seven-day period we included in our data set. For each of the 69 educational tweets identified, we requested the 100 most recent replies to the tweet and the profile details of the authors of these replies. This created a set of 2,407 replies to educational tweets, authored by 1,719 unique Twitter users. From this set, we removed comments that were not direct responses to the scientist, comments that were made by the original author, and comments that were not in English. This resulted in 1,404 replies responding directly to the educational tweets, authored by 1,297 unique users.

## Identifying members of the general public

To distinguish between scientists and non-scientists in the 1,297 users interacting with educational tweets in our dataset, we drew from the coding scheme developed by Côté and Darling (Côté & Darling, 2018) to categorize users based on their public Twitter profiles. Côté and Darling defined *science* as those whose profiles indicated they were science faculty, science students, science organizations, other scientists, and science associated groups. We follow that definition and also include health care workers as many tweets in our data set were related to health and medicine, warranting the inclusion of healthcare workers as professionals in the area of expertise. In addition, we used Côté and Darling's description of *outreach*, including educators, museums, and zoos, and *media*, referring to journalists and communications professionals. Users with non-English language profile descriptions were categorized as *unknown* and users that had no information in this profile were categorized as *blank*. The remaining users who did not fit into any of these categories were classified as *general public*. Côté and Darling's codes *applied* and *decision makers* were not well represented in our sample and these Twitter users were included in the general public category. The *general public*, *unknown*, and *blank* categories were mutually exclusive; for other categories, profiles that indicated multiple categories were coded for each. We used 203 user accounts (4.7% of the data set) to establish inter-rater reliability in applying the modified Côté and Darling coding scheme. Disagreements were resolved by discussion and the first author coded the remainder of the data set.

## Identifying learning talk

Open coding of a subset of tweets generated a preliminary codebook with fifteen categories including claims, comparisons, critique, sharing resources, questions, and social/political commentary. Our second cycle coding (Miles et al., 2020) refined these preliminary patterns into ten subcodes grouped into the categories *learning talk*, *social talk*, and *conflict talk*. We focus on these three high-level categories for the remainder of this analysis.

*Learning talk* responses demonstrated substantive engagement with the educational content of the tweet. Replies that included learning talk restated or summarized the content of the tweet, introduced new information, or critically engaged with the information presented. For example, “And the moon is separating from the earth's orbit at 3cm a year” or “still 7-10 days out so I expect changes will be likely” were counted as learning talk. *Social talk* identified comments that express appreciation, interest, humor, or wonder, without adding additional information. Statements like “Thanks for sharing!” or “I think I found my next tattoo” were counted as social. *Conflict talk* included a challenge, criticism or clash, such as “what a dummy” or “get a job”. These categories are not mutually exclusive; one tweet could include learning talk, conflict talk, and social talk. For example, “The Moon isn't even real in the first place, making your point moo(n)t.” includes learning talk evaluating claims about the moon, conflict talk challenging the assumption that the moon is real, and social talk in the humorous play on the word moon/moot. IRR ( $\kappa > 0.80$ ) was established between the authors using a subset of 83 replies from the larger dataset (5.92% of the reply dataset), which the authors discussed to resolve any disagreement. The first author then coded all 1,404 replies in the dataset.

## Findings

## Extent of engagement by the general public

While one might expect that scientists on Twitter are primarily talking to other scientists, our results replicate previous researchers' findings that popular scientists on Twitter reach a broad audience. Of the 1,404 replies, 840 (59.83%) were authored by members of the general public, 103 (7.34%) were authored by science-affiliated individuals, 69 (4.91%) were members of the media, 42 (2.99%) were from outreach, 101 (7.19%) had foreign language bios, and 275 (19.58%) had blank bio descriptions. Our results show that scientists and science organizations make up a minority of engagement with educational science content. Diverse non-scientist audiences are not only following popular scientists on Twitter (Côté & Darling, 2018), but they are also actively engaging with the educational content they share.

## Substance of replies made by the general public

Of the 840 comments authored by members of the general public, 62.56% included some form of learning talk. These responses displayed various ways of meaningfully engaging with educational content and revealed learning talk among the audience of the original tweet, though the level of learning displayed in these responses varied. In the simplest form, we found replies that summarized or restated information shared in the original tweet. Though summarizing requires only a basic command of the content, restating a tweet contradicts the purpose of the retweeting function on Twitter, suggesting that users who restate content are not only trying to spread the message but are processing the information into their own words for their audience. Comments that included both humor and learning talk indicated that humor was used to engage with the educational content of the original tweet. In one response to a tweet from astrophysicist Katie Mack on dark matter, one user wrote "Meanwhile, on the Dark Matter side: 'Huh, that's funny...It's like the universe is 15% heavier than it should be.'" The author of this comment successfully summarized the main point of the original tweet in an amusing and memorable way. The joke illustrated how humor can be not only a method of socializing but also an effective way of engaging with the material and bringing additional attention to the educational content.

Users also contributed new information in their replies. This form of learning talk required users to establish a relationship between existing knowledge and the educational information in the tweet. Users might share a personal belief ("All college and university rankings are a joke. Useless"), connect to a personal experience ("Winter is coming. My propane is \$1.00 more a gallon than last year"), or cite another source of information ("Total worker engagement is higher now than any year between 2000 and 2017."), or in some cases misinformation (e.g. links to non-scientific anti-vaccination blog posts).

Given the potential for misinformation, it is important to note that some replies were intellectually critical of information being presented by popular scientists. Asking questions and making evaluative claims did the work of critically engaging with the information shared. Evaluations of content appeared in our dataset as assessments of the reliability of the information presented or the methods used to get there. For example, in response to a summary of survey results on mothers' hopes for their young children, one user criticized the generalization saying that the 142 British mothers in the study hardly represent "mothers." In this case, the educational tweet is being critically evaluated by a member of the general public. Sometimes learning tweets, especially evaluative comments, included conflict talk, such as when critiques of methods worked to evaluate scientists' claims. One user pointed out an assumption at the root of a tweet from economist Kaushik Basu and asked "do you have data to prove this hypothesis?." A different user responded to the same tweet asking Dr. Basu to "explain with proper data." These kinds of productive conflicts can also be considered substantive learning talk demonstrating the ability of members of the general public to critically consume content on social media.

Finally, substantive learning talk in replies also included questions about the information presented. In response to the tweet about motherhood, one user asked "Do we also know what Fathers want?" Questions were sometimes directed toward the scientist who created the original tweet or to the general Twitter crowd. It is not clear how often or how well these questions are answered, and we propose this as a line of future work.

## Factors correlated with increased substantive learning talk on Twitter

Given that diverse audiences are substantively engaging with tweets, we wanted to know what factors increased substantive learning talk among members of the general public. Not surprisingly, tweets that received more comments generated more learning talk. Using a Pearson correlation coefficient, we found that the number of comments from the general public that included learning talk was strongly correlated with total number of responses collected from Twitter ( $r(67)=0.92, p<.001$ ). Additionally, our data suggests that regularly tweeting educational content is one way to increase the number of replies including learning talk. We used a two-sample one-tailed t-test assuming different variances to compare engagement with educational tweets from scientists who posted five or more educational tweets during the seven-day period with tweets from scientists who posted educational content less than five times that week. Scientists who frequently tweeted educational content received

significantly more comments that included learning talk from members of the general public on each post ( $M=9.43$  learning replies,  $SD=10.21$ ) than scientists who tweeted educational content less frequently ( $M=5.03$  learning replies,  $SD=7.77$ ) ( $t(66)=-2.03, p=0.023$ ). These scientists also received more comments in general ( $M=15.51$  total replies,  $SD=16.20$ ,  $M=7.46$  total replies,  $SD=11.53$   $t(67)=2.41, p=0.009$ ), but there was not a statistically significant difference in the number of likes or retweets their tweets received. This suggests that scientists who cultivate a reputation for posting educational content are more likely to attract an audience interested in participating in substantive learning talk.

## Discussion

### Postnormal science communication

Our analysis reveals that people are reacting to post-normal science communication practices on digital media platforms with a willingness to participate in substantive dialog around the information shared by professional scientists. Learning learning talk was entangled with social talk, and conflict talk, often in the same tweet; users shared humorous takes on educational content and criticized the methods and epistemologies of scientists. The coexistence of social talk, learning talk, and conflict talk points to the complexity of the role of the public in post-normal science communication. Just as scientists face increasing complexity and social-political entanglements in the face of post-normal science, the public responses from diverse audiences on Twitter reflect a similar intertwining of scientific knowledge with social and political conflicts.

While diverse audiences are engaging with and responding to educational material shared by scientists on Twitter, this form of participation in science remains limited. There is tension between a desire and expectation to expand the peer community and a resistance to meaningfully engaging non-scientists in scientific processes. While participation is encouraged in some cases, such as crowd work (Kittur et al., 2013) and online investigations (Belghith et al., 2022), there remains tension between participation and professional control (Lewis, 2012) in many domains. One example of the limitations of participation from our data is when people directly address the scientist author of the educational tweet with questions that go unanswered. Posing questions to scientists on social media is interesting as it indicates intellectual curiosity about science, an appeal to the authority of the scientist, and an implied expectation to respond. Participation in substantive conversations around educational content also opened the floor for misinformation, such as anti-vaccination comments to tweets about COVID-19 vaccinations. These comments reveal another tension within participation: risk of losing professional control of the norms and standards that retain the integrity of scientific knowledge production.

### Supporting science conversations on social media

The new norms of post-normal science communication have created opportunities for collaborative learning on social media. Individual scientists and social media platforms could increase substantive learning talk on social media by sharing frequent, engaging educational content and introducing feedback mechanisms in learning dialog.

#### Increasing learning talk

Individuals aiming to use social media as an educational platform might consider increasing learning talk by increasing engagement and regularly sharing educational content. Our results indicated that learning talk was highly correlated with the number of replies, suggesting that methods for increasing engagement can generate more learning talk around educational content. On Twitter, strategies for writing tweets that are clear, engaging, and informative, such as those described by Gero (Gero et al., 2021) may help scientists generate more learning talk. Additionally, providing links to content, using hashtags and mentions, using video or images, and tweeting at strategic times have been found to increase engagement on Twitter (Davis et al., 2019; Kocatepe et al., 2018; Siyam et al., 2020), and therefore have potential to generate more learning talk. Additionally, we found a statistically significant increase in learning talk in responses to tweets written by scientists who tweeted educational content five or more times in a seven day period. Frequently sharing science education content may be another method for individuals looking to support science conversations on social media.

#### Creating feedback mechanisms

A clear challenge that emerges from public responses to expert-generated science content is the lack of feedback for users. While some of the replies in the dataset clearly indicate attempts at extending or evaluating the claims made by the original tweet, correspond with higher learning outcomes in a traditional classroom, there is no reliable feedback from the platform or audience about whether the user's contribution is accurate. This lack of feedback is a challenge faced throughout informal learning interactions, such as in museums, where visitor



interpretations of an exhibit may or may not align with the curator's intentions. However, the open, public nature of Twitter amplifies misstatements, particularly when connected by an account with a large number of followers.

Following the lead of other sociotechnical systems, social media platforms could explore mechanisms for increasing feedback to participants in learning conversations. Massive Open Online Courses (MOOCs) scale up feedback and assessment mechanisms typically provided by instructors by leveraging peer feedback to increase student learning outcomes (Kulkarni et al., 2013). Peer feedback has also been an effective learning mechanism in fanfiction communities (Campbell et al., 2016; Evans et al., 2017). Content moderation, performed by other users (Crawford & Gillespie, 2016; Lampe et al., 2014), automated tools (Jhaver et al., 2019) or a specialized team (Jhaver et al., 2019; Wohn, 2019) provides an alternative strategy for increasing the quality of contributions to science conversations in social media. Developing features that generate meaningful feedback to user contributions in science conversations has the potential to increase discussion quality and support learning.

### Study limitations and future work

While our data helped answer questions about who is engaging with educational content on Twitter and how they respond, some claims are beyond the scope of this study. First, we focused on popular scientists to ensure significant activity collected in our data, but we could not confirm if the patterns described hold true for scientists who have smaller Twitter followings. Similarly, more work is needed to analyze non-English tweets and conversation threads generated by educational tweets. Our analysis of user profiles was limited to interpreting how Twitter users self-described themselves, meaning that some individuals with science PhDs may have been miscategorized as general public. Furthermore, we did not assess the quality of contributions to science conversations, choosing to focus on the prevalence rather than the quality of learning talk. Future work could assess the quality of responses to educational tweets in order to support higher-quality contributions to science conversations on social media. Finally, data for this study was collected in 2022, before the platform was purchased, reorganized, and rebranded by Elon Musk. Changes in the site policies have led to observed changes in user behaviors (Chapekis & Smith, 2023), and future work investigating patterns and prevalence of the learning, social, and conflict talk described here should take this new paradigm into account.

### Conclusion

This analysis contributes to the ongoing conversation around the enormous potential for Twitter to act as a site of informal learning. Our results extend previous work showing that experts can reach large audiences on Twitter (Côté & Darling, 2018) by showing that the audience reaches back. Our findings demonstrate that the majority of replies to educational tweets authored by scientists come from members of the general public, suggesting that popular scientists are able to engage a wide audience for science communication on Twitter. Furthermore, the content of replies indicates that the majority of users are publicly contributing substantive responses that indicate meaningful engagement with the educational content of the original tweets. Not only do our results indicate learning, they also identify these conversations as a site of social support and turmoil. The social nature of Twitter provides affordances for people to show support to scientists, express interest in science, and share humor. However, the broad network also gives rise to conflict. Learning talk, social talk, and conflict talk coexist and are often intertwined even in the same tweet, revealing the complex role of the public as an active audience to post-normal science communication. The COVID-19 pandemic is a compelling example of this phenomenon as the political polarization of masking, social distancing, and vaccination policies illustrated how public perceptions of science impact responses to urgent situations that engage scientific, political, and social domains.

We contribute findings to increase existing understandings of science communication in the context of sociotechnical systems. Our analysis suggests there is a relationship between how scientists use social media, for example, the frequency of posting educational content, and the response from public followers. Future work may be able to further detail any existing causal relationships to better inform scientists how they harness social media as an educational venue, and designers how they can support science conversation in sociotechnical systems.

### References

- Allen, S. (2003). Looking for Learning in Visitor Talk: A Methodological Exploration. In K. Crowley, K. Knutson, & G. Leinhardt (Eds.), *Learning Conversations in Museums* (0 ed., pp. 265–309). Routledge. <https://doi.org/10.4324/9781410606624-14>
- Ansari, J. A. N. (2020). *Exploring the role of social media in collaborative learning the new domain of learning*.
- Belghith, Y., Venkatagiri, S., & Luther, K. (2022). Compete, Collaborate, Investigate: Exploring the Social Structures of Open Source Intelligence Investigations. *CHI Conference on Human Factors in Computing Systems*, 1–18. <https://doi.org/10.1145/3491102.3517526>

- Brüggemann, M., Lörcher, I., & Walter, S. (2020). Post-normal science communication: Exploring the blurring boundaries of science and journalism. *Journal of Science Communication, 19*(03), A02. <https://doi.org/10.22323/2.19030202>
- Campbell, J., Aragon, C., Davis, K., Evans, S., Evans, A., & Randall, D. (2016). Thousands of Positive Reviews: Distributed Mentoring in Online Fan Communities. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, 691–704*. <https://doi.org/10.1145/2818048.2819934>
- Chapekis, A., & Smith, A. (2023). *How U.S. adults on Twitter use the site in the Elon Musk era*. Pew Research Center.
- Choo, E. K., Ranney, M. L., Chan, T. M., Trueger, N. S., Walsh, A. E., Tegtmeier, K., McNamara, S. O., Choi, R. Y., & Carroll, C. L. (2015). Twitter as a tool for communication and knowledge exchange in academic medicine: A guide for skeptics and novices. *Medical Teacher, 37*(5), 411–416. <https://doi.org/10.3109/0142159X.2014.993371>
- Côté, I. M., & Darling, E. S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops? *FACETS, 3*(1), 682–694. <https://doi.org/10.1139/facets-2018-0002>
- Couldry, N. (2016). *Life with the media manifold: Between freedom and subjection*.
- Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society, 18*(3), 410–428. <https://doi.org/10.1177/1461444814543163>
- Davis, S. W., Horváth, C., Gretry, A., & Belei, N. (2019). Say what? How the interplay of tweet readability and brand hedonism affects consumer engagement. *Journal of Business Research, 100*, 150–164. <https://doi.org/10.1016/j.jbusres.2019.01.071>
- Ebner, M., Lienhardt, C., Rohs, M., & Meyer, I. (2010). Microblogs in Higher Education – A chance to facilitate informal and process-oriented learning? *Computers & Education, 55*(1), 92–100. <https://doi.org/10.1016/j.compedu.2009.12.006>
- Evans, S., Davis, K., Evans, A., Campbell, J. A., Randall, D. P., Yin, K., & Aragon, C. (2017). More Than Peer Production: Fanfiction Communities as Sites of Distributed Mentoring. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 259–272*. <https://doi.org/10.1145/2998181.2998342>
- Funtowicz, S. O., & Ravetz, J. R. (1993). *Science for the Post-Normal Age*.
- Gero, K. I., Liu, V., Huang, S., Lee, J., & Chilton, L. B. (2021). What Makes Tweetorials Tick: How Experts Communicate Complex Topics on Twitter. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 1–26. <https://doi.org/10.1145/3479566>
- Gleason, B. (2013). #Occupy Wall Street: Exploring Informal Learning About a Social Movement on Twitter. *American Behavioral Scientist, 57*(7), 966–982. <https://doi.org/10.1177/0002764213479372>
- Hedman, U., & Djerf-Pierre, M. (2013). THE SOCIAL JOURNALIST: Embracing the social media life or creating a new digital divide? *Digital Journalism, 1*(3), 368–385. <https://doi.org/10.1080/21670811.2013.776804>
- Jhaver, S., Bruckman, A., & Gilbert, E. (2019). Does Transparency in Moderation Really Matter?: User Behavior After Content Removal Explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1–27. <https://doi.org/10.1145/3359252>
- Junco, R., Elavsky, C. M., & Heiberger, G. (2013). Putting twitter to the test: Assessing outcomes for student collaboration, engagement and success. *British Journal of Educational Technology, 44*(2), 273–287. <https://doi.org/10.1111/j.1467-8535.2012.01284.x>
- Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. (2013). The future of crowd work. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13, 1301*. <https://doi.org/10.1145/2441776.2441923>
- Knight, C. G., & Kaye, L. K. (2016). ‘To tweet or not to tweet?’ A comparison of academics’ and students’ usage of Twitter in academic contexts. *Innovations in Education and Teaching International, 53*(2), 145–155. <https://doi.org/10.1080/14703297.2014.928229>
- Kocatepe, A., Ulak, M. B., Lores, J., Ozguven, E. E., & Yazici, A. (2018). Exploring the reach of departments of transportation tweets: What drives public engagement? *Case Studies on Transport Policy, 6*(4), 683–694. <https://doi.org/10.1016/j.cstp.2018.08.007>
- Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction, 20*(6), 1–31. <https://doi.org/10.1145/2505057>
- Kumar, P., & Gruzd, A. (2019). *Social Media for Informal Learning: A Case of #Twitterstorians*. Hawaii International Conference on System Sciences.

- Lampe, C., Zube, P., Lee, J., Park, C. H., & Johnston, E. (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly*, 31(2), 317–326. <https://doi.org/10.1016/j.giq.2013.11.005>
- Lewis, S. C. (2012). THE TENSION BETWEEN PROFESSIONAL CONTROL AND OPEN PARTICIPATION: Journalism and its boundaries. *Information, Communication & Society*, 15(6), 836–866. <https://doi.org/10.1080/1369118X.2012.674150>
- Lundgren, L., Crippen, K. J., & Bex, R. T. (2022). Social Media Interaction as Informal Science Learning: A Comparison of Message Design in Two Niches. *Research in Science Education*, 52(1), 1–20. <https://doi.org/10.1007/s11165-019-09911-y>
- Marsh, O. M. (2018). “Nah, musing is fine. You don’t have to be ‘doing science’”: Emotional and descriptive meaning-making in online non-professional discussions about science [Doctoral, UCL (University College London)]. In *Doctoral thesis, UCL (University College London)*. UCL (University College London). <https://discovery.ucl.ac.uk/id/eprint/10044289/>
- Miles, Huberman, & Saldana. (2020). *Research Design and Data Management*. SAGE Publications.
- O’Keeffe, M. (2019). Academic Twitter and professional learning: Myths and realities. *International Journal for Academic Development*, 24(1), 35–46. <https://doi.org/10.1080/1360144X.2018.1520109>
- Parra, D., Trattner, C., Gómez, D., Hurtado, M., Wen, X., & Lin, Y.-R. (2016). Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 Computer Science conferences. *Computer Communications*, 73, 301–314. <https://doi.org/10.1016/j.comcom.2015.07.001>
- Rainey, S., Mormina, M., Lignou, S., Nguyen, J., & Larsson, P. (2021). The Post-Normal Challenges of COVID-19: Constructing Effective and Legitimate Responses. *Science and Public Policy*, 48(4), 592–601. <https://doi.org/10.1093/scipol/scab037>
- Roberts, J., & Lyons, L. (2017). *Scoring Qualitative Informal Learning Dialogue: The SQuILD Method for Measuring Museum Learning Talk*. 8.
- Siyam, N., Alqaryouti, O., & Abdallah, S. (2020). Mining government tweets to identify and predict citizens engagement. *Technology in Society*, 60, 101211. <https://doi.org/10.1016/j.techsoc.2019.101211>
- Turnpenny, J., Jones, M., & Lorenzoni, I. (2011). Where Now for Post-Normal Science?: A Critical Review of its Development, Definitions, and Uses. *Science, Technology, & Human Values*, 36(3), 287–306. <https://doi.org/10.1177/0162243910385789>
- Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512(7513), 126–129. <https://doi.org/10.1038/512126a>
- Walter, S., Lörcher, I., & Brüggemann, M. (2019). Scientific networks on Twitter: Analyzing scientists’ interactions in the climate change debate. *Public Understanding of Science*, 28(6), 696–712. <https://doi.org/10.1177/0963662519844131>
- Wohn, D. Y. (2019). Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3290605.3300390>
- Zhao, J., Wu, J., & Xu, K. (2010). Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E*, 82(1), 016105. <https://doi.org/10.1103/PhysRevE.82.016105>

# Exploring the Efficacy of an Online Collaborative Problem Solving Training Program

Jessica Andrews-Todd, ETS, [jandrewstodd@ets.org](mailto:jandrewstodd@ets.org)  
Emily Kerzabi, ETS, [ekerzabi@ets.org](mailto:ekerzabi@ets.org)  
Carolyn M. Forsyth, ETS, [cforsyth@ets.org](mailto:cforsyth@ets.org)

**Abstract:** Collaborative problem solving (CPS) competency is increasingly identified as important for current and future success. It is often expected that students know how to collaborate effectively, but they do have difficulty spontaneously engaging in beneficial CPS behaviors, thus signaling the need to support development of CPS skills. In the current study, we designed and piloted a three-session CPS training program that incorporated live team-based interactions in an online platform. A treatment group completed the full training program, and a control group only completed a pre- and post-assessment. Results showed the treatment group demonstrated greater use of beneficial CPS strategies than the control group at posttest. We discuss implications of the results for design of training activities and evaluation of CPS behaviors.

## Introduction and background

Collaborative problem solving (CPS) competency involves people working together to share information and pool their knowledge and effort to reach a solution to a problem (OECD, 2013). This competency has increasingly been identified as important for success in the 21<sup>st</sup> century workforce (Fiore et al., 2017; McGunagle & Zizka, 2020). Furthermore, research has shown particular benefits of engaging in collaborative activities for a number of outcomes (e.g., learning, performance, productivity) across different contexts, including school, the workplace, and the military (Graesser et al., 2018; Jeong et al., 2019; LePine et al., 2008; Lou et al., 2001). These factors have motivated interest in assessing CPS skills in an effort to prepare students for the increasing changes in how we live, work, and learn as we face new economic, social, and technological challenges. Moreover, it is often expected that students know how to collaborate effectively as evidenced by the vast amount of group work students engage in with little to no formal instruction on CPS skills. However, students do experience difficulty spontaneously engaging in beneficial collaboration (e.g., sharing important information with teammates and inviting contributions from teammates as opposed to dominating or contributing little to conversations) (Vogel et al., 2017), thus signaling a need to also support development of CPS skills. Limited research has been devoted to developing research-driven learning and formative assessment environments for learners in higher education and workplace contexts to develop and showcase CPS skills in particular. In the current study, we sought to design and pilot a CPS training program that incorporates live team-based interactions in an online platform.

## Collaboration scripts and training

Given that learners often have difficulty spontaneously engaging in effective collaborative behaviors (Andrews & Rapp, 2015; Kollar et al., 2006; Weinberger, 2011), there has been research devoted to supporting learners in collaborative contexts. One area of research has been on the development and use of collaboration scripts. Collaboration scripts are scaffolds or sets of rules that help structure interaction among learners (Mende et al., 2017). Collaboration scripts can be in the form of prompting relevant cognitive operations, creating sequences of activities to complete the goal, and distributing tasks across various team members (Kobbe et al., 2007). For example, facilitation may occur in the form of prompting a student to attempt to elaborate on a concept that another student has mentioned, which may aid in explaining (Webb et al., 2009), an important aspect of learning and achieving a shared understanding of the content in CPS (Andrews-Todd & Forsyth, 2020). Collaboration scripts have been created to aid collaborative learning in various domains, including life sciences (Noroozi et al., 2013) computer science (Demetriadis et al., 2011), economics (Huang et al., 2012), and psychiatry (Rummel & Spada, 2005) to name a few. A recent meta-analysis of 22 articles comparing collaborative learning with collaboration scripts to those without scripts discovered that collaboration scripts benefit student's collaboration skills and specific domain knowledge compared to collaborative environments without such scripts (Vogel et al., 2017).

There has also been research on developing trainings for collaboration and related skills. A large segment of this work focuses on training for teamwork skills specifically (Cortez et al., 2009; Lacey, 1996), with much of the work occurring outside of academic settings (e.g., in clinical or military settings (Capella et al., 2010;

Dalenberg et al., 2009)). In this work, different approaches to teamwork training are utilized. These include didactic lectures where learners are taught about targeted skills in a classroom context (Hobgood et al., 2010), simulation training where learners engage in simulated activities that require enacting targeted skills (Hedges et al., 2019), interactive, workshop style training contexts (e.g., working through case studies together) (Ellis et al., 2005), and approaches in which team reviews occur while performing tasks allowing teams to monitor their performance on an ongoing basis (McEwan et al., 2017; Villado & Arthur, 2013). A recent meta-analysis of teamwork trainings showed that teamwork interventions resulted in significant positive effects on team performance for all of the aforementioned methods of teamwork training; however, significant effects on teamwork were found for all training approaches except didactic education (McEwan et al., 2017).

## The current study

As much of the prior work on supporting development of collaborative skills have focused on collaborative learning and teamwork skills, we sought to design and pilot a CPS training program, with the goal of developing training that would focus on skills particular to CPS. In evaluating the CPS training program our research questions are: 1) Does the CPS training program improve learners' skills related to collaborative problem solving and use of CPS strategies? and 2) How do learners perceive the CPS training program?

## CPS training program design

The CPS training program includes three training modules programmed into the web based ETS Platform for Collaborative Assessment and Learning (EPCAL) (Hao et al., 2017). Each module consists of three components – (1) a short learning video that provides information relevant to CPS concepts, (2) hands-on practice activities to support application of knowledge gained from the learning materials and improve awareness of good practices around CPS, and (3) reflection activities for team members to discuss any issues they encountered during group activities, how issues were resolved, and things that went well. The three training modules cover topics associated with an introduction to CPS, social aspects of CPS (e.g., communication), problem solving in teams, and managing conflict in teams. The training program also contains a pre- and post- CPS assessment to evaluate changes in students' CPS behaviors from the beginning of the training to the end.

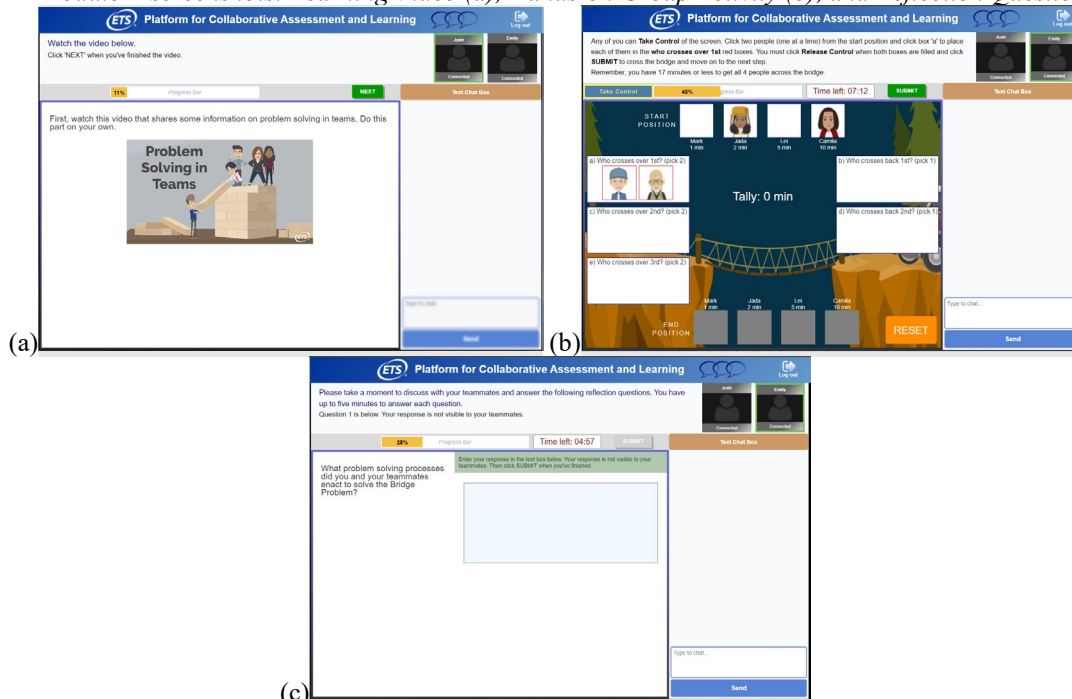
The first training module, Module A, included a learning video that introduced learners to CPS, including emphasizing its importance, how it is used in different contexts, and how it is defined. The second video in the module covered managing conflict in teams, providing information to learners about different types of conflict in teams and strategies for resolving conflict. The hands-on group task was a negotiation task in which teammates were tasked with organizing a fundraiser for charity. They had to negotiate plans for the fundraiser around five issues (i.e., what type of event to plan, the fundraising goal, the beneficiary, when to meet to plan, and what time in the semester the event will be run). Points were available to each team member for options associated with each issue. The goal was for teammates to reach the best agreement on each issue to earn the most points for themselves. This would require teammates to negotiate options and put into practice ways to manage conflicts in option selections. Module reflection questions included questions such as “How did you resolve any conflicts or differences of opinion that you experienced?” and “What are examples of positive and negative collaboration experiences you have had? What made them positive and negative?”

The second training module, Module B, included a learning video on problem solving in teams which provided information about problem solving processes to use in teams to support reaching optimal solutions (e.g., representing the problem, formulating hypotheses, planning, monitoring). The hands-on group activity was a problem-solving activity in which teammates were presented with a scenario of four campers trying to cross a bridge in 17 min or less. There were factors that needed to be considered in figuring out how to get all campers across the bridge within 17 min (e.g., each camper had a speed at which they could cross the bridge, only 2 campers could cross the bridge at one time). Module reflection questions included questions such as “What problem-solving processes did you and your teammates enact to solve the Bridge Problem?” and “What problem-solving processes (if any) did you not use that you now think would have been helpful?” See Figure 1 for screenshots from this training module.

The final module, Module C, included a learning video on social aspects of collaborative problem solving which provided information about communication (e.g., grounding) and principles for information sharing in groups. The hands-on group activity was a hidden-profile, or *jigsaw*, decision making task. In the task, teammates were part of a hiring committee for a company and tasked with rank ordering three job candidates based on their strengths and weaknesses. Each teammate was given information about each of the three job candidates. Some of the information was shared across all teammates and other pieces of information about the candidates were only provided to one teammate. Selecting the best candidate required sharing all available information to the group. This task also served as the post-assessment. Reflection questions included questions such as “What are some

things you noticed about the collaborative behaviors needed to solve the Job Candidate problem?” and “What were some challenges you experienced during the group activity?”

**Figure 1**  
Module B screenshots: Learning Video (a), Hands-on Group Activity (b), and Reflection Questions (c)



Copyright © 2023 ETS. www.ets.org.

## Pilot study

### Participants

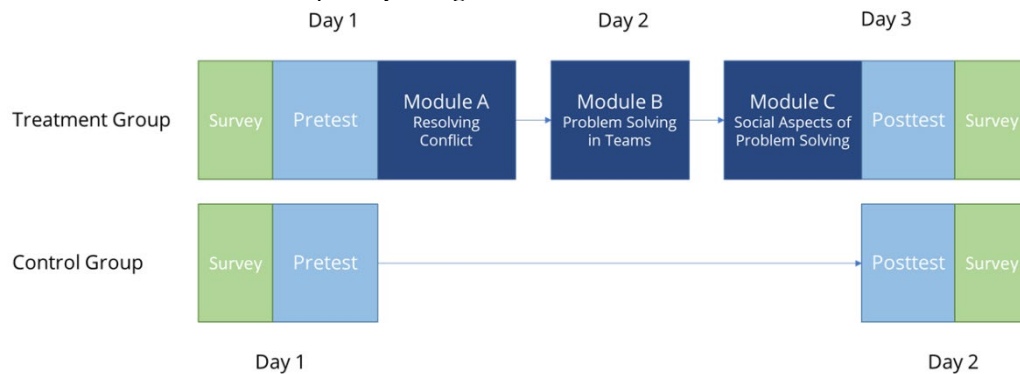
A total of 50 participants were recruited through three universities in the United States. Although some declined to provide an age, the vast majority indicated they were students between the ages of 18 and 31 ( $M = 22$  years). There were 27 people in the treatment group (9 female, 14 male, and four who declined to respond) and 23 in the control group (10 female, 10 male, and three who declined to respond). There was some attrition, with 10 participants (7 from the treatment group and 3 from the control) dropping out before the posttest was administered. Since the treatment group also had to complete the training (a greater commitment), it could be expected that more from this group would drop out than would those in the control group.

### Procedure and study design

Data were collected during the Spring of 2023. All testing sessions were completed online using the EPCAL platform, and participants communicated via a text chat box. Online collaboration was used to accommodate participants being in different locations and to support capturing log data of participants' actions and communications while completing the collaborative tasks for analyses. All participants were paid for their participation. Treatment and control group assignments were determined during participant recruitment. Participants were surveyed regarding the days and times they would be able to complete the training program and were split into similarly sized treatment and control groups based on their availability. Within each testing condition, participants were divided into teams of three or four for collaborative training and assessment. Treatment group participants completed a collaborative pretest before beginning the training program and a parallel posttest following the training program. Those in the control group took the same pretest and posttest, at the same time interval as the treatment group, but without completing the training program. Figure 2 shows the study design for the treatment and control groups. Pre and post surveys were administered to all participants,

which covered demographic information (such as age and gender), as well as their perceptions of their CPS skills before and after the collaboration.

**Figure 2**  
*Treatment and Control Group Study Design*



The pretest and posttest tasks were parallel versions of a hidden-profile task, in which each person is given both common and unique information about three different option categories. Participants discussed three different apartments available for rent in the pretest and three different job candidates in the posttest. For the treatment group, this task served as the hands-on group activity for Module C.

## Measures

Measures of group success were directly linked to the training program. Finding the best group solution in a hidden-profile task requires all team members to share and evaluate the relevant knowledge collectively, which relates to the training tenants of how to effectively share information: 1) share your unique perspectives, 2) invite others to share their knowledge, and 3) actively consider the information that is shared. An additional feature is planning out how to make the group decision and executing the strategy. To identify if participants exercised these strategies, chat logs for each participant were evaluated by two independent coders, who were not provided with the participants' group assignments. The coding rubric is shown in Table 1. This rubric focuses on conversation contributions related to sharing information, a major component in decision making (Andrews-Todd & Forsyth, 2020; Kerzabi et al., 2023; Reimer et al., 2007; Scoular et al., 2017). Sharing information is both rooted in decision making literature as a frequently observed skill strategy and was readily observed in the conversation data we collected. To facilitate coding, logfiles were subset such that all chats from a single participant were gathered and could be evaluated collectively with a single strategy code. Coders were trained on the coding rubric, and coder agreement was calculated after the initial round of coding. Coders reached a pairwise agreement of 75% ( $\kappa = .601$ ). All coding discrepancies were then revisited and reconciled.

**Table 1**  
*CPS Strategy Coding Rubric*

Score	Strategy	Description	Example Statements from an Individual
2	Attempt to fully explore options	Provide unique information from multiple options for the group to consider; ask teammates for their unique information	...lets discuss the pros and cons... ...A Pros: ... B Pros: ... [shares details]... C Cons... ...I felt that C was better than A because...
1	Attempt to partially explore options	Offer opinions and contribute to group discussion appropriately, but do not explore alternatives to their own preconceptions; may share some unique information, but generally as a justification for their own ranking decisions	...I had BAC... ...C was the worst. ...i put C at last because of... ...are we all in agreeance that C is the worst then?

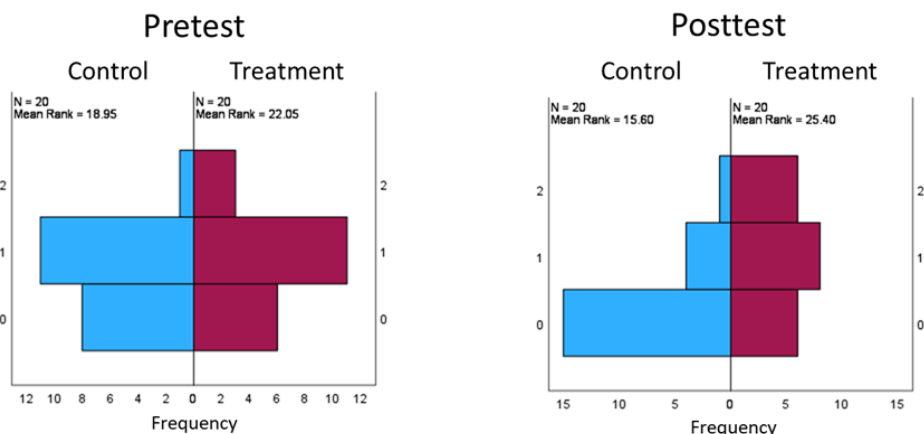
0	Do not attempt to explore options	Offer few opinions; do not explore options	...BCA is my order. ...what do you have in ranking? ...let submit then.
---	-----------------------------------	--	---

## Results

To address our research question on if the training program improved learners' CPS skills relevant for the assessments, CPS strategy usage was analyzed for the treatment and control groups between the pretest and posttest. In the control group, only three out of 20 participants showed an improvement in their CPS strategy usage between the pretest and posttest, and eight used the same strategies; only one participant attempted to fully explore the content options in the posttest. In the treatment group, 11 out of 20 showed an improvement, and three showed no change in their strategy usage; six participants attempted to fully explore options in the posttest. We also observed that there was very little variation in skill usage within a team in either condition group, on either test, meaning that team members generally used similar strategies, such that they all tried to explore options to some extent (all team members scoring 1 or 2) or they did not try to explore options (all team members scoring 0 or 1). This indicates that teams took cohesive approaches, and there was little social loafing.

Because the data were ordinal, with three CPS strategy categories and only 20 participants in each group in the posttest condition, a Mann-Whitney U test was used to compare the two groups on both the pretest and posttest scores. In the pretest, both treatment (mean rank = 22.05) and control (mean rank = 18.95) groups were statistically similar ( $U = 231.00, p = .414, d = 0.268$ ), with the majority of participants in both groups partially exploring options. For the posttest, there were statistically significant differences between the treatment (mean rank = 25.40) and control (mean rank = 15.60) groups in their use of CPS strategies ( $U = 298.00, p = .007, d = 0.923$ ). Specifically, the treatment group showed greater use of higher scoring CPS strategies than the control group. Figure 3 shows the frequencies of strategies between the pretest and posttest for these groups.

**Figure 3**  
*Frequencies of CPS Strategy Use in the Pretest and Posttest between Groups*



Survey data regarding participant experiences and perceptions were also reviewed. Participants were asked to agree or disagree (on a 5-point Likert scale) with 14 statements about their own collaborative behaviors in both the presurvey and postsurvey – this was a self-report for CPS behaviors; for example: “I share information that can help to solve the problem” and “I consider how my actions can contribute to solving a problem.” There were very few differences between how participants in either group assignment rated themselves in the presurvey. However, one notable difference was that relative to the treatment group, participants in the control group were more likely to agree or strongly agree with the statement “I am strategic in solving the problem with my teammates rather than using a trial-and-error approach.” In the postsurvey, there were more statements to which there was a difference between groups, with notable differences being that those in the treatment group were more likely to agree or strongly agree with the statements “I help to develop a plan to solve a problem” and “I help execute a plan for solving a problem” than participants in the control group.



To address our research question on learners' perceptions of the training program, those in the treatment group also responded to questions regarding the training program. Out of 20 participants that completed the full training program, 19 stated that the program was *engaging* or *very engaging*, with many citing the chat features and the ability to interact with other people while practicing CPS skills. The training was also considered *helpful* or *very helpful* in developing CPS skills by 18 participants, with one commenting that it “was helpful to learn about the important skills that will help us to succeed in our professional career journey.” Further, 16 indicated that they took from the training program *a moderate* to *a great deal* of learning, with the specific feedback that “it taught me how to communicate with people in an organized and deliberate way.”

## Discussion and implications

Human-to-human collaboration is still a novel area of measurement. As a training pilot study, we took a somewhat exploratory approach of how to effectively measure CPS as a result of training. We considered a few different approaches to viewing the acquisition of learned material via conversation chat data before deciding on strategy use, ruling out easily produced statistics like word and chat counts, as well as discretized CPS skill counts, derived from coding individual lines of chat (which tends to be a more common measure). We utilized a measure directly linked to the CPS training material. The results suggest that participants indeed show more evidence of productive strategy use in the treatment condition (in which they were taught specific CPS skills) than in the control condition. This introduces the possibility of investigating how students are engaging in CPS behaviors after learning about CPS skills themselves, which is an important addition to the current conceptual representation of CPS.

While our results may have been anticipated (the treatment group exhibited more scores of 1 and 2 on the posttest than the control group), it was not expected that the control group would more frequently exhibit strategies that were worse aligned with the task (more scores of 0 on the posttest), even worse than what was used on the pretest. We reviewed process data for additional insight. Process data has already been shown to enlighten researchers to gain valuable insights into how students are thinking about problem solving processes (Kerr & Chung, 2012). By looking at the timing variable in our logfiles, we discovered that the control group sessions tended to be shorter than the treatment group sessions on the posttest, with control group teams taking about four minutes to discuss and reach a consensus, while treatment group teams tended to take just over seven minutes. This likely indicates that members of the control group were trying to finish quickly – by why? It could have been a collective action issue, that all participants expected a short second session; however, not all control group members seemed to want to push through. We observed that some in the control group began the session by trying to explore options but buckled under social pressure from others on their team to reach an agreement without exploring all options. The take-aways we see are that individual contributions to a group may often be confounded by the contributions of other team members and that response time and other process data are important pieces of information to review when trying to get a better understanding of collaborative processes.

The value of learning CPS skills did seem to be recognized. Survey responses indicated that the training was well received by participants, with the vast majority feeling that the training was helpful and impactful for their future. Further, following the posttest, treatment group participants were more likely to agree with statements related to planning and plan execution. This may be attributable to the approach taken in this training, which was to scaffold CPS behaviors and then provide opportunities to practice them. A recent study found “an astonishing regularity” in the number of skill practice opportunities a person has and skill mastery across several domains (Koedinger et al., 2023), with more practice opportunities generally leading to skill acquisition and mastery. Although both the treatment and control groups had two opportunities to practice specific CPS strategies targeted for the hidden-profile task, only the training group exhibited strategies appropriately aligned with the task at the second practice opportunity. This finding implies that many of the skills and strategies used in successful collaboration still require support and should be paired with practice opportunities, making this particular CPS training a good model for future designs.

## Limitations and future research

There are a few areas identified for improvement. First, we received some feedback from participants that their collaborations would have been made easier if they could just talk to one another. The training was conducted through the online EPCAL platform, which, in addition to the text-based chat, does allow for participants to see and hear one another during collaboration if those features are enabled, which would have made collaboration more natural. However, for the purpose of this study, we wanted to collect all communications between team members for analysis, so only the text-based chat feature was enabled. This is a limitation of our technology; ideally, the system should include a speech-to-text feature, which would dually function to subtitle other

participants' communications and enhance the user experience while logging spoken communication as text for analysis and reporting. Further, because this was designed as a pilot study, with a modest sample of participants, our statistical inferences are limited. For example, we chose to use a simple scoring rubric, and although hierarchical in nature, the data are ordinal, which means we were precluded from using some of the more robust pretest-posttest analyses. In the future, more detailed measures of human-to-human collaboration should be sought, perhaps with more nuanced rubrics for coding individuals' strategy usage.

The training design, pairing content material with practice opportunities, seems to be a good model and will likely be the basis of larger-scale initiatives. Future CPS curricula could incorporate more opportunities to practice targeted CPS skills and, because individuals vary in their contributions, it may also be useful to have different team combinations across practices. As technologies continue to develop, it may also be possible to incorporate detectors into collaborative activities to allow researchers and teachers to know when students are performing useful or counterproductive strategies and automatically intervene when necessary. With this knowledge of group interaction and how people's strategies transform after CPS training, we can improve CPS training methodologies and measurement practices appropriate for human-to-human collaboration.

## References

- Andrews, J. J., & Rapp, D. N. (2015). Benefits, costs, and challenges of collaboration for learning and memory. *Translational Issues in Psychological Science, 1*, 182–191.
- Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior, 104*, 105759. <https://doi.org/10.1016/j.chb.2018.10.025>
- Capella, J., Smith, S., Philp, A., Putnam, T., Gilbert, C., Fry, W., Harvey, E., Wright, A., Henderson, K., & Baker, D. (2010). Teamwork training improves the clinical care of trauma patients. *Journal of Surgical Education, 67*(6), 439–443.
- Cortez, C., Nussbaum, M., Woywood, G., & Aravena, R. (2009). Learning to collaborate by collaborating: A face-to-face collaborative activity for measuring and learning basics about teamwork. *Journal of Computer Assisted Learning, 25*(2), 126–142. <https://doi.org/10.1111/j.1365-2729.2008.00298.x>
- Dalenberg, S., Vogelaar, A. L. W., & Beersma, B. (2009). The effect of a team strategy discussion on military team performance. *Military Psychology, 21*(sup2), S31–S46. <https://doi.org/10.1080/08995600903249107>
- Demetriadis, S., Egerter, T., Hanisch, F., & Fischer, F. (2011). Peer review-based scripted collaboration to support domain-specific and domain-general knowledge acquisition in computer science. *Computer Science Education, 21*(1), 29–56. <https://doi.org/10.1080/08993408.2010.539069>
- Ellis, A. P. J., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology, 58*(3), 641–672. <https://doi.org/10.1111/j.1744-6570.2005.00617.x>
- Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., Massey, C., O'Neil, H., Pellegrino, J., Rothman, R., Soulé, H., & von Davier, A. A. (2017). *Collaborative problem solving: Considerations for the National Assessment of Educational Progress* (pp. 1–83). National Center for Education Statistics.
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest, 19*(2), 59–92.
- Hao, J., Liu, L., von Davier, A. A., Lederer, N., Zapata-Rivera, D., Jakl, P., & Bakkenson, M. (2017). *EPCAL: ETS Platform for Collaborative Assessment and Learning* (RR-17-49; ETS Research Report Series, pp. 1–14). <https://onlinelibrary.wiley.com/doi/10.1002/ets2.12181>
- Hedges, A. R., Johnson, H. J., Kobulinsky, L. R., Estock, J. L., Eibling, D., & Seybert, A. L. (2019). Effects of cross-training on medical teams' teamwork and collaboration: Use of simulation. *Pharmacy, 7*(1), 13.
- Hobgood, C., Sherwood, G., Frush, K., Hollar, D., Maynard, L., Foster, B., Sawning, S., Woodyard, D., Durham, C., & Wright, M. (2010). Teamwork training with nursing and medical students: Does the method matter? Results of an interinstitutional, interdisciplinary collaboration. *Quality and Safety in Health Care, 19*(6), e25–e25.
- Huang, H.-W., Wu, C.-W., & Chen, N.-S. (2012). The effectiveness of using procedural scaffoldings in a paper-plus-smartphone collaborative learning context. *Computers & Education, 59*(2), 250–259.
- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review, 28*, 100284.
- Kerr, D., & Chung, G. K. W. K. (2012). Identifying key features of student performance in educational video games and simulations through cluster analysis. *Journal of Educational Data Mining, 4*, 144–182.

- Kerzabi, E., Lawless, R., & Wang, Y. (2023, April). Annotating CPS Dialogue: Rubric Development and Category Refinement. National Council on Measurement in Education (NCME), Chicago, IL.
- Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, 2(2–3), 211–224. <https://doi.org/10.1007/s11412-007-9014-4>
- Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences*, 120(13), e2221311120. <https://doi.org/10.1073/pnas.2221311120>
- Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts—a conceptual analysis. *Educational Psychology Review*, 18(2), 159–185.
- Lacey, P. (1996). Training for Collaboration. *British Journal of In-Service Education*, 22(1), 67–80. <https://doi.org/10.1080/0305763960220107>
- LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61(2), 273–307.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449–521.
- McEwan, D., Ruissen, G. R., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The effectiveness of teamwork training on teamwork behaviors and team performance: A systematic review and meta-analysis of controlled interventions. *PLoS One*, 12(1), e0169604.
- McGunagle, D., & Zizka, L. (2020). Employability skills for 21st-century STEM students: The employers' perspective. *Higher Education, Skills and Work-Based Learning*, 10(3), 591–606.
- Mende, S., Proske, A., Körndle, H., & Narciss, S. (2017). Who benefits from a low versus high guidance CSCL script and why? *Instructional Science*, 45(4), 439–468. <https://doi.org/10.1007/s11251-017-9411-7>
- Noroozi, O., Teasley, S. D., Biemans, H. J. A., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning*, 8(2), 189–223. <https://doi.org/10.1007/s11412-012-9162-z>
- OECD. (2013). *PISA 2015 collaborative problem solving framework*. OECD Publishing.
- Reimer, T., Kuendig, S., Hoffrage, U., Park, E., & Hinsz, V. (2007). Effects of the information environment on group discussions and decisions in the hidden-profile paradigm. *Communication Monographs*, 74(1), 1–28. <https://doi.org/10.1080/03637750701209947>
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, 14(2), 201–241. [https://doi.org/10.1207/s15327809jls1402\\_2](https://doi.org/10.1207/s15327809jls1402_2)
- Scoular, C., Care, E., & Hesse F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement*, 54(1). <https://doi.org/10.1111/jedm.12130>
- Villado, A. J., & Arthur, W. (2013). The comparative effect of subjective and objective after-action reviews on team performance on a complex task. *Journal of Applied Psychology*, 98(3), 514–528.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477–511. <https://doi.org/10.1007/s10648-016-9361-7>
- Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., & Melkonian, D. K. (2009). 'Explain to your partner': Teachers' instructional practices and students' dialogue in small groups. *Cambridge Journal of Education*, 39(1), 49–70. <https://doi.org/10.1080/03057640802701986>
- Weinberger, A. (2011). Principles of transactive computer-supported collaboration scripts. *Nordic Journal of Digital Literacy*, 6(3), 189–202. <https://doi.org/10.18261/ISSN1891-943X-2011-03-06>

# I Know Where You are Looking! Broadening the Notion of Annotation in a Collaborative Immersive Simulation Environment

Yiqiu Zhou, University of Illinois Urbana-Champaign, yiqiu3@illinois.edu  
Jina Kang, University of Illinois Urbana-Champaign, jina.kang@illinois.edu

**Abstract:** Building on collaboration as a reciprocal process, this paper proposes *re-annotation* as a lens to examine engagement with digital annotations in computer-supported collaborative learning environments. Shifting from traditional annotation metrics, such as quantity and quality, we focus on dynamics of how annotations are attended to and revisited. Using a cross-device simulation platform, this study analyzes re-annotation behaviors through four temporal dimensions (frequency, breadth, contiguity, and recurrence) across groups with varying learning performances. Findings highlight shorter intervals between annotations and subsequent re-annotations in high-performing groups, indicating prompt response and a potential measure of attention coordination. Additionally, high-performing groups revisited their own annotations more frequently, indicative of reflective activity. Extending the investigation beyond annotation creation, this work enriches the theoretical understanding of social annotation and informs the design of annotation tools. While providing promising insights, future research is needed, incorporating qualitative insights to fully unpack the role of (re-)annotations in collaboration dynamics.

## Introduction

Computer-supported collaborative learning (CSCL) environments hold great promise for facilitating deeper learning through social knowledge construction. One of the key affordances of these environments is the ability to externalize and share thoughts through annotations (Chan & Pow, 2020; Krouska et al., 2018; Novak et al., 2012). Social annotations not only help express their evolving understanding through external representation but also bridge the transition from individual to collective understanding. Previous studies have extensively examined annotation in CSCL environments, primarily focusing on the quantity and quality of annotations, and their impact on the overall collaboration quality or learning results (Sun et al., 2023). However, these investigations have centered around the initiator of annotations (i.e., annotator). How annotations are subsequently received, attended to, and interpreted by their peers has remained relatively underexplored. It is often taken for granted that learners are generally attending to and engaging with peers' annotations. However, this assumption does not always hold (Wise et al., 2014) and overlooks another important dimension: the manner in which learners engage with existing annotations made by peers. Studies in online discussion platforms have paid attention to the recipients of online posts and identified the existence of a silent cohort, termed as *lurker*, who do not actively engage with peers' posts (Sun et al., 2014). Similarly, within other CSCL environments, varying degrees of interaction with peer's ideas and contributions likely occur. Since collaboration is a social process that involves constructing shared meaning through reciprocal engagement (Akkerman et al., 2007; Gerry et al., 2006), there is a need to analyze how annotations are responded to.

Bridging this gap, our study shifts the attention towards the followers' reactions to annotated objects. Specifically, we focus on the behavior of learners in a cross-device astronomy simulation platform when they revisit and re-highlight existing annotations on celestial objects. We term this phenomenon as "re-annotation." This behavior signifies that previous annotations have been noticed, which is indicative of group awareness and could potentially pave the way for the further establishment of shared understanding.

Since time plays a fundamental role in unpacking collaboration as an analytical tool, analyses that take temporal properties into account are critical for generating an understanding of how collaboration does (and does not) take place (Reimann, 2009). One important contribution of temporal analyses is that they can help unpack and explain the possible reasons behind differences found in group learning outcomes (Kapur, 2011). As such, this paper will examine key temporal properties of re-annotation, including frequency, temporal contiguity/lag, and recurrence. Our investigation of temporal metrics aims to discern the variations among various performance groups and provide deeper insights into the notion of annotation. Ultimately, this richer understanding of how groups leverage annotations to construct knowledge offers the potential to support such collaborative processes. Accordingly, our research questions are as follows: RQ1: How do the *quantity* and *breadth* of annotation as well as re-annotation behaviors vary between high and low-performing groups? RQ2: How does *temporal contiguity* of re-annotation (i.e., time gap between annotation and subsequent re-annotations) vary among high and low-

performing groups? RQ3: How does the *revisitation* on previous annotations vary among high and low-performing groups?

## Literature review

### Social annotation

Social annotation (SA) tools within CSCL settings, evolving from digital annotation tools, enable users to create annotations such as comments, notes, explanations, or other types of external remarks on online resources, and share these information within a community (Krouska et al., 2018). SA technology emerges as a multi-purpose system that facilitates both information sharing and social interaction, by allowing learners to engage with the annotated resources without altering the original material (Novak et al., 2012). These tools allow for the addition, editing, and modification of information, thus cultivating a richer understanding through the aggregation of diverse insights from various learners. The collaborative nature of SA tools encourages annotation sharing with peers or group members, catalyzing discussions and enhancing knowledge sharing as learners collaboratively explore and exploit valuable ideas (Yang et al., 2011). Therefore, SA is not only about annotating but also about sharing and interacting with annotations in a way that enhances knowledge construction. From this perspective, SA tools extend beyond mere information-sharing platforms, serving as social platforms that enable meaningful interactions and knowledge co-construction within groups.

Empirical studies have demonstrated the promising effects of SA within CSCL contexts. Annotation tools have been found to promote knowledge sharing by facilitating communication and collaboration in online group reading activities (Yang et al., 2011), enhance reading comprehension through quality annotations (Jan et al., 2016), improve learning outcomes in collaborative inquiry-based learning (Chan & Pow, 2020), foster high-level cognitive and metacognitive activities in group inquiry projects (Li et al., 2015), and support information organization and idea argumentation in collaborative writing (Passig & Maidel-Kravetsky, 2016). While these studies illustrate SA's potential advantages across various learning settings, the majority of these papers primarily focus on analyzing the quantity and quality of annotations produced by initiators (i.e., annotators). It is essential to delve deeper into activities beyond the generation of initial annotations.

In particular, there are limited studies that have examined how learners, other than annotators, interact with peer annotations. Only a handful of studies have analyzed annotation replies, underscoring the significance of further exploration in this area. For example, a web-based annotation system, PAMS, enables students to query about related annotations, and students have confirmed discussing with other annotators in the study group through related annotations in the questionnaires (Yang et al., 2011). Another study reported the statistical values of annotations and replies in a self-developed SA tool named the Web Annotation and Sharing Platform. Interestingly, their results suggest that both the mean and maximum number of annotations and replies were good indicators of the interactions (Chan & Pow, 2020). This finding suggests that analyzing responses subsequent to the initial annotation can serve as a window into the collaborative dynamics within SA-enabled learning environments. More research is needed to comprehensively explore the annotation reception and exchange. As Gao (2013) observed, there is limited understanding regarding the optimal conditions and contexts for implementing these SA tools.

This calls for a more in-depth examination of the reception of annotations - how learners engage with, interpret, and respond to annotations made by their peers. Delving into this can provide insights into how the exchange of annotations contributes to the collective knowledge construction. By broadening our focus beyond the mere adoption of SA, this study aims to achieve a more holistic understanding of collaborative learning.

### Conceptualizing re-annotation behavior

Grounded in socio-cognitive and sociocultural perspectives, collaboration is positioned as a social process centered on constructing and maintaining shared meaning interactively (Akkerman et al., 2007; Gerry et al., 2006). A key aspect of productive collaborative learning is engagement in shared knowledge practices that involve reciprocal efforts, where learners build on idea expressions made by others (Hakkarainen et al., 2013). These expressions can manifest in various forms such as verbal articulation, discussion posts, or annotations, supporting learners to become aware of, and attend to, the ideas forwarded by their peers. Knowledge construction occurs through such iterative cycles of individual processing and contribution, coupled with mutual engagement and response that negotiates shared meaning (Järvelä & Hadwin, 2013).

Building upon this foundation understanding of collaborative learning, this paper proposes the concept of *re-annotation* as a lens through which to analyze responsive behavior to digital annotations. This draws inspiration from the notion of online listening post revisiting behavior as described by Wise et al. (2013, 2014). Authentic discussion and collaborative knowledge construction is an iterative cycle of call and response, with

turn-taking and attention being integral to meaningful exchange (Wise et al., 2013). Re-annotation extends this notion of participation into the realm of annotated-enabled CSCL environments. Regardless of the varying terminologies used in different contexts, these concepts distinguish between those who “produce” and “consume” ideas in an online space (Muller et al., 2010), highlighting diverse ways in which learners navigate and contribute.

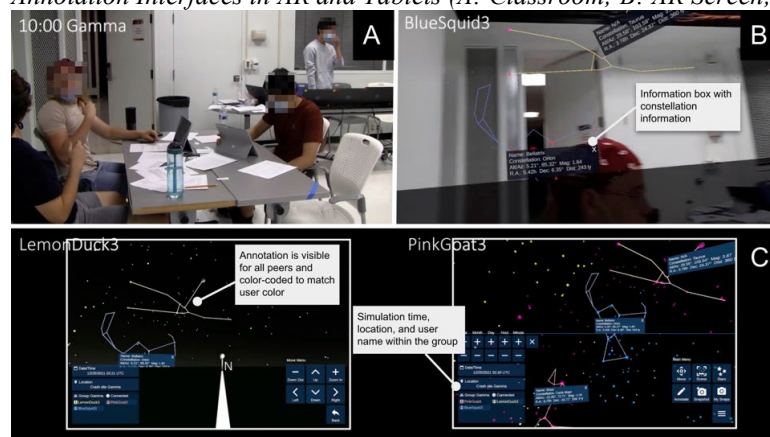
In the context of SA-enabled environments, re-annotation emerges as a critical form of participation that is complementary to annotation production. If we consider annotation as the externalization of one’s attention or ideas, then re-annotation reflects the acknowledgment of a peer’s attention and/or the internalization of a peer’s ideas. The act of re-annotation, similar to revisiting and responding to online posts, signifies a deeper level of group interaction, denoting a reflective or responsive engagement with the peer-generated annotations. This contrasts with the “lurking” behavior as described in earlier work by Nonnecke et al. (2004) and Rafaeli et al. (2004), where individuals passively consume content in online spaces without actively responding to and building on it. Re-annotation manifests as an active acknowledgment and engagement with the shared ideas, conveying “I know where you are looking at” and more. This behavior suggests that previous annotations have been acknowledged, thereby confirming shared attention, an indicator of group awareness (Dourish & Bellotti, 1992). Re-annotation further helps establish the common ground where a higher level of communication such as negotiation, elaboration, and co-construction can occur, foster social regulation activities to reduce divergent knowledge and increase collaboration quality and learning performance (Bodemer & Dehler, 2011).

## Methods

### Research context

The learning platform, Connections of Earth And Sky with Augmented Reality (CEASAR), is designed to provide an immersive and interactive learning experience through exploration of digital celestial objects to support astronomy education. It provides access to a star and constellation database and allows for location and time modification, facilitating observations from three different perspectives: Horizon, Star, and Earth. This cross-device system is networked between tablet and AR, allowing all group members to annotate the night sky by highlighting specific constellations of interest, share location information, and transition to a peer’s locations and perspectives by clicking on their username. Annotation by selecting the star will trigger an information box next to it, as shown in Figure 1. This box displays its name, position magnitude, and constellation membership if applicable. When the selected star belongs to a constellation, the entire constellation gets automatically highlighted and visible to the whole group. To foster accountability and recognition within the group, each group member was assigned a unique color to track their star selections and constellation highlights. Utilizing this platform, students were expected to solve a multi-task problem called “Lost at Sea”, where students are challenged to estimate the approximate landing site of a crashed space capsule based on celestial navigation. Annotations are instrumental for solving four sub-tasks including identifying the hemisphere (Task-1), locating directional references (Task-2), and estimating latitude and longitude (Task-3&4). These highlighted stars or constellations serve as reference points accessible to the whole group, enhancing spatial understanding and task coordination.

**Figure 1**  
*Annotation Interfaces in AR and Tablets (A: Classroom; B: AR Screen; C: Tablet Screen)*



### Participants

The participants included 77 undergraduates enrolled in an introductory astronomy course from a mid-western university in the United States. This course is designed for non-majors to fulfill general education requirements, with one main lecture and seven smaller discussion sections. To ensure diverse representation, participants were randomly selected from three of these seven discussion sections. Students participated in over three weekly 50-minute lab sessions: 2 introductory and 1 immersive learning simulation sessions. This study only focused on the simulation session where the participants engaged in the Lost at Sea activity. Each self-assigned group of 3-4 students was given one AR headset and two touch-based tablets, resulting in a total of 25 groups. They were expected to utilize the simulation platform to solve tasks collaboratively during the last session. Participants had the option to work at unmonitored tables. Due to data unavailability, 16 groups remained for further analyses.

## Data

### Data source

Data for this study is derived from log files and assessment scores. Log data was collected to capture students' annotation-relevant behaviors. Interaction logs were recorded as rows of events and generated each time a student interacts with the platform, where event = {Username, Groupname, Device, Activity, Event, UTC time, Heading vectors, Simulation time, Crashsite, Location, Scene, Selected object, Selected star}.

Individual pre-/post-paper-based assessments were collected to measure learners' conceptual knowledge before and after completing the Lost at Sea activity, each taking approximately five minutes to complete. Students received an open-ended question evaluating their understanding of latitude and longitude calculation "Write as much as you know about the steps for calculating the latitude and longitude based on the stars visible in a given location." Responses were scored from 0 to 2 based on completeness and accuracy. Normalized learning gains were calculated from the pre-/post-tests:  $(\text{post} - \text{pre}) / ((\text{post}_{\text{max}}) - \text{pre})$ . Analysis of group average gains revealed a gap between 0.1667 and 0.3125, which potentially indicates a divergence in knowledge gain. We subsequently categorized 16 groups into low-achieving ( $n = 7$ ) with a range of  $[-0.2222, 0.1667]$  and high-achieving groups ( $n = 9$ ) with a range of  $[0.3125, 0.6875]$ .

### Annotation and re-annotation identification

The CEASAR platform supports two forms of annotation: highlighting and line drawing. Given the infrequent use of line drawing in our dataset, this analysis focuses on highlighting as the primary form of annotation. Re-annotations are identified by repeated annotations or highlights made on the same constellations by different devices within a certain time threshold. Given the unique setting of CEASAR, where only one constellation can be highlighted per device at any given moment, setting a time threshold helps identify re-annotations that are potentially related to the annotations. The threshold is set at 120 seconds, a decision informed by our previous study on episodes (i.e., continuous verbal discussion with a pause of less than 20 seconds) (Planey et al., 2023; Zhou & Kang, 2022). Our analysis revealed an average episode length of 96.5 seconds, with 75% of the episodes lasting under 120 seconds. Although a threshold of 180 was considered, we selected the 120-second threshold to focus on shorter episodes to derive meaningful re-annotations, minimizing the potential noise that may arise from a larger threshold. This threshold ensures relevance between annotations and re-annotations while minimizing random coincidences.

## Data analysis

The overarching goal of this analysis is to capture various behavioral dimensions and related metrics that reflect how students interact with annotations within CEASAR. We examined the following key dimensions of re-annotation behavior: quantity, breadth, temporal contiguity, and reflectivity inspired by the framework of online discussion participation (Wise et al., 2014).

First, since previous studies found out positive relationships between the overall *quantity* of annotations generated in the online space and the learning performance (Chan & Pow, 2020), there can be variations in the quantity of re-annotations. We quantified the annotations and re-annotations by aggregating their total frequency throughout the session to index the interaction volume within groups, which may indicate engagement level and collaborative efforts. Mann-Whitney U tests were then employed to examine significant differences between high and low-achieving groups in terms of annotation and re-annotation frequency.

Second, in an open-ended environment where students navigate the simulation to highlight constellations of interest, the variety of constellations annotated and re-annotated can serve as a proxy for *breadth* of engagement. We calculated the total number of unique annotated and re-annotated objects each group interacted with, evaluating the extent of their celestial exploration. Comparative analyses (e.g., Mann-Whitney U tests) were utilized to discern any differences in the breadth of platform engagement across performance categories.

Third, the *temporal contiguity* during which students attend to already annotated/highlighted resources can indicate the degree to which they follow other’s attention. We calculated the time gaps between initial annotation and subsequent re-annotation to gauge the immediacy of responses to peer annotations. The distribution of time gaps was visualized to depict the temporal dimension of group interactions. Mann-Whitney U tests were then applied to evaluate the potential correlations between temporal contiguity and performance.

Last, we tracked the frequency with which students revisited previously annotated constellations, interpreting this as an indicator of *reflectivity*. This pattern can suggest a reflective process where students re-examine and consolidate their understanding, or it may indicate potential ambiguity, leading to a reassessment of previous notes. We measured the average count of revisits per user to quantify such behavior. Descriptive statistics were then reported to characterize the prevalence of such revisits among high and low-performing groups, informing whether reflectivity indicator generally leads to improved clarity and learning outcomes.

## Results

### RQ1: Quantity and breadth of annotations

The first research question examined differences in annotation quantity and breadth between high and low-performing groups. Table 1 summarizes the statistical results regarding a total and unique number of annotations, as well as re-annotations for high and low-performing groups. After checking the normality, we employed the Mann-Whitney U as a non-parametric test. We also reported the rank-biserial correlation as effect size, providing insights into the magnitude and direction of the observed difference.

In comparing annotation behaviors between the two performance categories, we observed differences in both total and unique numbers of annotations. As shown in Table 1, high-performing group showed larger means for a total number of annotations and unique annotations compared to the low-performing group. However, these differences were not statistically significant for either total or unique annotations. While observable quantitative differences existed between groups, the non-significant statistical outcomes may have been due to the small sample size and insufficient statistical power.

**Table 1**

*Metrics of Annotation and Re-annotation between High and Low-Performing Groups*

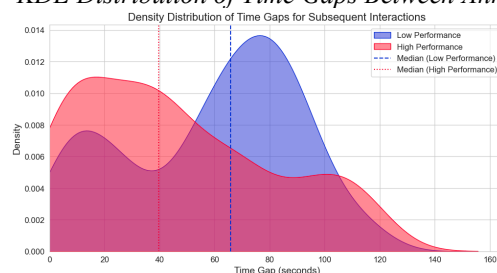
Metric	High	Low	Statistics
Total annotation (count)	76.3 <sup>a</sup> (63.0 <sup>b</sup> )	48.3 (19.4)	$U^c = 21.5, p = .314, r_{rb}^d = .303$
Unique annotation (count)	26.2 (17.3)	17.7 (4.6)	$U = 22.5, p = .366, r_{rb} = .275$
Total re-annotation (count)	22.9 (49.5)	45.9 (113.5)	$U = 22.5, p = .365, r_{rb} = .275$
Unique re-annotation (count)	3.2 (2.5)	1.7 (1.6)	$U = 20.0, p = .234, r_{rb} = .343$
Time gap of re-annotation (seconds)	46.2 (34.9)	58.4 (31.7)	$U = 40374.0, p < .001, r_{rb} = .216$

Note. <sup>a</sup> average; <sup>b</sup> standard deviation; <sup>c</sup> Mann-Whitney U test was employed as our data does not follow a normal distribution; <sup>d</sup> rank biserial correlation (effect size).

### RQ2: Temporal contiguity between annotations and re-annotations

Our second research question looks at the temporal contiguity of re-annotations. Through Mann-Whitney U test, significant differences were found across two performance categories (refer to ‘time gap of re-annotation’ in Table 1). For a more granular visualization of the time gap distribution, we employed the Kernel Density Estimate (KDE) plot. Figure 2 presents a smoother depiction of the time gap between annotations and subsequent re-annotations. Dashed lines mark the median time gap for each performance category, representing the central tendency. Notably, a shorter time gap was observed in the high-performing group's re-annotation. This implies students were more closely following and acknowledging the contributions of their peers, potentially leading to idea exchanges and enhanced group cohesion.

**Figure 2**  
*KDE Distribution of Time Gaps Between Annotations and Re-annotations*





### RQ3: Revisitation on annotations

The third research question examines the reflective behavior as students revisited their own previously annotated constellations. Table 2 details the descriptive statistics of annotation revisiting behaviors. Notably, high-performing groups are characterized by higher average, minimum, and maximum values. Despite the larger standard deviation, the overall tendency to exhibit more annotation revisiting behaviors is apparent among students in high-performing groups. The prevalence of revisitation behavior potentially serves as a problem-solving strategy that supports higher-order cognitive processes. This includes the validation of previous ideas, re-evaluation based on peer feedback, or modifications of current understanding.

**Table 2**  
*Descriptive Statistics of Annotation Revesting Behavior (count)*

Performance	Mean	SD	SE	Q1	Q2	Q3	Minimum	Maximum
High	14.5	13.0	4.3	5.5	22.0	13.5	5.0	44.5
Low	5.9	3.3	1.3	4.0	5.5	6.0	3.0	13.0

### Discussion

This study aims to provide an in-depth analysis of collaborative interactions within the CEASAR simulation, offering a nuanced understanding of different group's patterns of re-annotation in a cross-device simulation platform. Although no significant differences emerged in annotation and re-annotation frequency or breadth across performance categories, potentially due to our limited sample size, the descriptive analysis suggests a trend. Specifically, high-performing groups tended to annotate a wider variety of constellations more frequently and revisit their previous annotations more often.

Beyond the overall frequency or breadth, we observed notable differences in the temporal continuity with which students attend to and re-highlight the peer-generated annotations. The metric of temporal contiguity, gauged by the time lag between initial annotations and subsequent re-annotations of celestial constellations, acts as a proxy for students' attentiveness and responsiveness to their peers' inputs. In face-to-face discussions, prompt responses often signify active engagement and attentiveness to the ongoing discourse. Similarly, in virtual spaces, timely responses to annotations could foster a more coherent discussion, catalyzing a cumulative building of ideas and insights through idea exchanges. On the contrary, delayed responses or lack of re-annotations may hinder the flow of interaction and discussion, possibly leading to fragmented discussions.

This temporal dimension echoes research on learners' perception of interactivity in virtual learning environments, where response times correlate with reduced perceived interactivity, operationalized as perceptions of navigation, responsiveness, and the overall interactivity of the online space (Barboza & Da Silva, 2016; McMillan & Hwang, 2002). Such delays can lead to learners feeling less connected and responsive, potentially impacting the overall learning experience. While the study primarily focused on student-instructor relationships, the importance of timely responses may extend to student-student relationships as well. Moreover, the notion of temporal distance becomes critical in asynchronous settings, where delays in responses can lead to frustration and extended problem solving time, resulting in a less effective collaboration (Morrison-Smith & Ruiz, 2020). Synchronous settings, however, do not automatically solve the temporal challenge nor guarantee a consistent awareness of peer actions within a shared problem space. Despite the opportunity for real-time interaction, the "attention distance" still exists, similar to the challenge of maintaining group awareness (Nicolaescu et al., 2013). Even in CSCL environments with sophisticated communication tools, desirable peer interactions and social dynamics cannot be taken for granted (Kirschner et al., 2015; Kreijns et al., 2003). This issue becomes more pronounced in CEASAR where students have the freedom to share and switch devices among them. Such a situation introduces a layer of complexity as multiple learners engage with multiple devices, using diverse tools and resources to navigate the space and solve problems (Martinez-Maldonado et al., 2016). Our study suggested that the time lag in revisiting peer annotations may serve as an indicator of coordination efforts to reduce attention distance and achieve mutual awareness. To further validate this interpretation, we plan to include video analysis for more nuanced insights in future studies.

Furthermore, the interplay between technological affordances and emergent collaborative dynamics merits a deeper exploration. In our context, annotation becomes the means through which students express their understandings and build on each other to develop collective understanding. The observed lag in re-annotation reflects collaboration dynamics, where learners interact with and through these digital tools (i.e., annotation), offering insights into group engagement and coordination. Prompt re-annotation suggests efficient coordination, while longer gaps or lack of re-annotation point out areas where the intended affordances may not be fully realized, impeding knowledge co-constructions. These observations underscore the important role of technology in shaping

collaboration. Future study could benefit from theoretical perspectives such as distributed cognition (Hollan et al., 2000) to further examine how collaboration is mediated by technology. Applying such a framework can deepen our understanding of how specific designs facilitate or hinder productive collaboration among groups of learners. Embedding new design features that facilitate the visibility and salience (e.g., notification for new annotations or updates to existing ones), allow for historical tracking and exporting annotations, and support interactive functions like comments or upvotes could impact the collaborative interactions.

## Conclusion

This study provides novel insights into annotation and re-annotation within a cross-device simulation platform. Specifically, our study makes a unique contribution by explicitly examining the often-overlooked aspect of re-annotation. While prior research primarily focused on the creation of annotation itself and its impact on learning outcomes, our investigation delves into the dynamics of how annotations are received and revisited by peers. We observed that students engaged with peer annotations not just as passive viewers but as active participants, re-highlighting the same constellation. This interaction possibly reflects underlying collaboration mechanisms, including the confirmation of shared perception, acknowledgment of peer's attention, and the orchestration of collaborative efforts within the group. While no significant differences emerged in the overall frequency or breadth of (re-) annotations, temporal contiguity proved distinguishable among high and low-performing groups. This time gap between annotation and subsequent re-annotation may serve as a proxy for how closely students are following their peers' attention and contributions. Notably, high-performing groups exhibited significantly shorter gaps, suggesting a correlation between prompt interactions and productive group dynamics. Such behavioral pattern suggests that a higher level of synchrony and group cohesion can lead to more fluid discussion flow and efficient problem-solving. This exploratory study provides preliminary insights for future research to employ qualitative approaches (e.g., video analysis and group interview) to fully unpack the intention and role of the (re-) annotations.

## References

- Akkerman, S., Van den Bossche, P., Admiraal, W., Gijsselaers, W., Segers, M., Simons, R.-J., & Kirschner, P. (2007). Reconsidering group cognition: From conceptual confusion to a boundary area between cognitive and socio-cultural perspectives? *Educational Research Review*, 2(1), 39–63. <https://doi.org/10.1016/j.edurev.2007.02.001>
- Barboza, E. J. S., & Da Silva, M. T. (2016). The Importance of Timely Feedback to Interactivity in Online Education. In I. Nääs, O. Vendrametto, J. Mendes Reis, R. F. Gonçalves, M. T. Silva, G. Von Cieminski, & D. Kiritsis (Eds.), *Advances in Production Management Systems. Initiatives for a Sustainable World* (Vol. 488, pp. 307–314). Springer International Publishing. [https://doi.org/10.1007/978-3-319-51133-7\\_37](https://doi.org/10.1007/978-3-319-51133-7_37)
- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, 27(3), 1043–1045. <https://doi.org/10.1016/j.chb.2010.07.014>
- Chan, J. W. W., & Pow, J. W. C. (2020). The role of social annotation in facilitating collaborative inquiry-based learning. *Computers & Education*, 147, 103787. <https://doi.org/10.1016/j.compedu.2019.103787>
- Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work - CSCW '92*, 107–114. <https://doi.org/10.1145/143457.143468>
- Gao, F. (2013). A case study of using a social annotation tool to support collaboratively learning. *The Internet and Higher Education*, 17, 76–83. <https://doi.org/10.1016/j.iheduc.2012.11.002>
- Gerry, S., Koschmann, T., & Suthers, D. (2006). Computer-supported Collaborative Learning: An Historical Perspective. In *Cambridge Handbook of the Learning Sciences* (Vol. 2006, p. 409).
- Hakkarainen, K., Paavola, S., Kangas, K., & Seitamaa-Hakkarainen, P. (2013). Sociocultural Perspectives on Collaborative Learning. In *The International Handbook of Collaborative Learning*. Routledge. <https://doi.org/10.4324/9780203837290.ch3>
- Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2), 174–196. <https://doi.org/10.1145/353485.353487>
- Jan, J.-C., Chen, C.-M., & Huang, P.-H. (2016). Enhancement of digital reading performance by using a novel web-based collaborative reading annotation system with two quality annotation filtering mechanisms. *International Journal of Human-Computer Studies*, 86, 81–93. <https://doi.org/10.1016/j.ijhcs.2015.09.006>

- Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, 48(1), 25–39. <https://doi.org/10.1080/00461520.2012.748006>
- Kirschner, P. a., Kreijns, K., Phielix, C., & Franssen, J. (2015). Awareness of cognitive and social behaviour in a CSCL environment. *Journal of Computer Assisted Learning*, 31(1), 59–77. <https://doi.org/10.1111/jcal.12084>
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19(3), 335–353. [https://doi.org/10.1016/S0747-5632\(02\)00057-2](https://doi.org/10.1016/S0747-5632(02)00057-2)
- Krouska, A., Troussas, C., & Virvou, M. (2018). Social Annotation Tools in Digital Learning: A Literature Review. *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–4. <https://doi.org/10.1109/IISA.2018.8633609>
- Li, S. c., Pow, J. w. c., & Cheung, W. c. (2015). A delineation of the cognitive processes manifested in a social annotation environment. *Journal of Computer Assisted Learning*, 31(1), 1–13. <https://doi.org/10.1111/jcal.12073>
- Martinez-Maldonado, R., Goodyear, P., Kay, J., Thompson, K., & Carvalho, L. (2016). An Actionable Approach to Understand Group Experience in Complex, Multi-surface Spaces. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2062–2074. <https://doi.org/10.1145/2858036.2858213>
- McMillan, S. J., & Hwang, J.-S. (2002). Measures of Perceived Interactivity: An Exploration of the Role of Direction of Communication, User Control, and Time in Shaping Perceptions of Interactivity. *Journal of Advertising*, 31(3), 29–42. <https://doi.org/10.1080/00913367.2002.10673674>
- Morrison-Smith, S., & Ruiz, J. (2020). Challenges and barriers in virtual teams: A literature review. *SN Applied Sciences*, 2(6), 1096. <https://doi.org/10.1007/s42452-020-2801-5>
- Muller, M., Shami, N. S., Millen, D. R., & Feinberg, J. (2010). We are all lurkers: Consuming behaviors among authors and readers in an enterprise file-sharing service. *Proceedings of the 16th ACM International Conference on Supporting Group Work - GROUP '10*, 201. <https://doi.org/10.1145/1880071.1880106>
- Nonnecke, B., Andrews, D., Preece, J., & Voutour, R. (2004). OnlineLurkersTellWhy. *New York*.
- Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education*, 15(1), 39–49. <https://doi.org/10.1016/j.iheduc.2011.09.002>
- Passig, D., & Maidel-Kravetsky, J. (2016). The impact of collaborative online reading on summarizing skills. *Education and Information Technologies*, 21(3), 531–543. <https://doi.org/10.1007/s10639-014-9337-5>
- Planey, J., Rajarathinam, R. J., Mercier, E., & Lindgren, R. (2023). Gesture-mediated collaboration with augmented reality headsets in a problem-based astronomy task. *International Journal of Computer-Supported Collaborative Learning*. <https://doi.org/10.1007/s11412-023-09398-w>
- Rafaeli, S., Ravid, G., & Soroka, V. (2004). De-lurking in virtual communities: A social communication network approach to measuring the effects of social and cultural capital. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of The*, 10 pp. <https://doi.org/10.1109/HICSS.2004.1265478>
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257. <https://doi.org/10.1007/s11412-009-9070-z>
- Sun, N., Rau, P. P.-L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, 38, 110–117. <https://doi.org/10.1016/j.chb.2014.05.022>
- Wise, A. F., Hausknecht, S. N., & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning*, 9(2), 185–209. <https://doi.org/10.1007/s11412-014-9192-9>
- Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, 41(2), 323–343. <https://doi.org/10.1007/s11251-012-9230-9>
- Yang, S. J. H., Zhang, J., Su, A. Y. S., & Tsai, J. J. P. (2011). A collaborative multimedia annotation tool for enhancing knowledge sharing in CSCL. *Interactive Learning Environments*, 19(1), 45–62. <https://doi.org/10.1080/10494820.2011.528881>
- Zhou, Y., & Kang, J. (2022). Characterizing joint attention dynamics during collaborative problem-solving in an immersive astronomy simulation. *Proceedings of the 15th International Conference on Educational Data Mining*, 406–413. <https://doi.org/10.5281/zenodo.6852988>

# Examining Coordinated Computer-Based Fixed and Adaptive Scaffolds in Collaborative Problem-Solving Game Environments

Daeun Hong, Chen Feng, Xiaotian Zou, Cindy E. Hmelo-Silver  
dh37@iu.edu, carrfeng@iu.edu, xz107@iu.edu, chmelosi@indiana.edu  
Indiana University

Krista Glazewski, Tianshu Wang, Bradford Mott, James Lester  
kdglazew@ncsu.edu, twang43@ncsu.edu, bwmott@ncsu.edu, lester@ncsu.edu  
North Carolina State University

**Abstract:** The study aims to examine the effectiveness of the coordinated computer-based scaffolds mediating students' collaborative problem-solving (CPS) practices and science learning. 24 students in six groups participated in CPS activities related to middle school life science, conducted in a scaffolded collaborative game. We analyzed their CPS practices during 31 collaborative periods across three quests using latent profile analysis, chi-square tests, and interaction analysis. We found certain clusters associated with higher-performing groups and patterns of their CPS engagement. Furthermore, we found that these CPS engagement patterns linked to higher performance were synergistically mediated by the coordination of fixed and adaptive scaffolds. The study also discusses implications and outlines future research directions.

## Introduction

Collaborative game-based inquiry environments are contexts for Computer-Supported Collaborative Learning (CSCL) and enable students to learn as they collaboratively engage with complex problems (Jeong et al., 2019). However, in such environments, students can face multiple challenges such as navigating a complicated problem-solving process (Savery, 2015). Successful learning in such circumstances largely hinges on skills in collaborative problem solving (CPS). Individuals with well-developed CPS are able to solve a shared problem by effectively sharing and negotiating ideas and experiences, jointly coordinating behaviors, and learning activities, and applying social strategies to sustain positive communication (Liu et al., 2016). As such, success at addressing authentic ill-structured problems in CSCL settings could be achieved if the students are provided with appropriate scaffolding to extend and enhance such capabilities as they engage with the target problems (Belland, 2017).

In CSCL environments, beyond support from teachers and peers, diverse forms and types of computer-based scaffolds have been embedded (Puntambekar, 2022). Two forms of computer-based scaffolds, fixed and adaptive scaffolds, have varying functions and play different roles in the learning process (Puntambekar, 2022). Computer-based adaptive scaffolds are supports that are offered contingent on students' previous learning performance or actions. In contrast, fixed scaffolds provide support that helps learners deal with problem complexity regardless of learner actions (Reiser, 2004). Previous studies suggest that it is necessary to seamlessly integrate the two different scaffolds in a scaffolding system to successfully design CSCL environments (Puntambekar, 2022). Such well-coordinated scaffolding is effective in synergistically addressing student's learning needs (Tabak, 2004). Despite some research studies examining coordinated scaffolds (e.g., Martin et al., 2019; Saleh et al., 2020), there is limited research that addresses how to coordinate fixed scaffolds with adaptive scaffolds in CSCL. Furthermore, little is known about how effectively coordinated computer-based scaffolds may or may not be taken up by students, thereby contributing to students' practices in CPS and disciplinary learning in science. As such, this study aims to understand how computer-based fixed and adaptive scaffolding may or may not effectively mediate student disciplinary learning in life science and CPS within a scaffolded game-based science learning environment. Specifically, the current study seeks to answer the following research questions within the context of collaborative problem-solving activities:

- RQ1: What CPS patterns do students who received both computer-based fixed and adaptive scaffolds show, compared to those who received fixed-only scaffolds?
- RQ2: What CPS patterns are related to higher and lower performance in science inquiry?
- RQ3: How do the coordinated scaffolds mediate learners' CPS practices in ways that may be associated with learning performance, if at all?

Throughout the paper, scaffolding is defined as the process of providing support, whereas scaffolds are the tools and artifacts that actually provide the support (Puntambekar, 2022). Adaptive scaffolds refer to a form of support that is only delivered to certain groups or students based on their prior gameplay actions or responses

to student actions. Fixed scaffolds refer to a form of support that is provided to all students regardless of their previous performances once they complete an activity or reach a certain point.

## Literature review: Scaffolding collaborative problem solving (CPS)

Collaborative problem solving (CPS) is described as a complex process in which two or more individuals attempt to solve a problem by sharing and jointly consolidating their knowledge, skills, and endeavors to reach a desired solution (Grasser et al., 2018; OECD, 2017). CPS encompasses both cognitive and social dimensions, which are closely intertwined with each other (Sun et al., 2022). Practices in CPS can be observed in discursive settings such as face-to-face conversation and text-mediated communications. Specifically, Liu et al. (2015) proposed a CPS framework focusing on the practices from a discursive perspective that consists of four major categories: (a) sharing ideas, (b) negotiating ideas, (c) regulating problem solving, and (d) maintaining communication. Under each category, they specified observable discursive practices. However, in problem-based learning (PBL) settings, CPS demands students engage in these discursive practices to understand a problem but also regulate their actions to solve it. Therefore, scaffolding is necessary to support students to overcome the challenges, productively engage in CPS and disciplinary learning, and ultimately, to be successful in PBL.

Computer-based scaffolding refers to support by a computer system that promotes student engagement with content and performance that are beyond their independent capacities (Belland et al., 2017). In CSCL classrooms, two types of computer-based scaffolds, fixed and adaptive scaffolds, can be offered to support students' CPS and content learning. The pedagogical framework itself (e.g., PBL cycle and inquiry process) also can be embedded as a fixed scaffold in CSCL environments (Martin et al., 2019). Both types of scaffolds benefit students by posing challenges to engage students with problems, decomposing a complex task to make it more manageable, and maintaining a forward direction (Reiser & Tabak, 2014). Indeed, computer-based scaffolding demonstrates substantial positive effects, as compared to comparison conditions (Belland et al., 2017).

Multiple forms of support are necessary to afford opportunities to engage students in learning activities and perform successfully (Reiser & Tabak, 2014). Tabak (2004) proposed the notion of *Synergistic scaffolds* as one of the patterns of how different types of scaffolds can support learners. The pattern involves multiple supports targeting the same needs such that the various supports complement the effects of other scaffolds that might not be sufficient alone. However, previous studies raised questions concerning the effectiveness of combining multiple forms of scaffolds, as varying features, functions, and goals of scaffolds can be dissonant and even conflicting in supporting students' learning (Zydney, 2010). Furthermore, one of the forms of scaffolds can be ignored, and their effects become diminished (Tchounikine, 2017). Although some previous studies examined how to distribute scaffolds across people and computer systems (Martin et al., 2019; Puntambekar, 2022), how (or if) students take up and use computer-based adaptive and fixed scaffolds as intended in CSCL settings remains unclear. Furthermore, if the coordinated scaffolds have been taken up, few have examined the synergistic effects of the coordinated scaffolding on students' CPS and how it contributes to disciplinary learning in science. As such, this study aims to examine how coordinated computer-based scaffolds mediate students' practices in CPS and science learning in collaborative game-based science learning environments compared to fixed-only scaffolding.

## Methods and analysis

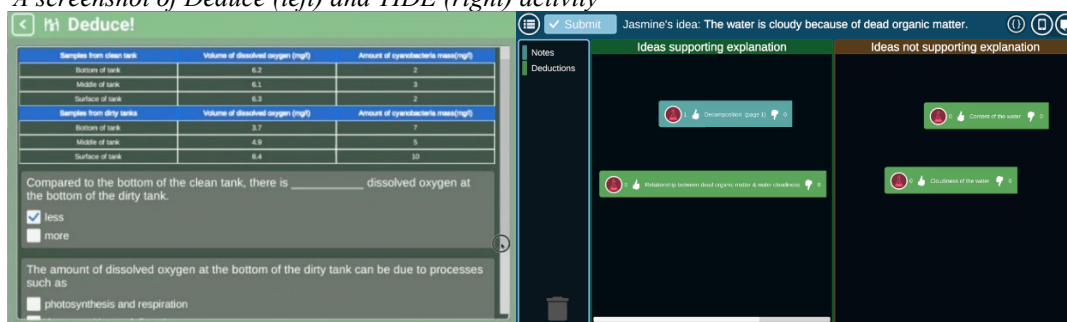
### Scaffolded game environment

CRYSTAL ISLAND: ECOJOURNEYS is a scaffolded collaborative game-based learning environment designed to support disciplinary learning in middle school life science concepts and CPS practices. Three to four students within a group collaboratively investigate why the tilapia fish at a farm are sick. The game contains a tutorial followed by three quests. In each quest, students individually collect data related to the fish ecosystems by collecting notes, talking to non-player characters (NPCs), and measuring water quality. Subsequently, students engage in CPS activities called Deduce and TIDE (*Talk, Investigate, Deduce, and Explain*), using what they have learned previously. In Deduce activities (see Figure 1 left), students as a group are required to answer aquatic ecosystem multiple-choice questions. In TIDE activities, they need to determine if the information collected (i.e., a note) supports a given claim or not, using the "TIDE board" (see Figure 1 right), which is a real-time collaborative whiteboard. They share their arguments on the claim by placing relevant notes as evidence into one of the columns on the board. During the activities, they discuss through in-game chat and in-person conversation. In this study, we only focus on the Deduce and TIDE activities, where students have to engage in CPS practices.

Fixed and adaptive scaffolds are embedded to support students' disciplinary learning in life science and CPS. Fixed scaffolds take the form of the Deduce App and the TIDE board. After entering answers for each question in Deduce, students in a group receive an update on their agreement status as a group for the answer choice. The TIDE board represents a fixed scaffold that models the collaborative inquiry processes by making

students' thinking visible and structuring their inquiry activity (Reiser & Tabak, 2014). They can justify their ideas with evidence by placing relevant data (i.e., a note) they have collected on one of the columns on the board. Students then read each note and express their opinions on its placement through voting, which facilitates them in negotiating ideas and reaching an agreement. During the discussion, they can use sentence starters based on the *Accountable Talk* framework, modeling quality collaborative discourse that helps students stay accountable for knowledge (e.g., "As I was playing, I learned that ..."), to reasoning (e.g., "My reason is ..."), and community (e.g., "Does anyone agree with my idea that ...") (Resnick et al., 2018). Following each group's submission in Deduce and TIDE activities, the game system also provides feedback on the accuracy of the answer to the group.

**Figure 1**  
*A screenshot of Deduce (left) and TIDE (right) activity*



Adaptive scaffolds are also included during the Deduce and TIDE activities. For example, in response to a group's prior chat contribution, an NPC delivers a message to the group at the start of either Deduce or TIDE (e.g., "Don't forget to use the group chat to talk out ideas together! Consider using sentence starters if you are stuck."). The group also receives a scaffold contingent on their level of satisfaction with prior collaboration and their identified CPS aspect that needs improvement, determined through a survey at the end of each quest. During the activity, the system also displays a prompt in the group chat that is adaptive to the group's condition (e.g., "[Player], can you share your thoughts about this question?" and "Have we heard from during Deduce? If not, let's make sure we hear from them.").

## Participants

Participants were 24 middle school students in six groups across three science classes. Before the gameplay, groups were randomly assigned to one of the two groups: fixed-only ( $n = 3$ , Fixed 1, 2, and 3) or fixed and adaptive ( $n = 3$ , Adaptive 1, 2, and 3). Each focal group consisted of four students, all of whom used individual laptops but played the game together. The unit of analysis is a collaborative activity period across three quests per group except for the tutorial. Excluding missing data, we used the 31 periods of Deduce and TIDE activities across the six focal groups.

## Data sources and analysis

For RQ1, we analyzed each group's in-game chat data and video data of in-person conversations to examine how the groups engaged in the collaborative activities (i.e., Deduce and TIDE) in each quest. After transcribing the in-person discussions, we collected a total of 3,455 utterances including the data from both videos (58%) and in-game chat (42% including 8% of spamming) across the 31 periods of the collaborative activities. In this analysis, each sentence is considered as one utterance. To classify higher and lower performing groups related to RQ2, we used individuals' final solutions, regarding the issues in the local aquatic ecosystem and the reasons for the tilapia fish becoming sick. For RQ3, we utilized the video data capturing focal group interactions during the 31 periods of the collaborative inquiry activities across the three quests. The length of the entire video corpus is approximately 372 minutes.

To analyze RQ1, two coders adapted the coding scheme by Liu et al. (2016) for our learning environments. Using the adjusted CPS coding scheme (see Table 1), they independently coded 691 utterances (about 20% of all utterances). This process yielded an inter-rater reliability of Cohen's  $\kappa = 0.81$ . Subsequently, one coder coded the remaining utterances. Following that, we summed the number of each code for each activity per group (i.e., the unit of analysis). As a result, excluding utterances coded as 'other', which were unrelated to CPS performance, we obtained the aggregate number of each CPS code for each activity per group (e.g., Group 3 in Quest 1 TIDE). We then converted the values of each CPS code to the z-standardized mean scale values for comparison and interpretation of indicator values and performed latent profile analysis (Spurk et al., 2020) to

cluster patterns of CPS. The number of profiles (i.e., clusters) was determined based on values of the Bayesian information criterion (BIC, Spurk et al., 2020), and the CPS clusters were extracted. Despite being aware of the small sample size, we performed a chi-square to determine if there was a significant difference in CPS patterns between the conditions.

**Table 1**  
*The Adjusted CPS Coding Scheme*

Code	CPS category	Description
Sharing	Sharing information and ideas	Share or ask to share ideas, knowledge, resources, and information regarding ongoing tasks or game-related functions
Negotiating	Negotiating ideas	(Asks to) Express one's opinion on ideas suggested by others or provide additional information or details to clarify, correct, or elaborate one's or teammate's ideas.
Regulating	Regulating problem solving	Identify goals and problems to work on, monitor and evaluate team's performance, and manage frustration and challenges during the activity
Maintaining	Maintaining positive communication	Respectfully communicate with others, maintain lively conversations, or encourage and support each other during the activity
Out of domain	Out of domain	Talk or work on something unrelated to the task
Other	Other	(Fixing) Typos, inaudible and incomprehensible utterances, self-talk, mumbling, words including gibberish words such as ah, uh, bdse, etc.

To identify CPS patterns associated with higher performance for RQ2, we classified the focal groups as either higher or lower performing based on their final solutions. Groups that explicitly connected the fish kill to aquatic ecosystem components and their relationships (e.g., “High levels of dead organic matter, which provides nutrients to cyanobacteria, and cyanobacteria reduces the availability of dissolved oxygen for fish, causing the fish to become sick.”) was assigned to higher-performing groups. Groups that provided a superficial description (e.g., “I think the cyanobacteria are impacting the tilapia”) or explanation that was not covered in the game (e.g., “cyanobacteria enter the fish gills and triggers an immune response making them sick”, etc.) were identified as lower-performing groups. The two coders reached a 100 percent agreement. Moreover, we probed the extent of each cluster's contribution to either higher or lower performance, computing Pearson residuals. Based on the results, we qualitatively examined student CPS practices in each category associated with either higher or lower performance. For RQ3, through interaction analysis (Hall & Stevens, 2015), we investigated how or if the coordinated scaffolds contributed to CPS practices associated with higher performance by comparing CPS practices of higher- and lower-performing groups in the adaptive condition. We selected and analyzed video segments that showed the contrasting patterns relevant to the scaffolds to identify possible scenarios of students taking up the scaffolds and engaging in productive CPS practices.

## Results

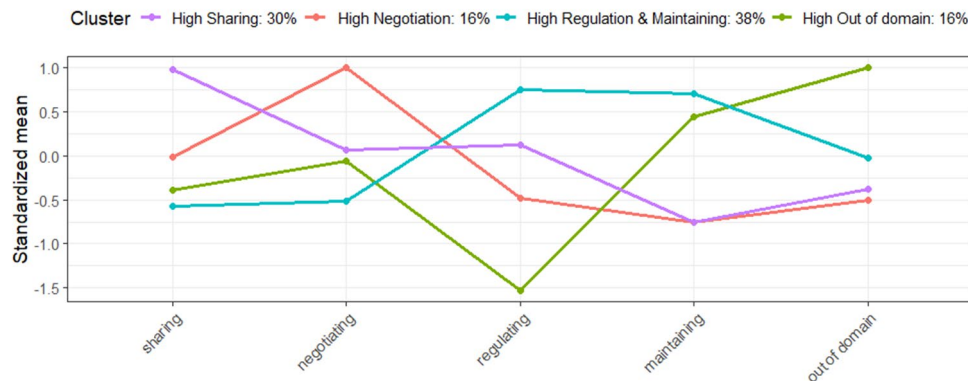
### RQ1: Collaborative problem-solving clusters

We identified four clusters (i.e., patterns) of student CPS patterns, selecting those with the lowest BIC values through the latent profile analysis (Figure 2). The first cluster was of medium size ( $n = 9$ , 30%) and showed a high level of sharing ideas and information, along with average levels of negotiation and regulation. This cluster can be characterized as the *high-sharing* (high shr) group. The second cluster was small ( $n = 5$ , 16%) and showed a high value in negotiating ideas and an average level of sharing. We labeled this cluster as the *high negotiating* (high ngt). The third cluster was of medium size ( $n = 12$ , 38%) and was characterized by above-average levels of regulating problem solving and maintaining positive communication. We named this the *high regulating and maintaining* (high rgl&mnt) group. The fourth cluster, *high out-of-domain* (high out), was small ( $n = 5$ , 16%) and showed high levels of maintaining positive communication and out-of-domain tasks. Throughout the 31 periods of the collaborative activities, groups across the conditions most frequently exhibited the high regulating and maintaining pattern, while high negotiating and high out-of-domain patterns were the least prevalent.

Furthermore, we compared the frequency distribution of each cluster between the two conditions. In the fixed and adaptive condition, the most frequent CPS pattern was the high negotiating pattern (80%), followed by

high regulating (50%), high sharing (44%), and high out-of-domain (40%). In the fixed-only condition, the most frequent CPS pattern was high out-of-domain (60%), followed by high sharing (54%), high regulating and maintaining (50%), and high negotiating (20%). Despite the trend towards variability in the frequency distribution, the results of the chi-square test indicated that there is no significant difference in the student CPS patterns between the conditions ( $\chi^2(3, n = 31) = 2.081, p > .05$ ). In other words, the groups who received both adaptive and fixed scaffolds did not exhibit distinctive CPS patterns compared to the other groups during the collaborative inquiry activities.

**Figure 2**  
*Four clusters of CPS patterns*



## RQ2: Association between CPS patterns and learning performance

To identify CPS practices associated with higher and lower performance in science inquiry, we classified the six groups based on the final solution. Adaptive 1, 2, and Fixed 2 were classified as higher-performing groups, and Adaptive 3, Fixed 1, and 3 were categorized as lower-performing groups. We examined the two groups' distribution of the frequency of each cluster (see Table 2). During the activities across three quests, 80 percent of the high-negotiating patterns were observed in the higher-performing groups, followed by 67 percent of high-sharing and 40 percent of high out-of-domain patterns. In the lower-performing groups, 67 percent of high-regulating and maintaining patterns were shown, followed by 60 percent of high out-of-domain and 33 percent of high-sharing patterns. In addition, we conducted a chi-square test to see if there was a significant difference in CPS patterns between the higher- and lower-performing groups, but the results showed no significant difference,  $\chi^2(3, n = 31) = 4.305, p > .05$ . We further computed standardized residuals to assess the degree of contribution of each cluster to higher and lower performance. The findings indicated that high sharing (.628), and high negotiating (.884) showed moderate and high associations with higher performance, whereas high regulating and maintaining (.910) and high out-of-domain (.373), showed large and small contributions to lower performance, respectively.

**Table 2**

*Each group's frequency distribution of each cluster and performance in science inquiry*

	Adaptive 1	Adaptive 2	Adaptive 3	Fixed 1	Fixed 2	Fixed 3
Performance	Higher <sup>1</sup>	Higher	Lower <sup>2</sup>	Higher	Lower	Lower
High sharing	0	4	0	1	2	0
High ngt	2	1	1	0	1	0
High rgl & mnt	2	1	2	3	0	3
High out-of-domain	1	0	1	2	1	2
Missing data	0	0	2	0	2	1

<sup>1</sup>Higher indicates that the group belongs to the higher-performing group.

<sup>2</sup>Lower indicates that the group belongs to the lower-performing group.

To better understand the relationship between the clusters and performance in science inquiry, we closely examined students' actual utterances. We observed nuances between the groups in their practices in each aspect of CPS, which might be related to the group's level of performance (see Table 3). In terms of practices in sharing ideas and information, both groups similarly shared ideas on potential answers. However, higher-performing groups tended to refer to more learning resources, which was subsequently connected to sharing their answers and negotiating their ideas. The other groups merely expressed ideas or typed answers for another, which did not lead



to further discussion (Fixed 3, see Table 3). In other words, the groups that engaged in limited aspects of CPS practices were still labeled as a lower performance group, even though they showed the high-sharing pattern, which had a positive correlation with higher performance. When negotiating ideas, higher-performing groups argued with their rationales and elaborated their arguments in detail. However, the lower performing groups were more likely to simply express whether they (dis)agree on one’s answer. Higher performing groups’ regulation tended to be tied to sharing or negotiating ideas, while the other groups mostly used the strategy to move to the next step. For instance, a student’s utterance, “Hanna, we would love it if you joined in” (Adaptive 1, see Table 3), led Hanna to express her (dis)agreement. Additionally, one of the utterances from Adaptive 2, “You have to say why”, was followed by an argument shared by a student and led the student to reflect on the argument and elaborate on its reasoning. Thus, even the high regulating and maintaining cluster, which was correlated to lower performance, contributed to higher performance, if students engage in multiple CPS strategies within scientific discussion. Lastly, lower-performing groups’ maintaining positive communication and out-of-domain tasks distracted their science discussion, while the other groups tended to quickly get back on task. Taken together, it can be inferred CPS practices organically connected to others, especially negotiating ideas, might have contributed to higher performance in collaborative science inquiry.

**Table 3**

*Examples of the Discourse from Higher and Lower Performance Group for Each CPS Category*

CPS category	Higher performance	Lower performance
Sharing ideas and information	“i think the answer is greenish brown and cloudy” (Adaptive1) “Hey, Yeah I looked at the notes and it said nothing about them swimming at the top because the water is cloudy.” (Fixed 2)	“What should I write?” (Fixed 3) “I typed the answer for you” (Fixed 3)
Negotiating ideas	“So, there’s Yes, there’s less dissolved oxygen because the bottom of the clean tank is 6.2 and the bottom of the dirty tank is 3.7. A 3.7 is less.” (Fixed 2) “no there’s less on the bottom than their swimming in the top because there’s more right?” (Adaptive 2)	“No, don’t try less. It’s more.” (Fixed 1) “I think it’s good.” (Fixed 1)
Regulating problem solving	“Hanna, we would love it if you joined in” (Adaptive 1) “You have to say why.” (Adaptive 2)	“Hit agree agree” (Adaptive 3) “What should we do?” (Fixed 1)
Maintaining positive communication	“Good job, Ivanna.” (Adaptive 2) “you can politely disagree” (Adaptive 1)	“ola” (Fixed3 C2G4) “Oh Agree. Yay.” (Fixed 3)

### RQ3: Coordinated scaffolds mediating student CPS

Utilizing interaction analysis, we examined how or if the coordinated scaffolds productively mediated the students’ CPS associated with higher performance as delineated in RQ2. Based on the results above, we differentiated between higher-performing (Adaptive 1 and 2) and lower-performing (Adaptive 3) groups in the fixed and adaptive condition, then compared their engagement in CPS during the Deduce and TIDE activities. Several segments within the videos revealed that the students in Adaptive 1 and 2 took up the computer-based scaffolds, which in turn led to an expanded science discourse during their collaborative endeavors. However, Adaptive 3 did not use scaffolds as intended. The following presents one example of each case where the students used or did not use the adaptive scaffolds as intended.

During the first set of questions in Quest 3 Deduce, Adaptive Group 2 encountered the adaptive scaffold (e.g., “Is there anyone we haven’t heard from during Deduce? If so, let’s make sure we hear from them.”), which actively encouraged participation from all group members. Subsequently, two students, Madelyn and Eliza, started to type potential answers for the questions in the chat (e.g., “The first one is A” [CPS code: sharing]), fostering initiation into negotiations concerning the correct answers. Additionally, as promoted by the adaptive scaffold (e.g., “Don’t forget to use the group chat to talk out ideas together! Consider using sentence starters if you are stuck.”), Madelyn leveraged one of the conversation starters, a fixed scaffold, by sending “Does anyone agree with my idea that ... it increases?” (i.e., negotiating). Their actions elicited other’s reactions (e.g., “Yeah, the first one is A, and the second one is B.” [negotiating] and “That’s what I said.” [negotiating]), and the group resumed their discussion. As such, the adaptive scaffold, which encouraged sharing ideas and the use of the fixed scaffold, facilitated a clear consensus on the idea and progressed to the next set of questions. This highlights the synergistic

effect of the coordinated scaffolds in promoting their CPS practices. Moreover, during the second set of questions, another adaptive scaffold (i.e., “Ivanna (one of the group members), can you share your thoughts about this question?”) encouraged Madelyn to recognize a disparity in group participation and reflect on her group’s conversation, “Oh we have to hear from everybody in the chat” (i.e., regulating). She pointed out the absence of Ivanna’s participation and invited Ivanna (i.e., “Ivanna, you have to chat something. Because we haven’t heard from you in a while” [regulating]). Initially, the group’s ideas converged to answer ‘a’ and ‘d’, but Ivanna expressed disagreement on the group’s ideas (i.e., “I thought it’s produce and photosynthesis” [negotiation]). This led to an extended negotiation to address the disagreement, during which they delved into a further discussion of the relationship between photosynthesis and dissolved oxygen to support her idea. These differing perspectives necessitated regulation and evaluation of the final decision, encompassing various aspects of CPS practices. As a result, along with the fixed scaffolds, the adaptive scaffolds played a pivotal role in enhancing the depth of their negotiation and engagement in diverse aspects of CPS practices, which are associated with higher performance.

In contrast, Adaptive 3 did not consistently adopt and employ the given scaffolds as intended, resulting in distinct patterns of their CPS from the higher-performing adaptive groups. As an illustrative example, during the Quest 1 TIDE, Sam was just looking at his screen and not participating in tasks at hand (e.g., “who’s doing it?”, “no it’s not me it’s not me” [regulating]), but only two students in the group were moving notes on the TIDE board. Although they noticed that an adaptive prompt was presented, they took no action in response to the adaptive scaffold. After all the notes were placed, they did not discuss if they agreed on each note’s placement but tried to proceed forward (e.g., “Pres, press submit” [regulating]). In this situation, students even unexpectedly used the sentence starters by leaving a blank as it is (e.g., “As I was playing, I learned that ...”) and putting irrelevant content (e.g., “Does anyone agree with my idea that Arron needs to put notes” and “Does anyone agree with my idea that whale noises[noises] help me sleep”). Although the adaptive scaffold was triggered, it did not affect their CPS engagement and redirect their focus toward the task at hand. As a result, their CPS practices were skewed to certain categories of CPS, particularly regulating and maintaining, with limited connection to other facets of CPS performance. Indeed, across activities, utterances coded as regulating are about what they are working on (e.g., “what should we do?” [regulating]), suggesting selecting another option without further discussion (e.g., “do ‘a’ do the first one” [regulating]), and fostering group members to proceed by hitting a ‘submit’ button (e.g., “Let’s go. Caitlin, Caitlin. Hit submit. Hit submit. Okay” [regulating]). Such participation limited their utilization of other available scaffolds and interaction with the learning environment and peers, potentially contributing to their learning outcomes. As such, compared to Adaptive 1 and 2, Adaptive 3, which underutilized the scaffolds, demonstrated less productive CPS practices, affording fewer opportunities for meaningful interaction with the learning materials and group members.

## Discussion

This study examined the effectiveness of coordinated computer-based scaffolds mediating students’ CPS practices and science learning in a game-based CSCL context, focusing on CPS clusters linked to student performance. The study found that although each of the four identified clusters shows a slight to moderate association with either higher or lower performance, focusing on one or two categories of CPS performance is associated with lower performance. In contrast, CPS practices characterized by interconnections with various CPS categories appear to be correlated with group performance levels, offering a potential explanation for the lack of significant difference in the frequency distribution of CPS clusters between high-performing and low-performing groups. Furthermore, the study presents scenarios wherein adaptive scaffolds collaboratively encouraged students in high-performing groups to actively engage in diverse facets of CPS practices alongside the fixed scaffolds. Considering collectively, our results suggest that the pivotal distinction in group performance levels may reside in the interconnection between different CPS performance types. Nevertheless, not all students consistently utilized the adaptive scaffolds, potentially contributing to the absence of significant differences in the frequency of each CPS cluster between the conditions. Providing group-level scaffolding presents challenges as the scaffolds may not align with various group dynamics, particularly those that hinder taking up and utilizing the scaffolds effectively. This implies that it may be imperative to implement multiple forms of computer-mediated scaffolding targeting the same needs, tailored to address distinct requirements across various contexts and audiences (Puntambekar, 2022). The findings also suggest involving human intervention in scaffolding and providing strategically distributed scaffolding across tools, peers, and teachers (Puntambekar, 2022) to deliver effective scaffolding (Tabak, 2004). Such approaches would lead to a more sophisticated and comprehensive scaffolding design.

Our findings contribute to the existing body of research on the coordination of different types of computer-based scaffolds and their interplay supporting students. In line with the previous studies, the study demonstrates that different forms of coordinated computer-based scaffolding can also produce synergistic effects on student learning outcomes. One of the ways to manifest the synergistic effects of the computer-based fixed and

adaptive scaffolds is to use adaptive scaffolds as a means of encouraging learners to employ fixed scaffolds, which subsequently leads them to further engage in productive CPS practices. This can offer insights for designing scaffolded computer-mediated learning environments that promote synergistic effects of coordinated scaffolding on student learning. Given the limited sample size and the aggregation of CPS performance categories, further study should incorporate a sequence of CPS practices to examine connectivity between different CPS categories and investigate relationships between student performance and CPS practices using a larger sample size. As a result, our study is an initial step in providing design implications for different forms of scaffolding.

## References

- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309-344.
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological science in the public interest, 19*(2), 59-92.
- Hall, R., & Stevens, R. (2015). Interaction analysis approaches to knowledge in use. In *Knowledge and interaction* (pp. 88-124). Routledge.
- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational research review, 28*, 100284.
- Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, J. D. (2016). A tough nut to crack: Measuring collaborative problem solving. In *Handbook of research on technology tools for real-world skill development* (pp. 344-359). IGI Global.
- Martin, N. D., Tissenbaum, C. D., Gnesdilow, D., & Puntambekar, S. (2019). Fading distributed scaffolds: The importance of complementarity between teacher and material scaffolds. *Instructional Science, 47*(1), 69-98.
- Organisation for Economic Co-operation and Development. (2013). PISA 2015: Draft collaborative problem solving framework. (Accessed 7 April 2015).
- Puntambekar, S. (2022). Distributed Scaffolding: Scaffolding Students in Classroom Environments. *Educational Psychology Review, 34*(1), 451-472.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304. [https://doi.org/10.1207/s15327809jls1303\\_2](https://doi.org/10.1207/s15327809jls1303_2)
- Reiser, B. J., & Tabak, I. (2014). Scaffolding. In *The Cambridge Handbook of the Learning Sciences, Second Edition* (pp. 44-62). Cambridge University Press.
- Resnick, L. B., Asterhan, C. S., & Clarke, S. N. (2018). Accountable talk: Instructional dialogue that builds the mind. *Geneva, Switzerland: The International Academy of Education (IAE) and the International Bureau of Education (IBE) of the United Nations Educational, Scientific and Cultural Organization (UNESCO)*.
- Saleh, A., Yuxin, C., Hmelo-Silver, C. E., Glazewski, K. D., Mott, B. W., & Lester, J. C. (2020). Coordinating scaffolds for collaborative inquiry in a game-based learning environment. *Journal of research in science teaching, 57*(9), 1490-1518.
- Savery, J. R. (2015). Overview of problem-based learning: Definitions and distinctions. In A. Walker, H. Leary, C. E. Hmelo-Silver, & P. A. Ertmer, P. A. (Eds.), *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows* (pp. 5–15). West Lafayette, IN: Purdue University Press.
- Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. *Journal of vocational behavior, 120*, 103445.
- Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education, 143*, 103672.
- Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *The journal of the Learning Sciences, 13*(3), 305-335.
- Tchounikine, P. (2017). Designing for appropriation: A theoretical account. *Human-Computer Interaction, 32*(4), 155-195.
- Zydney, J. M. (2010). The effect of multiple scaffolding tools on students' understanding, consideration of different perspectives, and misconceptions of a complex problem. *Computers & Education, 54*(2), 360–370.

## Acknowledgments

This study was supported by US National Science Foundation DRL-1561486, DRL-1561655, and DRL-2112635.

## What Are They Regulating? Research on Cognitive, Task and Emotional Regulation Patterns in CSCL

Shurui Gao, Weijia Chen, Zihan Guo, Fati Wu  
gsr97@foxmail.com, 202239010018@mail.bnu.edu.cn, 202221010197@mail.bnu.edu.cn, wft@bnu.edu.cn  
School of Educational Technology, Beijing Normal University

**Abstract:** In computer-supported collaborative learning (CSCL), the learners constantly regulate themselves to achieve task goals and make learning happen with the support of online tools and resources. This research examines the learners' regulation in a face-to-face discussion environment with access to the Internet for information. The research focuses on how learners carry out cognitive, task, and emotional regulation. The results indicate that CSCL learners mainly engage in cognitive regulation and can maintain regulation of task and emotional aspects throughout the discussion. At the individual level, learners are clustered into four types: Knowledgeable Learners, Curious Learners, Expressive Learners, and Passive Learners. At the group level, the high-performance groups emphasize in-depth discussions on perspectives, while the low-performance groups focus more on continuously expressing their views and rushing to complete tasks. The findings of this research serve to provide insights into learner grouping and the design of discussion activities with scaffolding in CSCL.

### Introduction

The COVID-19 pandemic has caused a massive shift in how learning activities are conducted. This move towards computer-supported collaborative learning (CSCL) has accelerated during the pandemic. Still, the question now is whether to return to face-to-face learning activities or whether hybridity should be embraced as the new norm. The advantages of CSCL for building shared meaning are clear, but the expansion also brings the challenges of increased complexity in diversity and learning environment design. This research aims to examine how learners regulate their participation during the discussion, including cognitive, emotional, and task-related aspects, offering insight into how to facilitate CSCL discussions to be inclusive and effective.

### Literature review

#### Collaborative learning occurs through social exchanges

The collaborative learning theory, rooted in social constructivism, provides valuable insights into the nature of learning as a social phenomenon (Vygotsky, 1978). This theory posits that learning is not an individual process but rather emerges from the interactions and collaboration among learners. It emphasizes the significance of reciprocal and supportive interactions in promoting the sharing of ideas, perspectives, and expertise among learners (Dillenbourg, 1999). In collaborative learning environments, learners actively engage with one another, participating in collective problem-solving, knowledge construction, and meaning-making. Thus, collaborative learning promotes the cultivation of critical thinking skills, the development of metacognitive abilities, and the enhancement of communication and collaboration skills (Gokhale, 1995; Laal, 2013).

Jean Lave and Etienne Wenger (1991) further argued that collaborative learning is a crucial aspect of social participation within a community of individuals who share a common interest or profession. Moreover, within the collaborative learning process, the role of the "knowledgeable other" assumes great significance (Vygotsky, 1978). This knowledgeable other can be a teacher, facilitator, mentor, or a more experienced peer who possesses a deeper understanding of the subject matter. They provide scaffolding by offering guidance, posing probing questions, and providing feedback, enabling learners to navigate challenging tasks, acquire new knowledge, and develop self-regulation skills (Cohen, 1994; Hadwin et al., 2017; Johnson & Johnson, 1974).

To facilitate collaborative learning, various models have been developed, such as Students Teams Achievement Divisions (STAD), which approach learning from a cooperative perspective (Slavin, 1991). However, collaborative learning goes a step further by placing greater emphasis on the significance of "conversational learning" through discourse and argumentation (Bruffee, 1984). Studies have demonstrated that engaging in argumentative discussions can enhance learners' critical thinking and reasoning abilities (Wason & Johnson-Laird, 1972). Additionally, Bruffee (1999) has put forth a conceptualization of knowledge construction that involves the processes of idea generation, idea organization, and intellectual convergence. For a group to be

productive in a collaborative learning process, group dynamics, individual accountability, and peer support, amongst other factors, were identified as significant contributing factors towards an effective team (Cohen & Lotan, 2014; Laal et al., 2013)

## Regulation in CSCL

CSCL introduces specific complexities that demand the regulation of the learning process to ensure the effective accomplishment of tasks (Koivuniemi et al., 2018; Malmberg et al., 2015). Regulated learning encompasses the deliberate negotiation of task objectives, the strategic selection and implementation of techniques to optimize task performance, the continuous monitoring of progress, and the flexibility to make necessary adaptations (Järvelä & Hadwin, 2013; Winne et al., 2013). Within this framework, three categories of regulation can be identified: self-regulation, co-regulation, and socially shared regulation. In the context of collaborative learning, each group member assumes responsibility for regulating their learning journey (self-regulation), contributing to the regulation of other group members' learning (co-regulation), and participating in the collective regulation of all group members' learning processes (socially shared regulation).

While prior research has extensively examined self-regulation within the domain of self-regulated learning (Winne, 2004; Zimmerman & Moylan, 2009), recent academic endeavors have increasingly directed their focus toward co-regulation and socially shared regulation. Current empirical studies suggest that all three forms of regulation remain active and influential throughout the collaborative learning process, guiding learners to navigate tasks, construct knowledge, and seek emotional alignment (Järvenoja et al., 2020; Zheng, 2016c). Furthermore, learners engage in social regulation to effectively address a wide array of challenges, encompassing cognitive hurdles, motivational barriers, task and time management, and technical difficulties (Malmberg et al., 2015). In the domain of CSCL in Chinese schools, research has been conducted on co-regulation and socially shared regulation using behavior traits and sequence analysis (Tian & Wu, 2022; Zheng, 2016a, 2016b). These behaviors are primarily manifested through conversations, specifically in exchanging ideas and conceptual exchanges.

However, few research has investigated the transitions between the regulation of knowledge building, task management, and emotional alignment within a CSCL activity.

## Research question

One notable finding from Zheng's research is that low-performing groups tend to have more off-topic discussions when it comes to socially shared regulation for task management (Zheng, 2016a, 2016b). This finding has inspired us: what may seem "off-topic" in terms of managing the learning task could serve a purpose in regulating the group's emotional tension. In the CSCL environment, learning happens in a continuous manner as ideas are generated, deliberated upon, accepted, or rejected. This learning process is influenced by the learners' emotional state and their perception of task progress.

In a typical online collaborative learning environment, learners' discussion often takes place through online channels such as forums (asynchronous) or chat boxes (synchronous). Either way, learners are usually engaged through written or typed exchanges rather than oral conversations. In the typing process, the learner may intentionally or unintentionally sieve out sentences that they deem "off-topic" to the learning task. Thus, this research proposes to examine the more real-time and authentic scenario of face-to-face discussion exchanges, using content analysis for all conversational exchanges.

While the current research mostly dissects regulation based on the "who" is being regulated, this research aims to categorize regulation based on "what" is being regulated. In this research, we define regulation behaviors as "*cognitive regulation*", "*task regulation*" and "*emotional regulation*", addressing research gap we identified in literature review. By examining patterns in the traits and sequence of regulating activities at both the individual and group levels, this study seeks to shed light on the multifaceted nature of regulation within CSCL.

Based on the above, this research proposes three research questions:

1. How are the three types of regulation behaviors distributed throughout the discussion process?
2. At the individual level, how many potential clusters can be formed based on the distribution of the three types of regulation behaviors?
3. How do behavior transition sequences differ between the high- and low-performing groups?

## Methodology and design

### Participants

Volunteers were recruited from a university, consisting of both undergraduate and postgraduate students, to perform a group discussion on the “Sponge Campus Design” project. Each group comprised of 3 to 4 participants. In total, 42 groups (127 participants) completed the experiment. To ensure a similar level of prior knowledge of the subject matter, volunteers from the Geography Faculty were excluded.

## Experiment setup

A dedicated discussion room on campus was used for the face-to-face CSCL task. Upon arrival at the experiment venue, the participants were briefed on the experiment proceeding. They received the list of learning tasks and a set of reading materials. The participants were also given a tablet for discussion, writing, and drawing. Besides the reading material, the participants were allowed to use their personal devices for research. The participants gave consent for the audio recording of their discussion process.

## Experiment design and data analysis

The participants were to complete the 1.5-hour CSCL task and produce a set of “Sponge Campus Design” plans. At the end of the discussion, one of the group members represented the team to present their design. Two subject matter experts from the Geography Faculty were invited to independently grade the group designs. The average of the two scores was used to measure the group performance. The discussion exchanges were transcribed into text for content analysis based on semantic units for mining any patterns of regulation behaviors exhibited in the CSCL process.

To address Research Question One, descriptive statistics were used to calculate the mean number of semantic units in each category and subcategory, and the Mann-Whitney U Test for analyzing the difference in distribution between the high- and low-performance groups.

To address Research Question Two, the K-means algorithm was used to partition the learners into clusters based on their regulation behavior during the CSCL, based on which learner types can be analyzed.

To address Research Question Three, the behavior patterns of the high- and low-performance groups were analysed using Lag Sequential Analysis for any statistically significant behavior transition sequence.

## Data analysis

The text was analyzed based on the coding framework consisting of three categories, namely Cognitive Regulation (CR), Task Regulation (TR), and Emotional Regulation (ER), and the respective categories as shown in Table 1.

**Table 1**  
*Coding Framework*

Category	Subcategory	Examples
Cognitive Regulation (CR)	Expressing views (C11)	<i>“I think having a pond here will help water retention on campus...”</i>
	Providing explanation (C12)	<i>“... because precipitation through rain can collect here and be gradually filtered for future use.”</i>
	Sharing information (C13)	<i>“There is a pond on the east side of the campus.”</i>
	Commenting on own views (C14)	<i>“Oh, I was wrong. I missed out that the soil here is not suitable for holding water.”</i>
	Commenting on others’ views (C15)	<i>“You may have missed out the point that simple filtration is not sufficient here.”</i>
	Seeking help (C16)	<i>“What is the national standard for tap water processing?”</i>
Task Regulation (TR)	Forming action plans (C21)	<i>“We need to examine three areas, starting from the teaching compound, followed by living quarters and other amenities. Let’s distribute the workload”</i>
	Taking on Tasks (C22)	<i>“I’ll research on the requirement for science lab water recycling protocol.”</i>
	Coordinating resources and action (C23)	<i>“let’s use the tablet for drawing now. you can use your phone for research.”</i>
	Time management (C24)	<i>“We have about 20 minutes left.”</i>

Emotional Regulation (ER)	Task monitoring (C25)	<i>"This looks good enough. Let's move on to the next" or "(after prolonged discussion), ok, let's move on and revisit this later."</i>
	Emotional comments (C31)	<i>"Yes, but...."</i>
	Emotional consensus (C32)	<i>"..... right?"; "....., agree?"</i>
	Positive emotions (C33)	<i>"Good job!"</i>
	Other emotions (C34)	<i>"Oh, this is so tough!"</i>

### Distribution of learners' regulation behaviors

The descriptive statistics of the semantic units in each subcategory show the distribution of learners' regulation behavior, as shown in Table 2. At a group level, the mean number of units under Cognitive Regulation, Task Regulation, and Emotional Regulation are 287.98, 57.38, and 55.07, respectively. Particularly under the Cognitive Regulation category, the learners appear to focus more on "expressing views" (group mean of 141.14 units) and "seeking help" (group mean of 54.19 units). Under the Task Regulation category, more semantic units fall under "task monitoring" (group mean of 24.52 units) and "coordinating resources and actions" (group mean of 21.81 units). Whereas for the Emotional Regulation category, most of the units fall under "emotional comments" (group mean of 43.57 units).

**Table 2**

*Descriptive Statistics of the Number of Semantic Units in the Subcategories (N=42)*

Subcategory	Group Mean of Units	Subcategory	Group Mean of Units	Subcategory	Group Mean of Units
C11	141.1429	C21	5.619048	C31	43.57143
C12	24.92857	C22	3.904762	C32	2.738095
C13	32.16667	C23	21.80952	C33	2.857143
C14	5.166667	C24	1.52381	C34	5.904761905
C15	30.38095	C25	24.52381		
C16	54.19048				
Cognitive Regulation Category Mean = 287.9762		Task Regulation Category Mean = 57.38095		Emotional Regulation Category Mean = 55.07142857	

A further exploration of the distribution between the high- and low-performance groups was carried out. Based on the average score by the two independent subject matter experts, the top 27% groups (11 groups) are identified as the "high-performance groups", and the bottom 27% as the "low-performance groups". As the data does not fulfill normal distribution, the Mann-Whitney U Test was performed to examine the differences in the number of units in the three categories between the high- and low-performance groups. The results indicate no statistically significant difference in the total number of units ( $U = 53.00, p = .652$ ), Cognitive Regulation ( $U = 55.00, p = .748$ ), Task Regulation ( $U = 59.00, p = .949$ ) and Emotional Regulation ( $U = 45.50, p = .332$ ) between the two groups, as shown in Table 3.

**Table 3**

*Mann-Whitney U Test Results between High- and Low-performance Groups*

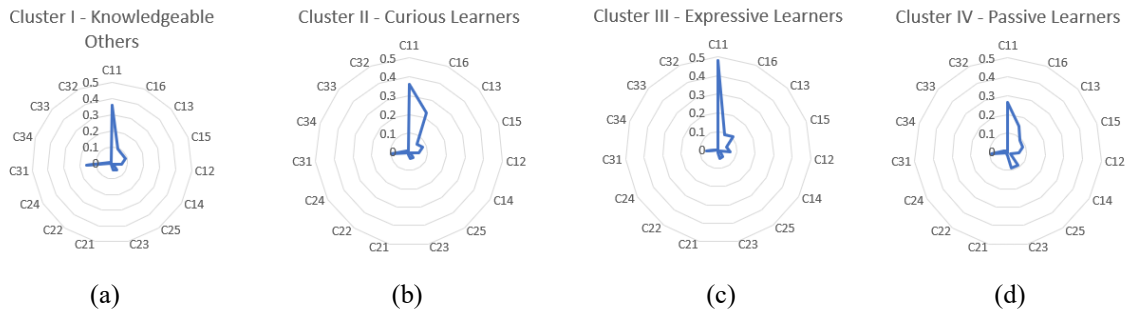
	Total Number of Units	Number of CR Units	Number of TR Units	Number of ER Units
Mann-Whitney U	53.000	55.000	59.000	45.500
Wilcoxon W	119.000	121.000	125.000	111.500
Z	-0.492	-0.361	-0.099	-0.986
Asymp. Sig. (2-tailed)	0.622	0.718	0.921	0.324
Exact Sig. [2*(1- tailed Sig.)]	0.652	0.748	0.949	0.332

### Clustering based on regulation behavior for learner types

Clustering methods were used to mine the potential types of learners in the CSCL process based on their regulation behavior traits demonstrated. To do so, the percentage of the semantic units generated by each member compared to the group was calculated. This percentage was calculated for each subcategory and used for the K-means

clustering method. When using four clusters, the analysis results achieved a silhouette coefficient of 0.20, indicating a reasonable separation of clusters. The characteristics of the four types of learners are illustrated in Figure 1.

**Figure 1**  
*Radar Charts for Cluster I (a), Cluster II (b), Cluster III (c) and Cluster IV (d)*



From the analysis, 38 learners fall under Cluster I, where the learner demonstrate consistent good contribution of views while providing constructive comments on others and positive emotional comments to encourage others' participation. As such, we coin them as the “Knowledgeable Learners”, to some extent playing the role of “knowledgeable others”.

Cluster II consists of 29 learners, who tend to be more active in stating their views and asking questions but less participative in the Task Regulation and Emotional Regulation aspects. This indicates that these learners may be driven by their desire to address questions arising from CSCL; hence, we coin them as the “Curious Learners”.

Another 25 learners fall into Cluster III, where learners are eager to express their views and comment on others. However, they do not seem to sufficiently substantiate their views with explanation or information, and hardly do they engage other members in Task Regulation or Emotional Regulation. Because of their eagerness to put forth their views, we coin them as the “Expressive Learners”.

The final 35 learners fall into Cluster IV, where the learners appear to be less participative in all subcategories under the Cognitive Regulation, especially in “expressing own views”. Nonetheless, they attempt to be part of group dynamics by staying active in “task monitoring” and “coordinating resources and actions”. It seems that the lack of prior knowledge and confidence has pushed them into a more passive stance, waiting to receive information. Hence, we coin them as the “Passive Learners”.

### Regulation behavior patterns of high- and low-performance groups

The software GSEQ 5.1 was used for the Lag Sequential Analysis for the difference in traits and patterns of regulation behavior between the high- and low-performance groups. Where the z-score is greater than 1.96, the behavior transition sequence is deemed statistically significant.

The z-scores for behavior transitions between the categories of Cognitive Regulation, Task Regulation and Emotional Regulation are shown in Table 4. From the analysis, both high- and low-performance groups demonstrated four significant transition sequences, which are CR→CR, CR→ER, TR→TR and ER→CR for the high-performance groups and CR→CR, CR→ER, TR→TR, and ER→TR for the low-performance groups.

**Table 4**  
*Z-scores of High- and Low-performance Groups at Category Level*

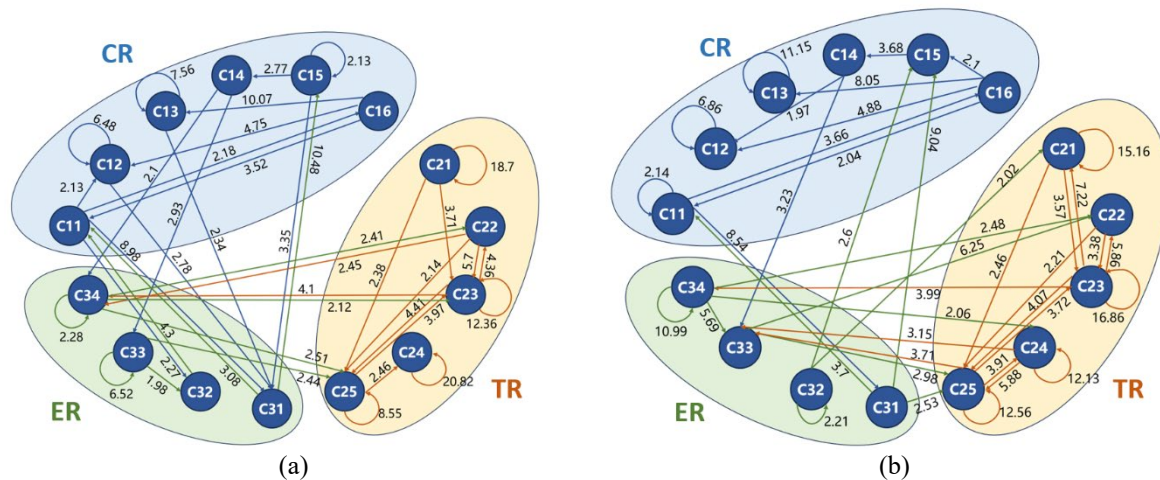
	High-performance groups			Low-performance groups		
	CR	TR	ER	CR	TR	ER
CR	<b>8.46*</b>	-18.41	<b>7.94*</b>	<b>15.33*</b>	-22.27	<b>3.63*</b>
TR	-14.67	<b>21.53*</b>	-3.01	-19.82	<b>25.1*</b>	-0.61
ER	<b>4.1*</b>	1.87	-7.3	0.97	<b>2.69*</b>	-4.29

A further examination of the behavior transition sequence at the subcategory level revealed more details on the different behavioral patterns between the high- and low-performance groups. The high-performance groups



demonstrate 40 statistically significant transition sequences, while the low-performance groups demonstrate 41. Among the sequences, 27 appear in both groups. Figure 2 illustrates the significant transition sequences with the respective z-score indicated. The high-performance groups demonstrate unique behavior sequences of C11 → C12, C11 → C32, C12 → C31, C13 → C31, C14 → C34, C15 → C15, C15 → C31, C22 → C34, C32 → C11, C33 → C32, C33 → C33, C34 → C23 and C34 → C25; whereas the low-performance groups demonstrate unique sequences of C11 → C11, C14 → C12, C16 → C15, C23 → C21, C24 → C25, C24 → C33, C25 → C33, C31 → C25, C32 → C15, C32 → C21, C32 → C32, C33 → C22, C34 → C24 and C34 → C33. In summary, the high-performance groups demonstrated 7 unique sequences led by CR category, 1 led by TR category and 5 led by ER category. In contrast, the low-performance groups demonstrated only 3 unique sequences led by CR category, but 4 led by TR category and 7 by ER category.

**Figure 2**  
*Significant Behavior Transition Sequences for High- (a) and Low-performance Groups (b)*



## Discussion

CSCL emphasizes knowledge building and common understanding via exchange and discourse (Bruffee, 1984). Based on the results, although some individuals appear to drift off towards “cooperative task completion”, most groups can regulate their cognitive activities to allow views to be deliberated and evaluated or questions raised seeking collaborative answers. Meanwhile, the positive emotional comments before expressing a critical view of others, such as “yes, but...” are also prominent throughout the discussion, indicating mature learners’ respect for others and consciously building a positive group dynamic. Generally, Task Regulation and Emotional Regulation appear to be of similar weightage in the collaborative learning process and reasonably maintained throughout, not necessarily limited to when facing challenges (Malmberg et al., 2015).

## Learner types and role of facilitators

At the individual level, the four types of learners, namely Knowledgeable Learners, Curious Learners, Expressive Learners, and Passive Learners, displayed their unique patterns in CSCL.

Possibly having advantages in prior knowledge, the “Knowledgeable Learners” often appear to play the leading role in the collaborative learning process. While their openness in sharing views and information may appear similar to the “Expressive Learners”, it is the depths of discourse with strong substantiation of their views that set the “Knowledgeable Learners” different. In contrast, “Expressive Learners” tend to list their views without dwelling on the details and sometimes may jump to conclusions. Distinguishing the “Expressive Learners” and guiding them to develop into the “Knowledgeable Learners” will be an important task for the discussion moderator in the event of a facilitated discussion.

The frequent switch between expressing views and seeking help suggests robust self-regulation of cognitive activities as the “Curious Learners” move through the “ask-and-tell” process. While they may be active and possibly effective learners, they tend to zoom too quickly into the details and miss the “big picture” or digress to explore topics less related to the task. For example, when discussing about the possible use of spaces between two dormitories, a “Curious Learner” jumped in to questions specific to the space, such as “What is this space for?” or “Can we store bicycles there?”, but not in relation to the dormitories. This observation is in line with the findings from Järvelä et al. (2016) that low-performance individuals may reflect regulation on their own learning

but without a joint understanding of the task, goal, or plans. In such a case, a firm “Knowledgeable Learner” may be able to bring the “Curious Learner” back on track, but more likely, a facilitator intervention may be helpful.

Finally, being quiet most of the time, it is easy to perceive the “Passive Learners” as detached from the group or not participative in the CSCL process. However, the reason for the passiveness may be multifaceted. Whether it is due to a lack of prior knowledge and confidence, lack of motivation, or other reasons, the facilitator needs to address the cause accordingly. Also, it must not be assumed that learning cannot take place passively, and it is possible that these learners prefer putting their cognitive resources to processing the new knowledge internally rather than demonstrating behaviors such as expressing and asking questions.

Unfortunately, the current research does not reveal the extent to which the learner type is resultant of personality traits or shaped by the group composition.

### Behavior pattern comparison between the high- and low-performance groups

The high- and low-performance groups show no statistically significant differences in the number of semantic units in each category, indicating that it is the quality rather than the quantity of conversational exchange that determines the learning achievement.

Based on the Cognitive Regulation behavior transition sequence, the discussions in the high-performance groups demonstrate more depth, allowing each viewpoint to be deliberated and evaluated, questions addressed, and conflicts resolved to achieve conceptual and emotional consensus among the group members. In contrast, although the low-performance groups have a similar quantity of semantic units related to the subject topic, they demonstrate a strong pattern of continuous expression and a broad spread of ideas without sufficient justification or questioning. The discussion stays at a relatively superficial level. Rushing for task completion becomes a higher priority than learning something.

Specific to Task Regulation, the difference between the high- and low-performance groups becomes more apparent. The only significant behavior sequence led by Task Regulation in the high-performance groups is from “taking on task” to “other emotions”, typically reflecting the individual expressing humility after taking the responsibility. On the other hand, the low-performance groups have four significant sequences led by Task Regulation, generally leading to “positive emotion” and “task monitoring”, reflecting the individuals being happy with their output and rushing to move on. This indicates a vast deviation of their perceived quality of work from the actual. This result is coherent with the previous finding that learning achievement is moderated by learning accuracy (Haataja et al., 2022). Having no significant difference in the distribution of the learner types between the high- and low-performance groups agrees with the finding that students with lower self-monitory accuracy can achieve higher learning when co-regulation is frequent in the group (Haataja et al., 2022).

A further comparison reveals that in high-performance groups, “forming action plans” (C21) and “time management” (C24) are only activated by behaviors within the TR category. In particular, the C21 is only activated by itself but leads to other critical behaviors in TR. This suggests that the high-performance groups are effective planners, allowing them to concentrate better on the construction of learning. In contrast, in the low-performance groups, C21 can be activated by “emotional consensus” (C32), indicating that follow-up plans being made based on emotional consensus without engaging in cognitive behaviors. Also in the low-performance groups, C24 can also be activated by “other emotions” (C34), meaning that low-performing groups appear to be frequently referring to time constraints to avoid emotional challenges. Hence, it appears the ability to effectively plan and accurately monitor tasks demonstrated in the high-performance groups help them to better focus on more thorough discussion of viewpoints; whereas to the low-performance groups, ensuring task completion within the specified time frame appears to be of higher priority.

In regulating emotions, the high-performance groups use emotional consensus and positive emotions to reinforce cognitive exchanges, whereas the low-performance groups tend to overwrite other emotions with positive emotions without addressing the cause of the concern.

### Conclusions

This research focuses on the regulation of behavior patterns in CSCL, specifically examining the learners’ Cognitive Regulation, Task Regulation, and Emotional Regulation, at the individual levels and between the high- and low-performance groups. The findings provide a scientific basis for student grouping and the design of learning scaffolds in CSCL, but due to physical constraints the research could not follow through learners over a more extended period of time for deeper mining of learner behavior patterns and validate the consistency. In future research, it is suggested to expand the sample size further, and conduct longer-term tracking to enhance the scientific rigor and applicability of the research. A further exploration on the relationship between group performance and the distribution of learner types can potentially reveal more insight into CSCL group design and facilitation strategies.

## References

- Bruffee, K. A. (1984). Collaborative learning and the "conversation of mankind". *College English*, 46(7), 635-652.
- Bruffee, K. A. (1999). *Collaborative learning*. Johns Hopkins University Press.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of educational research*, 64(1), 1-35.
- Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: strategies for the heterogeneous classroom third edition*. Teachers College Press.
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In: Oxford: Elsevier.
- Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Volume 7 Issue 1 (fall 1995)*.
- Haataja, E., Dindar, M., Malmberg, J., & Järvelä, S. (2022). Individuals in a group: Metacognitive and regulatory predictors of learning achievement in collaborative learning. *Learning and individual differences*, 96, 102146.
- Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *Handbook of self-regulation of learning and performance* (pp. 83-106). Routledge.
- Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational psychologist*, 48(1), 25-39.
- Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*, 42, 1-11.
- Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction*, 70, 101090.
- Johnson, D. W., & Johnson, R. T. (1974). Instructional goal structure: Cooperative, competitive, or individualistic. *Review of educational research*, 44(2), 213-240.
- Koivuniemi, M., Järvenoja, H., & Järvelä, S. (2018). Teacher education students' strategic activities in challenging collaborative learning situations. *Learning, culture and social interaction*, 19, 109-123.
- Laal, M. (2013). Collaborative learning; elements. *Procedia-Social and Behavioral Sciences*, 83, 814-818.
- Laal, M., Geranpaye, L., & Daemi, M. (2013). Individual accountability in collaborative learning. *Procedia-Social and Behavioral Sciences*, 93, 286-289.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.
- Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in human behavior*, 52, 562-572.
- Slavin, R. E. (1991). *Student team learning: A practical guide to cooperative learning*. ERIC.
- Tian, H., & Wu, F. (2022). The Multimodal Characterization and Analysis Path of Collaborative Cognitive Engagement in Blended Scenario. *Journal of Distance Education*, 40(4), 35-44.
- Vygotsky, L. (1978). *Mind in Society*. Harvard University Press.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.
- Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International journal of educational research*, 41(6), 466-488.
- Winne, P. H., Hadwin, A. F., & Perry, N. E. (2013). Metacognition and computer-supported collaborative learning. In A. O. D. C. Hmelo-Silver, C. Chan, & C. Chinn (Ed.), *International handbook of collaborative learning* (pp. 462-479). Taylor & Francis.
- Zheng, L. (2016a). Analysis of Co-regulation Behavioral Patterns by Cluster and Sequential Analysis in CSCL. In (pp. 49-63). Springer Singapore Pte. Limited.
- Zheng, L. (2016b). Analysis of Socially Shared Regulation in CSCL. In (pp. 65-81). Springer Singapore Pte. Limited.
- Zheng, L. (2016c). A Novel Approach to Analyzing Collaborative Knowledge Building in Collaborative Learning. In (pp. 3-18). Springer Singapore Pte. Limited.
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education* (pp. 311-328). Routledge.

## Student Perspectives on Learning and Teaching Data Ethics Through Speculative Game Design

Nisha Devasia, University of Washington, [ndevasia@uw.edu](mailto:ndevasia@uw.edu)  
Andrea Figueroa, University of Washington, [afigue@uw.edu](mailto:afigue@uw.edu)  
Sarah Evans, University of North Texas, [sarah.evans@unt.edu](mailto:sarah.evans@unt.edu)  
Bernease Herman, University of Washington, [bernease@uw.edu](mailto:bernease@uw.edu)  
Cecilia Aragon, University of Washington, [aragon@uw.edu](mailto:aragon@uw.edu)

**Abstract:** From our smartphones to our social media, artificial intelligence (AI) algorithms are becoming ubiquitous in our everyday lives. However, the conveniences that they bring come alongside many potential social and political harms. It is imperative that members of the public develop data ethics literacy to interpret AI's harms and benefits daily. The immersive and transformative nature of games may enable a wide range of people to explore complex ethical concepts in AI and data science through the lens of speculative design. In this project, we focus on the learning process of a diverse group of students from two universities as they embark upon a process of game design to teach ethical thinking in data science/AI. Through qualitative analysis of semi-structured interviews, we apply a speculative game design framework to identify aspects that aid student learning.

### Introduction

We interact with artificial intelligence and machine learning algorithms (AI/ML) multiple times every day – when we use facial recognition to unlock our phones, when we scroll social media, or when we tell our voice assistants to turn off our lights. However, the general public is often unaware that they are interacting with algorithms which may propagate data privacy issues. Given the pervasiveness of AI/ML in our lives, it is important that users understand its many risks and potential ethical concerns, such as algorithmic bias (Buolamwini & Gebru, 2018), misinformation, and discrimination (Mehrabi et al., 2022). Traditional computer literacy education currently does not address AI/ML's sociopolitical effects and fails to provide students with the skills to make informed choices about their individual usage of AI/ML enabled technologies (Touretzky et al., 2019). A more nuanced understanding of their human impact is needed in computer science and data science education (Aragon et al., 2022; Herman et al., 2020).

Given that AI/ML are complex topics for beginners to conceptualize, research in AI education has emphasized interest-driven learning: educational experiences that tap into students' personal passions and hobbies (Long & Magerko, 2020). To this end, games are often positioned as a salient approach to engage with students growing up in the digital age. More than 90% of children over 2 years old play video games (Alanko, 2023), and the games industry is one of the fastest growing markets in the world. Games are effective vehicles for creative learning (Koster, 2013), and have the unique quality of encouraging the user to give direct answers to questions posed through the choices they make in the game. Their sandbox-like nature especially suits them to engaging with complex problems: games can function as thought experiments that model such problems by situating them in simulated worlds (Schulzke, 2014). Speculative design, “an approach to design in which designers create a product or object connected to an imagined scenario” (Barendregt & Vaage, 2021), is a tool commonly used to engage in thought experiments. This approach is especially useful when considering “wicked problems” (Rittel & Webber, 1973), which lack defined aims or solutions, and defy definitive formulations. Notably, speculative design shares key traits with game design, such as prototyping as a method of inquiry, and using fiction to represent alternative futures (Auger, 2013).

Coulton et al. (2016) present a framework for using games as a method of speculative design, which allows designers and players to explore alternate presents and plausible futures. The framework suggests using the following elements: *plurality*, *plausibility*, *mimesis (enactment)* and *diegesis (narrative)*, *iteration*, and *the avoidance of reductionism*. Plurality implies that different worldviews should be incorporated within the design process and the game itself. Plausibility means that game scenarios should enable players to connect familiar elements of daily life with authoritative sources of data. Mimesis involves the player enacting the game through play, while diegesis is presented through cutscenes and character dialogue. Although iteration is a common design practice, Coulton et al. emphasize including all participants in the process and the need for reflection time. Finally, complex societal problems cannot be reduced to solutions addressable through minor behavioral changes or overly simplistic goals. In this work, we build upon this framework and draw from previous research that positions students as *designers* of games (Baradaran Rahimi & Kim, 2019; Kafai & Burke, 2015; Tan &

Kim, 2015). The game development process forces the creator(s) to set the rules that govern players' choices and follow through with the repercussions of their decisions. Gualeni described the process of self-transformation through game design, in which "designers inevitably self-fashion themselves in ways and in directions that are analogous to those that they intended to elicit in the recipients of their work" (Gualeni, 2015). Designing an educational game requires that designers become familiar with the topics the game aims to teach and the best ways of teaching it, which requires critical thinking skills and a deep understanding of the subject matter (Gee, 2008). Designing games allows for deep personal reflection and provides a space to form opinions on complex issues (Schwind & Buder, 2012). Furthermore, considering the recent framing of several issues surrounding AI/ML's impact on society as wicked problems (Holtel, 2016), combining educational game design with a speculative approach also allows for divergent exploration of this impact.

To explore this approach of designing educational games about AI/ML data ethics, we conducted a two-year long interdisciplinary study with students from two institutions (Byun et al., 2022). The instructors had expertise in data science, library science, and AI. Students ranged in age and experience from dual enrolled high schoolers to graduate students. The stated goal of the project was to collaboratively design games that teach critical and ethical thinking about data ethics in AI/ML and the effect that such technologies can have on our society. This paper explores the research question: What are the constraints and affordances of learning through speculative game design across two academic institutions? We contribute to the less studied area of *making* rather than *playing* games for learning (Kafai & Burke, 2015), and build on Coulton's framework of games as speculative design to gather insights into students' learning processes.

## Methods

The goal of this project was to design ethical games with a diverse group of students from two institutions. A listing describing the project was advertised through the University of Washington, and a researcher from the University of North Texas distributed the posting through student groups on campus. Sixteen students across both institutions from a range of college programs with interests in gaming, learning, data science, and/or research signed up to participate. Structured as a Directed Research Group (DRG), the course met weekly for 1.5 hours over two quarters or one semester, depending on the schedule of the student's home institution.

## Directed research groups

Directed research groups (DRG) are a decentralized learning experience that model the student-as-collaborator relationship (Larson et al., 2009; Turns & Ramey, 2006). They are intended to engage undergraduate and masters students in various phases of the research process, and were created in response to calls for more opportunities in undergraduate research, in accordance with the standards set by the Boyer Commission on Educating Undergraduates, the Council on Undergraduate Research, and the National Conference for Undergraduate Research (Hu et al., 2008). DRGs are normally led by doctoral students and faculty members, who mentor and empower students to both participate in and conduct their own research. DRGs encourage diverse collaboration among students and between students and academic leaders, emphasizing exploration and diversity as fundamental to research. This format draws inspiration from project-based learning (PBL) pedagogy, which de-emphasizes lecture-based content transmission in favor of interest-driven learning frameworks. Furthermore, research has shown that applying knowledge to real-world problems supports deeper learning than lecture-based learning (Miller & Krajcik, 2019).

The first academic term of the DRG was designed to establish a common base of knowledge for participants in data science, AI, ethics, research, learning, and gaming, as well as to design an initial prototype of a game by the end of 6 months (1 semester or 2 quarters). Students engaged with the material on various levels: 1) assignments and readings curated by the instructors, 2) self-directed meetings with a small group to work on a video game addressing a specific ethical issue in data science or AI/ML, and 3) weekly class time where students and instructors met over Zoom to participate in lectures, large group discussions, and joint exploration of serious games. Teaching materials included case studies of biased algorithms which resulted in actual human harms, immersive game design mechanics, and game studies literature on narrative elements. Students combined the technical learning from the data science lectures with techniques learned from game design to create multiple iterations of their game within each group (see Figures 1-4 for examples of student work).

## Participants

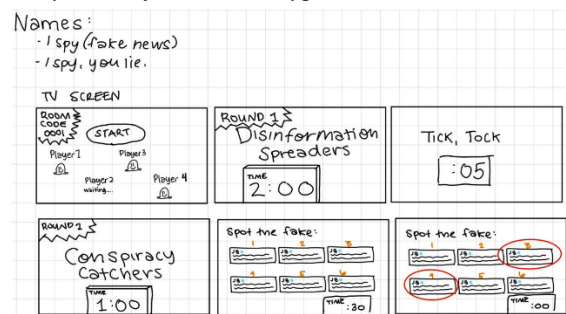
The participants consisted of 10 students from the University of Washington and 6 students from the University of North Texas. These participants expressed interest in joining the group based on the description and were selected based on their background and skills in topics such as gaming, education, data science, and research, with emphasis on recruiting underrepresented students. The group consisted of 9 undergraduates and 7 graduate

students. Two of the undergraduates were part of an accelerated program in which they finished the last two years of high school and the first two years of college concurrently. The majority of the students identified as members of groups underrepresented in STEM, including women and people of color (see Table 1).

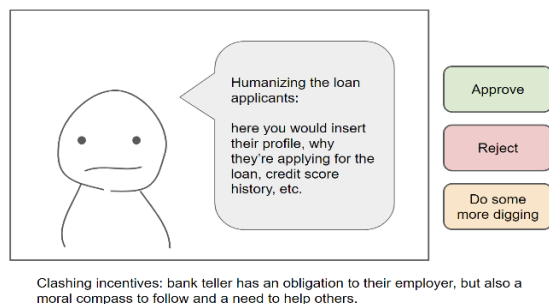
**Table 1**  
*Demographic Information of Participating Students in the DRG.*

	Gender	Ethnicity	Student Status
Univ. of Washington	M = 3	White = 4	Domestic = 9
	F = 7	Asian = 6	International = 1
Univ. of North Texas	M = 4	Black = 1	Domestic = 4
	F = 2	White = 1	International = 2
		Asian = 4	

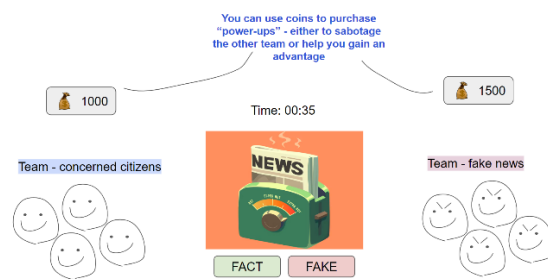
**Figure 1**  
*Storyboards from a Prototype Fake News Game.*



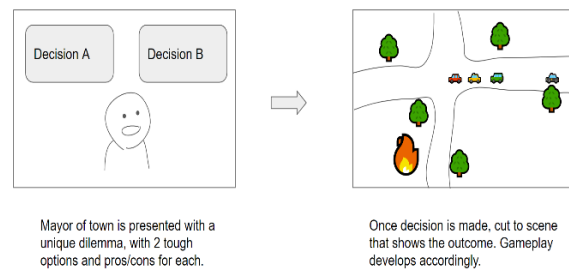
**Figure 3**  
*Storyboard from Prototype Game about AI in Banking.*



**Figure 2**  
*Paper Prototype for Misinformation Game Mechanics.*



**Figure 4**  
*Game Flow for Making Ethical Decisions In-Game.*



## Interviews

We conducted semi-structured interviews with each participant in a one-on-one setting, ranging from 7 to 30 minutes. Based on the research questions for the overall project, these interviews focused on student perspectives on data ethics, the experience of collaborative design, and their own relationship to the field of data science. Interviews were conducted virtually the week after the end of the first quarter and recorded and transcribed for later analysis. Examples of questions included:

- What were your expectations when you signed up for the DRG? How has the experience been different than you expected? In what ways has the experience been what you expected?
- Have any of your ideas changed around the topics of big data, ethics, research, learning, and gaming?
- What DRG activities have you enjoyed the most/least? What has been most/least useful?
- Do you identify with STEM as part of your studies and career?

## Analysis

The first and second author each independently performed a round of inductive coding on the interview transcripts to find emerging themes (Merriam & Tisdell, 2015). After developing an initial set of themes, we reanalyzed the transcripts with a focus on said themes. We focused on insights that spoke to what specific elements of the DRG helped students learn and why those particular elements were the most conducive to their learning.

## Findings

Among the interviews, different themes emerged about how the DRG elements impacted students' learning. We organized our findings around which activities and aspects students found the most helpful, applying Coulton et al.'s framework and including suggestions the students had on how to improve the DRG in future iterations.

### Group organization

*"It's more like an interactive class, where each one of us gets an opportunity to express our ideas and come up with new things."*

Guided discussions among both the entire group (the entire class and teachers) and the smaller groups (project groups of 3 or 4 students) were crucial to the students as they shared their views, heard different perspectives, and formed opinions on both ethical and technical topics. Educators must take care to not be reductionist when teaching about ethics; for example, believing that there is a "correct" or "preferable" solution to an ethical question. As Coulton et al. remark, 'preferable' should be a question the designers ask of themselves within the design activity rather than an aim of the design. The DRG structure greatly facilitated this self-exploration and group discovery, with students asking questions of each other and listening closely to diverse viewpoints. As Schwind and Buder point out, allowing time for reflection within the groups enabled students to thoughtfully consider opposing perspectives rather than "select[ing] information that confirms their prior perspectives".

### Small groups

*"I really enjoyed the small groups. I'm usually not a group person, I prefer working alone...but in our smaller groups I just really enjoy the idea sharing that happens there."*

Students spoke extremely positively of the small group discussions, with nine of them explicitly stating that it was the most useful aspect of the DRG organization. The ease of expressing one's own opinion was mentioned repeatedly in the interviews. This echoed Coulton's emphasis on inclusivity within the actual design, specifically the focus on democratizing the design process and encouraging the inclusion of all voices, not just those who are in the position of privilege or who agree with the dominant societal narrative. The DRG consisted of students from different educational levels, with age differences of five years or more. Younger students stated that they sometimes felt unqualified to express themselves effectively in the large group meetings, due to their perceived lack of experience. The more intimate setting promoted engagement, and allowed less-experienced students to openly share their views. Students repeatedly stated that they benefited from the open sharing of ideas and iteration that occurred within the smaller groups. The structure of the small groups ensured that the diverse group of students participating in the DRG saw each other as peers.

### Large groups

*"This is by far the most engaged discussion I've had so far in terms of a remote context."*

Large group discussions offered students numerous advantages that enhanced the learning experience and fostered a sense of community among participants. One student described the large group discussions as "more like an interactive class, where each one of us gets to express our ideas and come up with new things." Coulton points out that linking authoritative data sources (expert views from instructors) with both narrative discussion (diegesis) and student lived experience (mimesis) enables more effective exploration of complex societal issues. The three students that identified the large group discussions as the most useful aspect of the DRG stated that they felt their input was actively listened to and the interaction felt receptive and reciprocal. One student noted the sense of community and engagement that was promoted when instructors encouraged students to speak and to have their cameras on. Moreover, large group discussions allowed students to explore a wide range of perspectives, allowing participants to discover both commonalities and differences.

### Diversity of thought

*"It's crazy how...we can all come up with such interesting and different ideas. Coming from industry, where they harp so much on being diverse, you never really see it in action. But seeing [diversity] come to fruition and an actual diverse group, it's really neat."*

From physical geography to personal identity, the students and faculty that participated in the DRG were a striking example of diversity and intersectionality in action. This relates strongly to Coulton's framework element of plurality. The three instructors were women in STEM, bringing together 1. a Hispanic senior faculty member from the University of Washington, a large public university, 2. a Black early career data science researcher from the

same institution, and 3. a junior faculty member at the University of North Texas, a designated Hispanic-Serving Institution (HSI) in the American south. The two graduate teaching assistants were Hispanic women. The majority of the students identified as underrepresented in STEM. Three students were international, bringing non-American viewpoints to the discussions. The students spanned a wide range of specializations, including psychology, mechanical engineering, and business, with most majoring in human-computer interaction.

Students stated explicitly that the diversity in experience enabled them to generate a wide range of innovative ideas. By embracing a variety of perspectives, students found their collective brainstorming sessions to be particularly enriching. One student said: “I really love my teammates’ ideas and how differently they were thinking, and how that made it really interesting to meet up with them.” Another emphasized the collaborative nature of the discussions, stating: “We all try to come up with ideas together and try to take something from each person’s ideas.” The diverse environment fueled an equally diverse list of game ideas (see Table 2).

**Table 2**  
*A List of Game Ideas that Students Devised by the End of the First Academic Term.*

Game Idea	
Group 1	Players experience a day as a hacker. Players receive info on their phones that they must keep private from other players. Teams must decide if they can trust the information they've received and at what time to share it. Inspired by Bomb Corp.
Group 2	Players experience the flow of using AI to approve banking loans. The win condition involves collaborating with the AI rather than letting the AI make all the decisions on its own.
Group 3	Players roleplay as the mayor of a fictional town and have to make difficult ethical decisions to progress the game. Players are shown the effects of their actions upon the citizens of the town.
Group 4	Players engage in a family-friendly Jackbox-style game to both roleplay as a misinformation spreader and a receiver.

### Self-transformation

Several students reported that participation in the DRG had changed their personal beliefs or everyday actions. As Coulton mentions, the merging of mimesis, which often includes empathy, with narrative diegesis enables powerful and potentially life-changing personal exploration through game design. We identified two themes among these responses which were particularly salient to our understanding of what students took away from their experience.

#### Increased understanding of data science and ethics

*“Before, I didn’t see a lot of risks of future technology because I was so interested in the positives. But now I see how it can be used in ways where you’re like, okay we can still try to do this, but we need to have more control over it.”*

Combining technical lectures, group discussions, and active engagement in game design to teach data science led to tangible changes in students' perspectives and actions. The lectures taught students how to be “a little more careful on what to download...[we] have to be more careful of what [we] do with each person’s data and what [we] can use it for.” This is a clear example of the blending of mimesis and diegesis. The discussion helped broaden students’ perspectives on data science and ethics and had them confront the ways in which these topics affect their daily lives. As one student stated, “I didn’t think too much about [data science]...but I realized that everything I do is affected by it and learning about bias and how that can create prejudice towards people was really eye opening.” Students were also grateful for discussions in which instructors broke down real world examples of how data bias can affect others, which empowered them to explain these examples to non-experts in their lives. This relates to Coulton’s theme of plausibility and the importance of co-designing with experts.

#### Deconstructing negative notions of games

*“My insight of what a game is has completely changed.”*

The introduction of learning games within the educational scope of the DRG, as well as students’ active participation in both playing and designing games, were crucial to understanding the pedagogical potential of games. The majority of students in the DRG did not identify as gamers, either prior to or after the DRG. However, several students experienced a complete shift in their perceptions of video games, both recreationally and for learning purposes. Initially, some students held negative notions of both video games and their players. For example, one student stated: “My parents influenced me [to think] that only losers do that sort of thing.” Others believed that gaming was meant solely for entertainment, not learning. However, because of their participation in



the DRG, the majority of students came to view gaming in a different light. They were more conscious of the intention behind games, and approached game design more thoughtfully. Furthermore, some students stated that their conversations about gaming outside of the scope of the DRG also changed significantly, and that they are spreading their newfound perspectives to family members and friends. This iteration and reflection demonstrate that profound explorations can take place within the process of speculative game design.

When asked which aspect of the project they were looking forward to the most, an overwhelming majority of students (15 out of 16) responded that they were excited to work on the game prototypes, see other group's game ideas, and test out their game. This underscores the appeal of game development as a creative and engaging activity, transcending traditional gamer demographics. The development process itself became a powerful catalyst for enthusiasm and involvement, making it a valuable tool for fostering innovative thinking across diverse groups.

## Organizational complications

Coordinating students from two geographically disparate institutions led to some suboptimal methods of material organization. The following two subsections list student suggestions on how to improve the DRG.

### Clarifying expectations

*"My group is really good. I'm the only master's student in my group, all three of them are undergrads, so I think there was a tacit understanding that I was gonna lead. It just kind of happened that way, but I don't necessarily feel like I'm more experienced or anything like that, because we're all pretty much new to this field."*

The groups within the DRG were a mix of students from both institutions with varying degree levels, ranging from dual-enrolled high school students to master's students. However, master's students felt that there was an unspoken agreement in their groups that they were going to lead, even if they felt unprepared to do so. Furthermore, some master's students thought that they did not have as much experience with some of the topics as younger group members did. Although instructors stated at the outset of the DRG that more work was expected from graduate students, most of the graduate students did not interpret this as a call to leadership, and were under the impression that they would just have to do more assignments.

### Centralizing communications

*"At least the people I've spoken to, email is an announcement tool. [...] And then we also get so many emails [...] so that is an information overload. And then every single calendar invite is an email so everything is just like coming in. It's really hard to 'chat' with people through email."*

Students offered several suggestions on how to centralize the DRG materials and communications to serve their learning better. Originally, the primary form of group communication was email; messages were sent out weekly with group instructions and materials. However, because students received a plethora of different emails from their department, university, and scheduling tools, DRG related emails were often lost, leaving them unsure of what the deliverables were for the week. The additional overhead of the collaboration between two universities led to emails from unfamiliar domains being filtered or deprioritized. In addition, email communication between team members was thought unnecessarily formal, especially to students who didn't use it as frequently as instructors did. Many of the students resorted to using instant messaging tools to communicate among their small groups. Students self-organized to create a DRG Slack with specific channels for each different aspect of the game design, such as #programming, #art, and #data-science. They also suggested that the instructors upload all the materials in one place, such as Google Drive.

This decentralization of tools was a result of institutional restriction in place by software companies. While both campuses use Canvas as the primary learning management system (LMS), it was not possible to create a course shell for students and instructors in different universities. A Google Classroom was considered, but due to the University of Washington's Google license, students could not access it with their university emails. Having a common LMS would have enabled both discussion features and shared document management.

## Discussion

In this paper, we investigated the process of learning through speculative game design across two separate academic institutions. We structure our discussion around the impact of collaboration between two universities on specific components of the process and the influence of collaborative design in speculative games. We tie

our findings to Coulton's framework for games as speculative design tools, as we found this framework to be particularly helpful in considering the research question of exploring the constraints and affordances of learning through speculative game design in a diverse and geographically separated context.

The primary affordance of collaboration across separate institutions was diversity in lived experiences. Most of the students from the University of Washington, a large research institution on the West Coast, were US students majoring in human-computer interaction. The students from the University of North Texas brought a larger variety of academic and personal backgrounds. Instructors accounted for diversity and intersectionality across group members when assigning the smaller design groups, and students spoke about how this positively affected the ideas their groups were able to come up with. However, the incompatibility of learning management systems and accessibility barriers in decentralized communication tools often made collaboration difficult across the larger group.

Students universally praised the affordances of the speculative game design process, particularly citing themes related to Coulton's elements of *plurality*, *plausibility*, and *mimesis (enactment) and diegesis (narrative)*. Students explicitly stated the diversity (*plurality*) of the group and its contrasting opinions helped them develop their own views. Diversity fosters creativity by encouraging the collision of different ideas and approaches. In the context of speculative design, where innovation is key, diverse teams are more likely to generate imaginative concepts that push the boundaries of traditional thinking. Varied viewpoints and lived experiences, especially with regards to technologies becoming as ubiquitous as AI/ML, are more likely to lead to holistic design solutions for more positive speculative futures.

Students' descriptions of how the DRG had tangibly changed their actions, from discussions with family members about games and the class material to the deconstruction of negative notions of gaming, were evidence of *self-transformation* as described by Gualeni. Self-transformation encourages adaptability and equips students with the ability to navigate uncertainties and contribute to dynamic design processes. This self-transformation was facilitated and augmented by the affordances of the group involving *mimesis and diegesis*. Students, as they ideated and designed their games, were able to enact stories that reflected everyday experiences in their own lives, but also used rhetorical techniques and narrative to elaborate and reflect on those experiences in a way that deepened their own understanding of the game topics. Through designing to teach others about these topics, the students themselves internalized the potential harms of unethical data use and learned to think more critically about how they interact with these technologies every day. The *plausibility* of these everyday scenarios strengthened their understanding of potential harms and ethical concerns. Finally, the *iteration* in the game design process fostered time for reconsideration of design elements and the accompanying deep reflection on personal beliefs. By unpacking issues and investing themselves in the storylines, students gained substantial insight into the dangers of *reductionism* when exploring complex societal issues.

## Limitations

There were several limitations to this work. The size of the DRG was relatively small, with only 16 students and four groups. The data gathered is from the first period of the project and is representative of a snapshot of the entire game design process. The student group, although diverse, still lacked representation from key demographic groups. Although three members of the instructional team were Hispanic, none of the students identified as such. There was only one Black student and no representation from Native American or Pacific Islander students. Furthermore, the research team was not able to investigate how students' personal identities affected the ideas they contributed to the design process. We plan to address these limitations in future iterations of the DRG.

## Conclusion

This project presents qualitative findings from a project involving a diverse group of students and instructors collaboratively engaging in a speculative design process involving games about artificial intelligence, machine learning, and data ethics. The goal of the project was to produce games that inform non-experts about current issues related to AI/ML and data ethics, as well as potential futures that AI/ML could create. Evidence of student learning was gathered as the students collaborated, ideated, and designed four games. We found themes building on the five elements of Coulton's speculative design framework, particularly as related to a diverse and geographically separated group and listed key affordances and constraints in this environment. Our findings reflect the effectiveness of co-designing games to address complex, intersectional issues such as data ethics, encouraging further exploration in speculative game design with diverse collaborative groups. Envisioning a wide range of futures necessitates a wide range of lived experiences and ideas.

## References

Alanko, D. (2023). The Health Effects of Video Games in Children and Adolescents. *Pediatrics In Review*, 44(1), 23–32. <https://doi.org/10.1542/pir.2022-005666>

- Aragon, C., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-Centered Data Science: An Introduction*. MIT Press.
- Auger, J. (2013). Speculative design: Crafting the speculation. *Digital Creativity*, 24(1), 11–35. <https://doi.org/10.1080/14626268.2013.767276>
- Baradaran Rahimi, F., & Kim, B. (2019). The role of interest-driven participatory game design: Considering design literacy within a technology classroom. *International Journal of Technology & Design Education*, 29(2), 387–404. <https://doi.org/10.1007/s10798-018-9451-6>
- Barendregt, L., & Vaage, N. S. (2021). Speculative Design as Thought Experiment. *She Ji: The Journal of Design, Economics, and Innovation*, 7(3), 374–402. <https://doi.org/10.1016/j.sheji.2021.06.001>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Byun, M., Evans, S., Herman, B., Aragon, C., & Rojas-Ponce, D. (2022). Game Development as Speculative Design: Teaching Data Science Ethics Using Decentralized Research Groups. *Proceedings of the Association for Information Science and Technology*, 59(1), 630–632. <https://doi.org/10.1002/pra2.672>
- Coulton, P., Burnett, D., & Gradinar, A. (2016, June 25). *Games as Speculative Design: Allowing Players to Consider Alternate Presents and Plausible Features*. Design Research Society Conference 2016. <https://doi.org/10.21606/drs.2016.15>
- Gee, J. P. (2008). Video Games and Embodiment. *Games and Culture*, 3(3–4), 253–263. <https://doi.org/10.1177/1555412008317309>
- Gualeni, S. (n.d.). *SELF-TRANSFORMATION THROUGH GAME DESIGN*.
- Herman, B., Aragon, C., Evans, S., & Shanley, L. (2020). Advancing diversity in human centered data science education through games. *Proceedings of the Association for Information Science and Technology*, 57(1), e314. <https://doi.org/10.1002/pra2.314>
- Holtel, S. (2016). Artificial Intelligence Creates a Wicked Problem for the Enterprise. *Procedia Computer Science*, 99, 171–180. <https://doi.org/10.1016/j.procs.2016.09.109>
- Kafai, Y. B., & Burke, Q. (2015). Constructionist Gaming: Understanding the Benefits of Making Games for Learning. *Educational Psychologist*, 50(4), 313–334. <https://doi.org/10.1080/00461520.2015.1124022>
- Koster, R. (2013). *Theory of Fun for Game Design*. O'Reilly Media, Inc.
- Larson, J., Birge, C., Huang, Y.-M., Sattler, B., Turns, J., & Yellin, J. M. H. (2009). Directed Research Groups as a Means of Training Students to Become Technica Communication Researchers. *Technical Communication*, 56(2), 172–177. <https://www.jstor.org/stable/43092580>
- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376727>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 1–35. <https://doi.org/10.1145/3457607>
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons.
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169. <https://doi.org/10.1007/BF01405730>
- Schulzke, M. (2014). Simulating Philosophy: Interpreting Video Games as Executable Thought Experiments. *Philosophy & Technology*, 27(2), 251–265. <https://doi.org/10.1007/s13347-013-0102-2>
- Schwind, C., & Buder, J. (2012). Reducing confirmation bias and evaluation bias: When are preference-inconsistent recommendations effective – and when not? *Computers in Human Behavior*, 28(6), 2280–2290. <https://doi.org/10.1016/j.chb.2012.06.035>
- Tan, L., & Kim, B. (2015). *Learning by Doing in the Digital Media Age* (pp. 181–197). [https://doi.org/10.1007/978-981-287-326-2\\_12](https://doi.org/10.1007/978-981-287-326-2_12)
- Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), Article 01. <https://doi.org/10.1609/aaai.v33i01.33019795>
- Turns, J., & Ramey, J. (2006). Active and Collaborative Learning in the Practice of Research: Credit-based Directed Research Groups. *Technical Communication*, 53(3), 296–307. <https://www.jstor.org/stable/43089425>

## Acknowledgements

This work was funded under NSF award numbers 2127924 and 2127867.

## Use of Generative AI for Boundary Crossing in Interdisciplinary Collaborative Research

Shotaro Naganuma, Kyushu University, naganuma.shotaro.062@m.kyushu-u.ac.jp  
Tsubasa Minematsu, Kyushu University, minematsu.tsubasa.659@m.kyushu-u.ac.jp  
Kana Matsueda, Kyushu University, matsueda.kana.460@m.kyushu-u.ac.jp  
Jun Oshima, Shizuoka University, joshima@inf.shizuoka.ac.jp

**Abstract:** Ph.D. students in interdisciplinary education programs face challenges in achieving effective collaboration. To support them in overcoming their challenges, we designed a workshop using generative artificial intelligence (AI) as a broker device to create partially shared objects for boundary crossing. In the workshop, groups of students collaborated by coordinating research themes proposed by ChatGPT with the input of their research information. After their collaboration, experts evaluated improvements in the proposed themes. We then analyzed four high-outcome and three low-outcome groups to identify key discourse moves for successful knowledge-building discourse. Epistemic Network Analysis (ENA) and discourse analysis revealed that compared to low-outcome groups, high-outcome groups actively enhanced explanations, connected/synthesized their discourse, and shared relevant disciplinary domain and research-design knowledge to improve their themes without persisting with those that they initially judged as non-promising. We utilized these results to propose a conjecture map toward future systematic design-based research on interdisciplinary education.

### Introduction

In today's knowledge-based society, interdisciplinary research is expected to escalate innovations more than ever before to solve complex problems, such as global sustainability issues (Repko & Szostak, 2020). Interdisciplinary research is "any study or group of studies undertaken by scholars from two or more distinct scientific disciplines" (Aboelela et al., 2007, p.341). Unsurprisingly, interdisciplinary research is a difficult challenge for professional researchers because of different disciplinary cultures/paradigms and misunderstandings of mutual jargon (Fischer et al., 2011). Akkerman et al. (2006) draw attention to the problem that participants in an interdisciplinary research project do not explore one another's thought worlds. In response to such challenges, interdisciplinary education programs/practices were established to develop interdisciplinary researchers (Tobi & Kampen, 2018).

Many studies (e.g., Repko & Szostak, 2020; Stamp et al., 2015; Tobi & Kampen, 2018) have demonstrated challenges in designing interdisciplinary education programs. For example, in Japan, a program called the K-SPRING has been established for Ph.D. students to develop their skills and knowledge through interdisciplinary collaborative research. Anecdotal data from their previous experiences showed that Ph.D. students often faced considerable difficulties in setting interdisciplinary collaborative research-themes (IDCRTs), although they shared one another's disciplinary knowledge. In addressing such challenges in the interdisciplinary education program, we approached the interdisciplinary collaborative research from the perspective of boundary crossing in the activity theory (Akkerman & Bakker, 2011; Engeström & Sannino, 2021; Engeström, 2022). We designed an exploratory workshop using an emerging technology, generative artificial intelligence (AI).

### Theoretical background

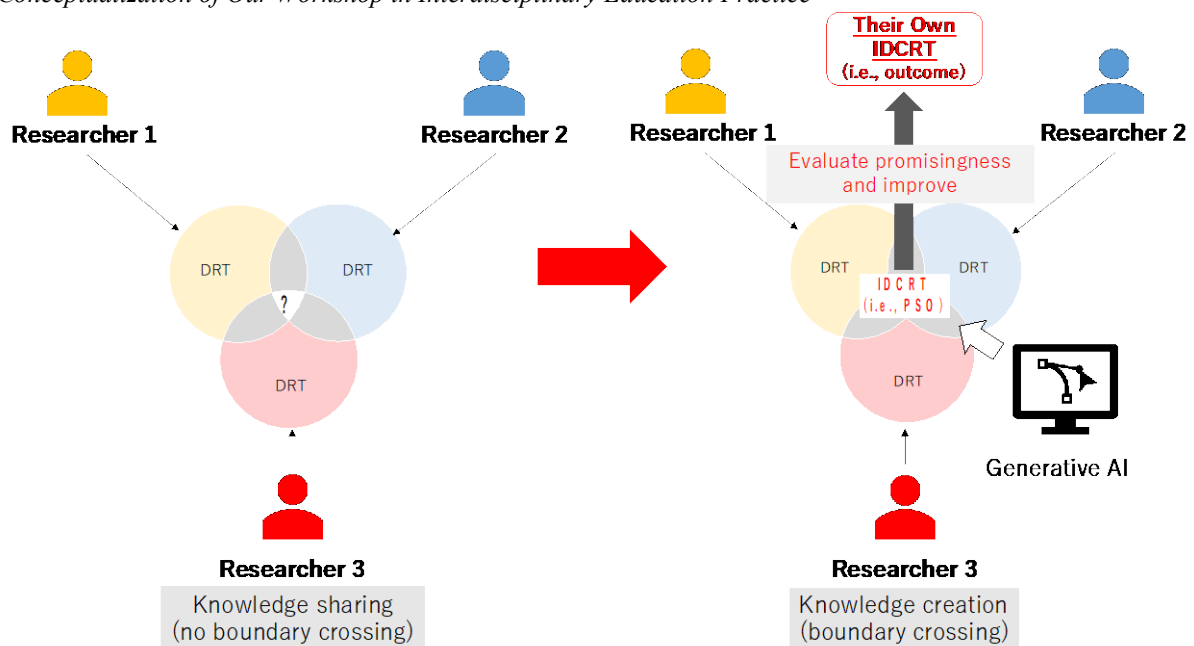
#### Interdisciplinary collaborative research as boundary crossing over activity systems

To design our workshop, we approached interdisciplinary collaborative research by referring to the third generation of the activity theory (Engeström & Sannino, 2021; Engeström, 2022) in the knowledge-creation metaphor (Paavola et al., 2004) (Figure 1). The reason for this approach is twofold. First, interdisciplinary collaborative research involves researchers in expansive learning because they engage in the collaborative advancement of knowledge beyond the knowledge of individual disciplines. Second, interdisciplinary collaboration can be explained as crossing boundaries over activity systems, to each of which a researcher in a specific discipline belongs. Through the conceptual application, we focused on two concepts in designing the workshop: *boundary crossing* and a *partially shared object (PSO)* (Akkerman & Bakker, 2011; Engeström & Sannino, 2021; Engeström, 2022).

Boundary-crossing is "a person's transitions and interactions across different sites" (Akkerman & Bakker, 2011, p.133). In the case of interdisciplinary collaborative research, each discipline has its boundary. Four learning mechanisms occur in crossing a boundary: identification, coordination, reflection, and transformation (Akkerman

& Bakker, 2011). We targeted the first two in this study. Identification is a mechanism through which one becomes “aware of one’s expertise, assumptions, values, and principles” (Fortune et al., 2023, p.4). We found that students engaged in this mechanism in our previous programs. However, students rarely engage in the coordination mechanism, such as “finding means and procedures to work together effectively” (p.4). To explain why coordination in boundary crossing is difficult in interdisciplinary education practices, another concept in the activity system theory is useful: a PSO. A PSO is a jointly constructed object among multiple activity systems that allows those systems to effectively collaborate beyond boundaries (Engeström, 2001; 2022). In the coordination mechanism of interdisciplinary education practices such as the K-SPRING program, students are required to construct IDCRTs as PSOs that enable coordination. As previously discussed, however, it is a difficult challenge for students to collaboratively construct such PSOs.

**Figure 1**  
*Conceptualization of Our Workshop in Interdisciplinary Education Practice*



*Note.* DRT: disciplinary research theme, IDCRT: interdisciplinary collaborative research-theme, PSO: partially shared object

### Creation of a PSO

A PSO is generally created in the intersection of objects in the involved activity systems (Engeström, 2001; 2022). Seilstad (2018), for instance, reported how a teacher and students in a classroom engaged in identification and coordination for boundary crossing in studying citizenship. In discussing citizenship, the classroom teacher presented the concept of a “jury.” The students then requested clarification, which prompted the teacher’s elaborate explanation and students’ deep discussion. Such interaction between the involved activity systems created a PSO, “understanding of a jury as a concept.” In this example, the classroom teacher played a critical role as a *broker* (Akkerman & Bakker, 2011). A broker is a person who attempts to bridge a gap between different activity systems using their knowledge and skills, which cross boundaries.

In interdisciplinary education practices, who should play a broker is another challenging but crucial problem to address for the following reasons. First, in some practices, faculty members function as brokers to prepare PSOs using their knowledge (e.g., Stamp et al., 2015; Tobi & Kampen, 2018). As Fischer et al. (2011) pointed out, however, it is difficult, even for professional researchers, to engage in interdisciplinary research with their own themes. Therefore, it seems profoundly difficult for them to prepare PSOs that fit all students’ expertise. Second, considering such difficulty even for professionals, it is also unrealistic to expect students to create PSOs because they are still at a developing stage of their own expertise. Indeed, our experience in the K-SPRING suggests that even when students with various expertise communicate with one another, the challenges of boundary crossing often result in mere knowledge sharing instead of knowledge creation (i.e., creation of PSOs). In the interdisciplinary education context, this problem is fatal as it will result in a poor coordination process. To solve this problem, we focused on generative AI as a broker to artificially create PSOs for boundary crossing.

## Using Generative AI to create PSOs

To design an effective workshop in interdisciplinary education, we used generative AI as a broker to create IDCRTs as PSOs for students to productively engage in their coordination mechanism in boundary crossing. The use of generative AI has the potential to propose PSOs for all students considering their expertise when we provide it with information on the students' research topics. Based on the input, it can provide research themes as likely word combinations in a massive amount of available online information (Radford et al., 2018). However, the quality of the PSOs prepared by the generative AI is not guaranteed. Therefore, Ph.D. students must conduct their discussion to improve the PSOs prepared by the generative AI through knowledge-building discourse to establish their own PSOs as outcomes.

## Purpose of the present study

Based on the conceptual framework of boundary crossing in the activity theory, we employed generative AI as a broker device to create PSOs (i.e., IDCRTs). We expected Ph.D. students from different disciplines to finally cross the boundary by engaging in the PSOs to construct and improve the PSOs. The following research questions (RQs) were set to guide our study.

When providing PSOs by generative AI in the interdisciplinary collaborative research practice,

- RQ1: Can Ph.D. students improve the PSOs generated by AI?
- RQ2: What key discourse moves lead groups to construct and improve the PSOs?

By answering these RQs, we aimed to develop a conjecture map (Sandoval, 2014) for systematic design-based research on interdisciplinary education in the future.

## Method

### Context and workshop participants

Our workshop was conducted at a national university in Japan. The participants were recruited from the K-SPRING program, which is a funding program that offers opportunities for Ph.D. students to engage in interdisciplinary research practice. Participation in this workshop was voluntary. The workshop was conducted in July 2023, with three sessions involving 56 students from various academic departments (e.g., Engineering, Dental Science, Education, Literature). Each session lasted 100 minutes, and there were no duplicate participants. Groups were reconfigured twice during each session. The participants were divided into groups of three to four, each with at least one member from a different academic department.

### Workshop design

#### IDCRTs generated by AI as PSOs

We used the prompts in Table 1 for AI to generate IDCRTs. Our engineered prompts had three characteristics. First, we asked ChatGPT4.0 to generate five research themes with students' research titles and abstracts as input. Generating five themes was expected to mitigate the risk that the proposed themes were inherently not promising. Second, our prompts included the term "moonshot" to create research themes. This term refers to an ambitious, exploratory, and ground-breaking project in the context of research funding. Third, we instructed ChatGPT to output each student's expected roles to enable their active participation. The generative AI's outputs were transferred to worksheets for students to share in their activities (Figure 2).

#### Instruction

In the workshop, the worksheets were distributed to students online. Initially, the participants were informed that AI generated the research themes on the worksheet based on their research themes. They were then instructed to evaluate the promisingness of the themes and to improve the promising ones. This task was inspired by the concept of *promisingness judgment* (Chen et al., 2015). Promisingness means "deserving of further investment in development" (p.347). Here, we broke promisingness down into *feasibility* and *attractiveness* for the students' easier comprehension and asked them to rate the themes (high, medium, and low) from the two perspectives. Finally, the participants were required to report their final judgment and improvement ideas on the worksheets (Figure 2). The descriptions on the worksheets' memo spaces were later assessed by experts as learning outcomes.

### Analysis

To answer RQ1, two raters with experience in evaluating the quality of interdisciplinary research assessed whether the Ph.D. students could improve the PSOs generated by AI by examining the memo spaces in the 38 worksheets.

Initially, the two raters referred to the conceptualization of *idea improvement* by Hong and Sullivan (2009), which includes idea diversification/exchange and idea (co)elaboration. They then independently assessed whether the

**Table 1**

*Prompts for ChatGPT*

Q1.	What do each researcher's areas of expertise and specific themes have in common?
Q2.	Please consider five moonshot research themes that can only be addressed by combining the areas of expertise and research methods and approaches used by all researchers listed above.
Q3.	Please describe the role expected of each researcher in addressing the above moonshot themes, based on the specific research methods and approaches used, for all of the above researchers. If you have proposed research themes that some researchers will not be involved in, please also tell us which researchers will not be involved.

*Note.* Students' research titles and abstracts (max 250 words) were entered into ChatGPT with the above prompts. Students' tasks were mainly related to the outputs for Q2 and Q3.

**Figure 2**

*Image of Worksheet Used in the Workshop*

Answers for Question 2 & 3. Joint Research Theme Proposal and Each Role						Feasibility (Low, Medium, High)	Attractiveness (Low, Medium, High)	Memo space (enhanced and improved theme/content)
No.	Theme	Content	Role of Researcher 1	Role of Researcher 2	Role of Researcher 3			
1	Enhancing STEM Education through Deep Learning	Develop deep learning-based tools to improve science education by analyzing student attitudes and optimizing teaching methods.	Conduct qualitative research to understand students' attitudes toward science education.	Develop deep learning models for analyzing educational data, focusing on identifying patterns in student attitudes.	Provide insights into the cultural and historical context of education in Japan and how it relates to literature.	Medium	High	Engage in an activity to read a chapter of a recent sci-fi novel ("The Three-Body Problem"), identify its STEM elements, and use Deep Learning to analyze learners' multimodal data.

*Note.* The red text is the output from ChatGPT. We asked students to enter results of their discussion in the blank space below the light-blue highlighted area. Themes from No.2 to No.5 continue below. Expertise of each Researcher 1 to 3 is education, informatics, and literature, respectively.

research themes proposed by the students were improved over those proposed by AI. If at least one of the five themes showed improvement, we considered that the group achieved "theme improvement."

To answer RQ2, we adopted a mixed-methods approach. First, we quantitatively analyzed the collaborative-discourse data by selecting the four high-outcome groups (three or more themes' improvements) and three low-outcome groups (no theme improvement) from the 38 group activities. These seven groups were selected for comparison because they all completed the promisingness judgments for the five themes within the allotted time, all comprised three members each, and they exhibited no significant omissions in their audio recordings of discourse. Table 2 shows each group member's discipline. The data from these groups were transcribed, and the transcripts were coded using a modified version of the coding framework developed by Tong and Chan (2023), which analyzed knowledge-building discourse moves in Knowledge Forum inquiry threads (Table 3). In addition to the three codes used in Tong and Chan (2023), we added the "Initial Impression" code because we recognized the influence of the initial impression of the research theme generated by AI on subsequent discourse. MAXQDA 2022 was used for the coding process. Two independent experts with experience in interdisciplinary research performed the coding. We then employed ENA (e.g., Shaffer, 2017) to visualize the discourse moves leading to different outcomes. Second, a discourse analysis was conducted. We examined the memo spaces of the four high-outcome groups' worksheets to identify the keywords that appeared in their improved PSOs. Subsequently, we focused on the discourse data to pinpoint how these keywords emerged.

**Table 2**

*Each Group Member's Discipline*

Group name	Discipline of Researcher 1	Discipline of Researcher 2	Discipline of Researcher 3
HOG1	Civil Engineering	Medical Science	Literature
HOG2	Education	Psychology	Literature
HOG3	Engineering	Medical Science	Dental Science
HOG4	Environmental Science	Chemistry	Engineering
LOG1		Chemistry	Dental Science
LOG2	Engineering	Engineering	Design Science
LOG3	Mathematics	Engineering	Dental Science

Note. HOG: high-outcome groups; LOG: low-outcome groups

## Results

### Assessment of worksheets (RQ1)

To examine whether the Ph.D. students could improve the PSOs generated by AI, two experts independently assessed the quality of the IDCRTs by reviewing the memo spaces on the worksheets. This assessment yielded 71.1% agreement and a kappa coefficient of .40 (fair level, Landis & Koch, 1977). When their assessments did not match, a third rater decided the final assessment. Consequently, out of the 38 worksheets, at least one improvement was observed in 13 worksheets (34.2%). Among the remaining 25 groups, the students rated all themes proposed by AI as “low” for promisingness in two worksheets. In other 23 groups, although their evaluations resulted in at least one perspective being rated as “Medium” or “High,” their comments were limited to negative evaluations or fragmented words, lacking details for improvements.

### ENA (RQ2)

We compared the four high-outcome groups from the 13 groups and three low-outcome groups from the 25 groups to identify key discourse moves leading groups to their high-quality PSOs. Two independent coders coded two of the seven groups’ discourse data (approximately 30%). The remaining data were coded independently by one researcher. The kappa coefficient ranged from .40 to .86 (fair to substantial level, Landis & Koch, 1977). In coding, distinguishing between “supporting an explanation” and “improving an explanation” was difficult because they both aimed to add another explanation. Therefore, these sub-codes were combined into a single sub-code, “enhancing an explanation.” Table 3 presents the coding table and average number of sub-code occurrences.

**Table 3**

*Coding Table and Average Frequency of Codes for High-Outcome (HO) and Low-Outcome (LO) Groups*

Codes	Sub-Codes	Descriptions	Mean (SD)	
			HO	LO
Questioning	Fact-seeking (FS)	Questions seeking factual information.	12.5 (5.45)	10.33 (4.51)
	Explanation-seeking (ES)	Questions seeking open-ended responses with explanations.	1.5 (2.38)	2.0 (1.0)
	Sustained inquiry (SI)	Asking further questions based on previous ideas and thus deepening the discussion deeper.	9.75 (13.0)	5.0 (4.36)
Theorizing	Simple claim (SC)	Simple (dis)agreement or repetition of a statement.	49.0 (12.68)	64.33 (32.72)
	Proposing an explanation (PE)	Proposing an explanation that explains certain phenomena for the first time	11.75 (2.75)	20.67 (6.43)
	Enhancing an explanation (EE)	Enhancing an already existing explanation through elaborations, specifying details, support, and using new evidence.	60.75 (12.71)	85.0 (39.13)
Community	Connection (CON)	Referencing to their own or others’ statements in different topics or quoting extra sources to advance community understanding.	11.75 (2.36)	2.0 (3.46)
	Synthesizing (SYN)	Synthesizing ideas from previous multiple <u>statements</u> and identifying gaps to engage in a high-level conceptualization.	6.5 (5.2)	1.0 (1.73)
Initial Impression	Promising (PRO)	Stating the promisingness of the proposed theme in the initial stage before conducting deep discussion. (e.g., this sounds interesting!).	2.75 (2.5)	7.0 (6.24)
	Non-promising (NON)	Stating the non-promisingness of the proposed theme in the initial stage before conducting deep discussion. (e.g., this does not make sense).	0.75 (0.96)	6.33 (4.04)
	Ambiguity (AMB)	Stating the ambiguous nature of the theme in the initial stage before conducting deep discussion (e.g., this is ambiguous).	1.50 (1.91)	4.33 (3.79)

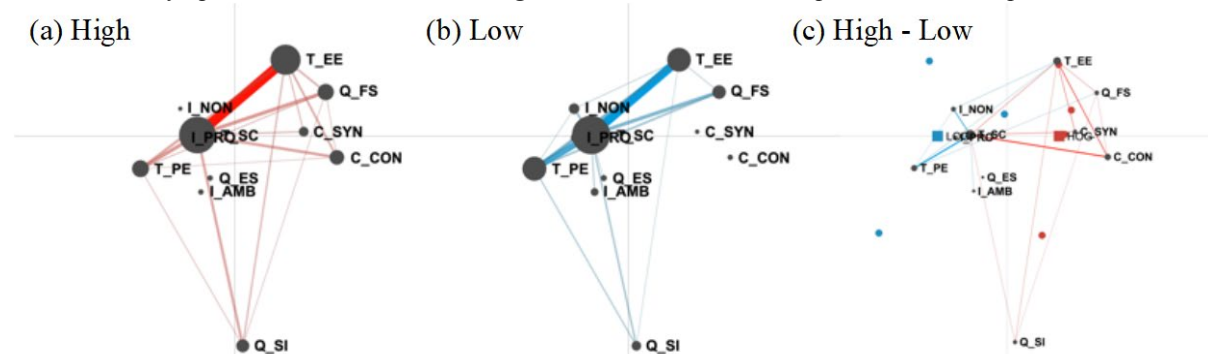


*Note.* Considering the context of coding the discourse instead of thread and of using AI to propose themes, some modifications were introduced into Tong and Chan’s (2023) original framework, which are underlined>.

The coded discourse data were analyzed using ENA (Figure 3). The stanza size was set to four, considering the variances explained. Figures 3a and 3b display the epistemic frames for the high- and low-outcome groups, respectively. They appeared similar. Figure 3c compares them. In the high-outcome groups, the four sub-codes—“simple claims (SC),” “enhancing an explanation (EE),” “connection (CON),” and “synthesizing (SYN)” —are more strongly linked. In the low-outcome groups, however, “non-promising (NON)” and “proposing an explanation (PE)” are more strongly connected with “simple claims (TS).” A Mann-Whitney test showed that the high-outcome groups (Mdn = 0.90, N = 4) were not statistically significantly different at the alpha = 0.05 level; however, they were marginally different from the low-outcome groups (Mdn = -1.28, N = 3) at the alpha = 0.10 level ( $U = 12.00, p = 0.06$ ).

**Figure 3**

*Visualization of Epistemic Network between High- and Low-Outcome Groups and Their Comparison*



*Note.* Q\_FS: fact-seeking, Q\_ES: explanation-seeking, Q\_SI: sustained inquiry, T\_SC: simple claims, T\_PE: proposing explanations, T\_EE: enhancing an explanation, C\_CON: connection, C\_SYN: synthesizing, I\_PRO: promising, I\_NON: non-promising, I\_AMB: ambiguity.

## Discourse analysis (RQ2)

Qualitative analysis identified two types of keywords that appeared in the improved PSOs: disciplinary domain knowledge (e.g., *The Wind Rises* (film title), mRNA) and disciplinary research-design knowledge (e.g., international or historical comparisons). For instance, a literature researcher’s domain knowledge about the film “*The Wind Rises*” facilitated the improvement of a literature-related IDCRT. Additionally, the presence of researcher(s) familiar with feasible research designs contributed to the theme improvement. For instance, in an infectious diseases-related IDCRT, a researcher’s suggestion to compare prevention measures across different eras improved the theme. Notably, these keywords exhibited an irregular pattern of appearing at the beginning, middle, or end of the discourse on the themes.

## Discussion

### RQ1: Can Ph.D. students improve the PSOs generated by AI?

We observed that 34.2% of the groups critically improved the PSOs generated by AI. As mentioned in Theoretical Background, interdisciplinary education practices have had a problem of creating PSOs. Moreover, the implementation of interdisciplinary research entails some challenges owing to the difficulty of boundary crossing. Our study suggests that with generative AI assistance, we can address some of the existing problems (e.g., Repko & Szostak, 2020; Stamp et al., 2015). Meanwhile, we observed no improvement in 25 cases, demonstrating that PSOs generated by AI alone were insufficient for boundary crossing.

### RQ2: What key discourse moves lead groups to construct and improve their PSOs?

Quantitative analysis with marginal significance suggests three key discourse moves: 1) enhancing an explanation; 2) connection; and 3) synthesizing. We also see “focusing only on the initially promising themes” as the fourth key move because persistence with non-promising themes may inhibit fruitful discourse from improving other promising themes (Figure 3c). Our findings are consistent with those of previous knowledge-building studies highlighting the importance of “improvement of explanation,” “connection,” and “synthesizing” in idea improvement (e.g., Tong & Chan, 2023; Yang et al., 2016).

Our qualitative analysis emphasized two important categories of disciplinary knowledge: domain and research-design knowledge. The above four key discourse moves would support the emergence of such knowledge but not necessarily guarantee it. Consequently, integrating the results from quantitative and qualitative analyses, it is essential to intentionally create a learning environment that not only increases the four key discourse moves but also enhances access to the two disciplinary-knowledge categories in the discourse.

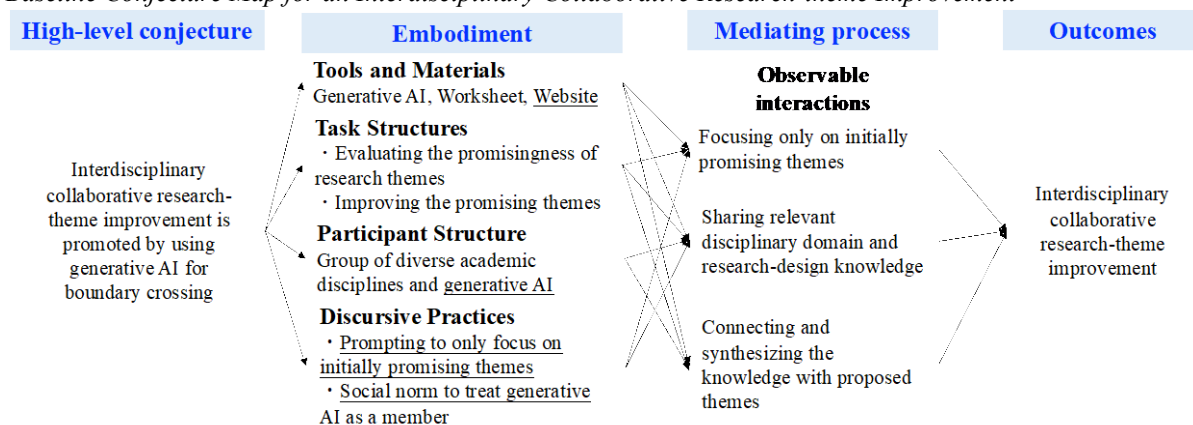
### Proposal of the baseline conjecture map

We constructed the baseline conjecture map with a view for future studies to design interdisciplinary education practices in which students can effectively engage in knowledge-building discourse to improve the IDCRTs (Figure 4). First, the results in this study yielded three potentially key mediating processes: 1) focusing on initially promising themes; 2) sharing relevant disciplinary domain and research-design knowledge; and 3) connecting the knowledge with the proposed themes and synthesizing it. Second, we integrated components of embodiment into the map, effectively realizing those mediating processes. While the workshop in this study already incorporated some of these components, we recognized the need for incremental enhancements. We introduced additional components to the tools and materials, participant structures, and discursive practices to address this need.

In the tools and material, we introduced a “website” because we noticed that the participants were not inclined to access extra resources during their collaboration. In the participant structures, “generative AI” was added. In this study, the use of generative AI was limited to providing students with PSOs for them to start their discussion. AI can be constructively used for students to search and share relevant disciplinary domain and research-design knowledge or connect and summarize their discourse (Scardamalia & Bereiter, 2022). Therefore, we should regard generative AI not only as a tool/material but also as a participant that continuously joins the discourse. In the discursive practices, we then added two components: “prompting to only focus on initially promising themes” and “social norm to treat generative AI as a member.” The former reflects the need to emphasize discarding non-promising themes; the latter requires participants to use generative AI for searching relevant knowledge and connecting/synthesizing their discourse. This map appropriately reflects the design principle we initially conceived: “Interdisciplinary collaborative research-theme improvement is promoted by using generative AI for boundary crossing.” This map should be used as a baseline for future study, and it will require validation and updating through iterative processes.

**Figure 4**

*Baseline Conjecture Map for an Interdisciplinary Collaborative Research-theme Improvement*



*Note.* Additional components to the embodiment are underlined.

### Conclusion

Our study suggests that by using generative AI as a broker for boundary crossing in an interdisciplinary education context, we can solve the problem of creating PSOs. We also identified key discourse moves that were likely to contribute to theme improvement, which we incorporated into a baseline conjecture map (Figure 4).

Our innovative practice contributes to the interdisciplinary education by offering a novel approach that enables smoother coordination. This approach applies to other interdisciplinary education practices because the primary resources required are students’ research information and the prompts in Table 3. Additionally, generative AI can offer insights for other pedagogical practices, especially collaborative learning (Cress & Kimmerle, 2023), to create objects considering group members’ interests or backgrounds. Finally, beyond the

educational contexts, practices using generative AI for boundary crossing is highly anticipated to contribute to the evolution of future interdisciplinary research and other practices that contain multiple activity systems.

## References

- Aboelela, S. W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S. A., Gebbie, K. M. (2007). Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health Services Research, 42*(1), 329–346.
- Akkerman, S. F., Admiraal, W., Simons, R. J., & Niessen, T. (2006). Considering diversity: Multivoicedness in international academic collaboration. *Culture & Psychology, 12*, 461–485.
- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169.
- Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer Supported Collaborative Learning, 10*, 345–366.
- Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*. Advance online publication. <https://doi.org/10.1007/s11412-023-09409-w>
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133–156.
- Engeström, Y. (2022). Learning in activity. In K. R. Sawyer (Ed), *The cambridge handbook of the learning sciences* (pp. 134–155). Cambridge University Press.
- Engeström, Y., & Sannino, A. (2021). From mediated actions to heterogenous coalitions: Four generations of activity-theoretical studies of work and learning. *Mind, Culture, and Activity, 28*(1), 4-23.
- Fischer, A. R. H., Tobi, H., & Ronteltap, A. (2011). When natural met social: A review of collaboration between the natural and social sciences. *Interdisciplinary Science Reviews, 36*(4), 341–358.
- Fortuin, K. P. J., Gulikers, J. T. M., Uiterweer, N. C. P., Oonk, C., & Tho, C. W. S. (2023). Developing a boundary crossing learning trajectory: Supporting engineering students to collaborate and co-create across disciplinary, cultural and professional practices. *European Journal of Engineering Education*. Advance online publication. <https://doi.org/10.1080/03043797.2023.2219234>
- Hong, HY., & Sullivan, F.R. (2009). Towards an idea-centered, principle-based design approach to support learning as knowledge creation. *Educational Technology Research and Development, 57*, 613–627.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
- Paavola, S., Lipponen, L., & Hakkaraine, K. (2005). Models of innovative knowledge communities and threemetaphors of learning. *Review of Educational Research, 74* (4), 557-576.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. <https://openai.com/research/language-unsupervised>
- Repko, A. F., & Szostak, R. (2020). *Interdisciplinary research process and theory (fourth edition)*. SAGE Publications, Inc.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36.
- Scardamalia, M., & Bereiter, C. (2022). Knowledge building and knowledge creation. In Sawyer, K. R. (Ed), *The cambridge handbook of the learning science third edition* (pp.385-405). Cambridge University Press.
- Seilstad, B. (2018). Partially shared objects and the (Elusive) potential of expansive learning: The case of “jury” in community-based citizenship classes for Nepali-speaking Bhutanese refugee elders. *Mind, Culture, and Activity, 25*(4), 308–323.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.
- Stamp, N., Tan-Wilson, A., & Silva, A. (2015). Preparing graduate students and undergraduates for interdisciplinary research. *BioScience, 65*(4), 431–439.
- Tobi, H., & Kampen, J. K. (2018). Research design: The methodology for interdisciplinary research framework. *Quality & Quantity: International Journal of Methodology, 52*(3), 1209–1225.
- Tong, Y., & Chan, C. K. K. (2023). Promoting knowledge building through meta-discourse and epistemic discourse understanding. *International Journal of Computer-Supported Collaborative Learning, 18*, 353–391.
- Yang, Y., van Aalst, J., Chan, C. K. K., & Tian, W. (2016). Reflective assessment in knowledge building by students with low academic achievement. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 281–311.

# The Role of Individual Preparation on Coordination in Computer-Supported Collaborative Learning: A Neuroscience Perspective on Learners' Inter-Brain Synchronization

Wenli Chen, Qianru Lyu, Mei-Yee Mavis Ho, Jessica Tan, Wei-Peng Teo, Siew Cheng Aileen Chai, Guo Su  
wenli.chen@nie.edu.sg, nie20.lq@e.ntu.edu.sg, mavis.ho@nie.edu.sg, jessica.tan@nie.edu.sg,  
weipeng.teo@nie.edu.sg, aileenchai7@gmail.com, nie22.sg4560@e.ntu.edu.sg  
National Institute of Education, Nanyang Technological University, Singapore

**Abstract:** Individual preparation (IP) is often applied to support collaborative learning. However, there exist mixed results of this pedagogical approach's effectiveness. This study aims to expand the current understanding of how IP influences social coordination during collaboration. A total of 78 university students (male = 30, female = 48) aged between 21 to 40 years old collaborated in dyads in this study. Functional infrared spectroscopy (fNIRS) was used to measure brain activity in the orbitofrontal cortex (OFC) for every individual during two conditions: immediate collaboration without IP (control condition) and IP before collaboration (experimental condition). Inter-brain synchrony (IBS) between dyads was derived and compared between two conditions. Results revealed that significantly higher levels of IBS could be observed in the control-experimental comparisons. These findings suggest that introducing individual preparation can facilitate social coordination during subsequent collaborative learning.

## Introduction

Collaborative learning allows students to expand knowledge and co-construct new knowledge through productive peer interactions (Janssen & Kirschner, 2020). As technologies are pervasively integrated into daily learning environments, students co-construct knowledge in a complex dynamic technology-supported environment (Stahl, 2010; Authors, 2011). However, technologies themselves do not promise productive collaborative learning and students still face challenges with engaging in high quality knowledge exchange, negotiation and co-creation (Barron, 2003). Providing pedagogical support for productive collaborative learning remains a key question for the field of Computer-Supported Collaborative Learning (CSCL).

One of the pedagogical designs of collaborative learning is individual preparation before collaboration, inspired by the Preparation for Future Learning paradigm (PFL; Bransford & Schwartz, 1999). Individual preparation (IP) before collaboration refers to "providing learners with time to perform activities directed at processing the instructional material on their own before the collaboration" (Mende et al., 2021, p30). This preparation activity is introduced to CSCL environments to promote students' readiness for learning in the subsequent task. This pedagogical design was implemented in multiple lines of research such as collaboration scripts and group ideation studies. Though being applied in various CSCL contexts, mixed results were reported in empirical studies. It was found that as students became more cognitively prepared after IP was provided, they were also likely to encounter coordination difficulties during the subsequent collaboration (Mende et al., 2021). Based on the existing studies, it remains unclear whether and how IP brings coordination challenges to students' collaborative learning, which poses questions for optimizing its design and implementation in CSCL activities.

When existing IP studies mostly focused on collaboration artefacts quality and discussions as evidence of coordination, the physiological dimension of collaborative learning process was barely discussed. In recent years, the development of neuroimaging techniques has allowed for a simultaneous recording of brain activity from two or more interacting individuals, otherwise known as hyperscanning (Tan et al., 2023). This method supports a shift towards a multi-brain paradigm and has been adopted in a myriad of studies to determine the influence of demographic characteristics and situational settings on collaborative outcomes (Liu & Pelowski, 2014). With hyperscanning, researchers have also been able to use Inter-Brain Synchrony (IBS) as valuable evidence for understanding social coordination and interaction (Czeszumski et al., 2020). A fine-grained examination of IP activity can therefore benefit from a close examination of students' IBS under different conditions: immediate collaboration and IP before collaboration. From this perspective, this study aims to expand the current understanding of IP by examining its impact on students' social interaction and coordination informed by their IBS.

## Literature review

### Individual preparation before collaborative learning

IP has been adopted in various collaborative learning designs (e.g., Chen et al, 2021). Though being widely applied in collaborative learning design, whether the IP contributes to students' collaborative learning remains unclear. Mixed results were found in existing literature. Some studies reported that IP before collaboration contributed to collaboration outcomes (e.g., Farrokhnia et al., 2019, Lyu et al., 2023), knowledge learning, socially shared regulation and the quality of knowledge co-construction discourses). With the benefits being said, researchers also identified the coordination challenges brought by the IP activities. For example, Tsovaltzi et al. (2015) found that students became less open to alternative perspectives after the provision of IP before collaborative argumentation activity. Compared to immediate collaboration where students partake in a continuous exchange of ideas, the IP activity may lead to distinct ideas from different learners that require substantial efforts to coordinate with each other, integrate different ideas and reach consensus (Mende et al., 2021).

To summarize, though IP allows for cognitive readiness for collaboration, researchers suggest the need to consider its coordination challenges. What remains unclear however, is whether and how IP poses coordination challenges for the subsequent collaboration. The current limited understanding of IP can benefit from a close examination of students' social coordination processes in two collaborative conditions: with and without IP. Though most CSCL investigated students' demonstrated behavior (e.g., verbal discussions; online posts; online chat), researchers found value in examining the physiological responses during social interaction and coordination. Recent advancements in hyperscanning techniques have revealed the close relationship between students' IBS during social coordination and interaction by allowing simultaneous examination of multiple brains using neuroimaging procedures.

### Inter-brain synchrony (IBS) for social interaction

Advancements in neuroscience technologies have made it feasible to adopt a multi-brain framework in naturalistic settings. With this approach, greater emphasis is placed on the interactions between multiple brains instead of single-brain activity (Schilbach et al., 2013). IBS refers to the strength of synchronizations between neural regions of two or more interacting brains (Li et al., 2021) and is commonly studied in hyperscanning studies. Existing literature has highlighted the importance of IBS for elements that drive collaborative processes. This includes joint attention, interpersonal communication and coordination, cooperation, and decision-making (Czeszumski et al., 2020). While the definitive role of each brain region remains a puzzle, previous studies investigating the neural mechanisms of collaboration have managed to identify several neural regions that are crucial for cognition and social interaction. Of which, the findings seem to display some overlap in illustrating the emergence of IBS in the frontal area of the brain (Xue et al., 2018). One example is in the orbitofrontal cortex (OFC), which has been shown to be vital for the regulation of appropriate social behavior (Jonker et al., 2015).

In the context of education, neural substrates-based evidence presents value in the understanding of peer interaction practices in the classroom (Xue et al., 2018; Tan et al., 2023). For instance, Dikker et al (2017) examined IBS at varying levels (group synchrony, student-group synchrony, student-student synchrony) and contrasted these findings against the students' self-reported ratings of overall lesson enjoyment. Results demonstrated a strong positive correlation between the two measures, suggesting the importance of social dynamics in supporting collaborative learning. With information on IBS, researchers are also able to gain clarity on the efficacies of various collaborative learning strategies (Arioli & Canessa, 2019). In the same exploratory study by Dikker et al (2017), IBS levels under different teaching conditions were compared against the students' self-reported scores of interpersonal closeness with their classmates and teachers. From these results, the authors were not only able to identify the students' preferred method of learning, but also draw associations between social closeness, classroom engagement and effective teaching styles.

### Research questions

In summary, there is much value in integrating the neuroscience and behavioral science fields when attempting to understand the underlying processes of collaborative learning. Physiological markers represented by brain activity help to supplement the findings from conventional demonstrated behavioral. Furthermore, existing literature has displayed that IBS is closely related to students' coordination of social behaviors in collaborative learning settings. Thus, this study aims to investigate the effect of IP on collaborative learning at brain level to reach a more comprehensive understanding. The research questions is: What is the impact of IP on IBS levels during collaboration?

## Methods

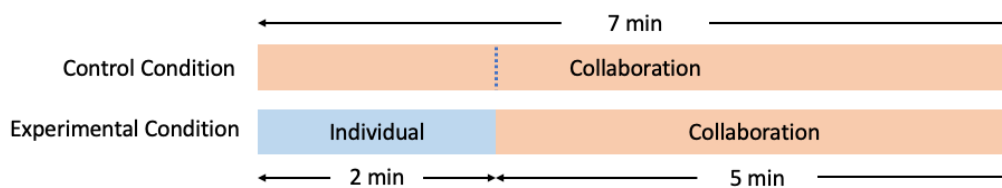
### Participants and learning context

This study was conducted at a university in Singapore. Appropriate ethical approval was granted by the university's Institutional Review Board, and all experimental procedures followed the relevant guidelines and regulations. A total of 78 participants (male = 30, female = 48) aged between 21 to 40 years old were involved in the study. They were graduate students from different disciplines, such as Education, Engineering, Mathematics, etc. All participants were right-handed with normal or corrected vision. No restrictions on gender, age, or academic backgrounds were imposed when participants were assigned into their respective dyads. However, to minimize the possible influence of interpersonal closeness on collaboration outcomes, the authors of this study took measures to ensure that participants were paired up with someone unfamiliar to them.

The experiment occurred in a laboratory with two computers and an fNIRS-system set-up that could capture dyads' brain data throughout. A crossover design was introduced, where participants sat separately as they worked on two product design ideation tasks in a single session. These tasks were assigned in a random sequence but were of comparable difficulty levels. They were required to design simple everyday items such as lunchboxes and schoolbags. Guiding prompts were included in the task interface to stimulate the provision of more details, and 7-minutes was given to complete each condition (see Figure 1). In the control condition, each dyad collaborated for the whole 7 minutes. In the experimental condition, every dyad spent 2 minutes on IP first and another 5 minutes on collaboration. In the 2-minute IP phase, participants could only view their own onscreen behavior and were not allowed to converse with one another.

**Figure 1**

*The Experimental Design*

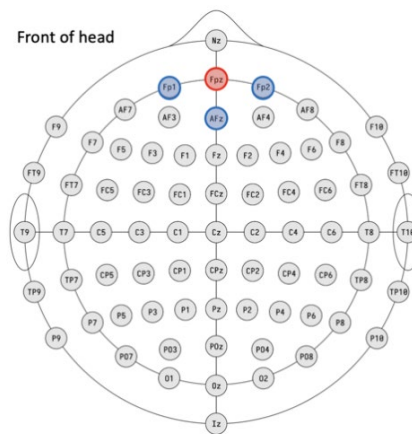


### Data collection

Changes in hemodynamic activity within the brain provide valuable insights on brain activation patterns during individual and collaborative work. This study employed an fNIRS-based hyperscanning approach where hemodynamic responses over the OFC were recorded with a sampling rate of 10Hz (Oxyton MkIII, ArtiniseMedical Systems, The Netherlands). Eight optodes (2 transmitters x 6 receivers) were used for each dyad, operating at 760nm and 850nm wavelengths to detect relative changes in hemoglobin levels within the brain. Transmitters and receivers were split equally between pairs and placed according to the international 10-10 EEG placement system, resulting in three measurement channels for each individual (Figure 2). Transmitter-receiver pairs were located 3cm apart. Individual differential path-length factors (DPF) were also calculated using the system's inbuilt formula to account for age differences.

**Figure 2**

*Transmitter (in Red) and Receiver (in Blue) Optode Placements over the OFC Region for Each Participant. Channels are Formed between Each Transmitter and Receiver Pairing, Thus Resulting in Three Measurement Channels.*

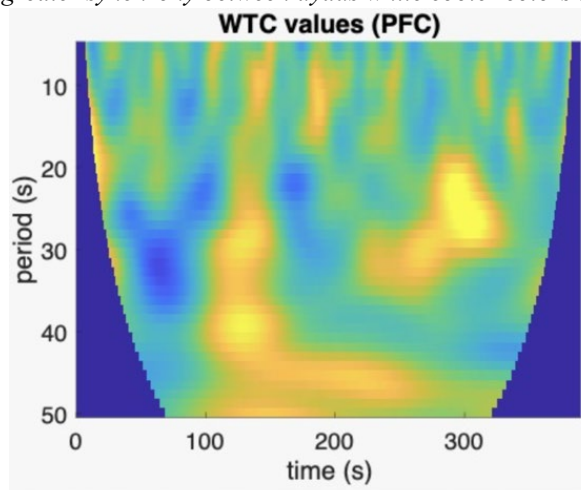


### Data analysis method

Raw data was recorded and stored in a laptop with the Oxysoft software version 3.2.72 (Artinis Medical Systems, The Netherlands). With reference to the event markers that indicated the start of each trial, data files were then cropped and extracted into three 7-minute recordings to capture the entire course of each condition. This study focused on the oxygenated hemoglobin (O<sub>2</sub>Hb) data due to its high sensitivity towards alternations in cerebral blood flow (Hoshi Y., 2003). Prior studies have also expressed the benefits of using wavelet transform coherence (WTC) analysis in hyperscanning research to quantify neural synchrony (Grinsted et al., 2004). Hence, a customized MATLAB-based script encompassing a preprocessing stream and WTC calculations was subsequently employed for IBS analysis between dyads. From which, WTC values within a task-related frequency band of 0.02Hz to 0.2 Hz (Jiang et al., 2015) were selected for comparisons between conditions. Figure 3 displays an example of the phases of coherence that appear within the dyads' OFC during task, derived from the time-based and task-related WTC.

**Figure 3**

*Correlation Heat Map Resulting from a Series of Paired Samples T-Tests Using Time-Based and Task-Related WTC Values. Warmer colors indicate periods of greater synchrony between dyads while cooler colors indicate lower synchrony.*



### Results

Table 3 displays the mean O<sub>2</sub>Hb values of the OFC region and their standard deviations (SD) during the control and experimental conditions respectively. A close examination of the control condition against the experimental helps to confirm the presence of IBS during IP. As such, time-based values yielded by WTC analysis were averaged across all three channels and a repeated measures ANOVA was conducted. Significance level was set at  $p < .05$ .

The results displayed a significant difference in IBS between the control and experimental conditions ( $p = .034$ ), indicating that students presented higher inter-brain synchronization when they had individual preparation before collaboration than without individual preparation.

**Table 3**  
*Descriptive Analysis Results of O<sub>2</sub>Hb Values in the Two Conditions*

Conditions	n	Mean	SD	t	p value
Control (immediate collaboration)	39	0.3100	0.0322	-2.41	.034
Experimental (Individual preparation before collaboration)		0.3297	0.0366		

## Discussion

Collaboration has been found to involve a series of executive and mentalizing functions, where both contribute towards the regulation of social behaviors (Decety et al., 2004). As previously identified by existing literature, the OFC region plays a critical role in one's ability to manage these functions in a social setting (Jonker et al., 2015). Building on scripted collaborative learning studies, this study examined the effect of IP on the subsequent collaboration with neural evidence.

The results from the control-experimental comparison indicated a significant increase in IBS within the OFC when IP was provided, signaling longer periods of similar cognitive loads between participants in each dyad. The increased IBS and the effectiveness of collaboration are regulated by collective knowledge structures by which the collaborative task is provided (Van den Bossche et al., 2006). Greater IBS also suggests that the experimental condition induced an increase in participants' effort to maintain a socially acceptable presentation of self and an understanding of their peers. This finding is interesting as previous studies highlighted the coordination challenges brought by the IP activity (Mende et al., 2021). Transactive activities or transactive costs refer to the cognitive load incurred by an individual as they initiate and maintain communication with their counterparts (Kirschner & Zambrano, 2018). Compared to immediate collaboration which allows for ongoing communication, it was suggested that while IP may present information processing advantages, it also reduces coordination opportunities. This in turn creates higher transaction costs and imposes higher collaborative cognitive loads on the individuals (Mende et al., 2021; Janssen et al., 2010; Kirschner et al., 2018). The increased transaction cost, at an extraneous level, would thus make it challenging for the collaborating individuals to understand and coordinate with each other (Kirschner et al., 2018). In this study however, students in the IP before collaboration condition presented higher levels of synchronization in terms of brain activity than the immediate collaboration condition. Hence, based on the brain data, instead of facing difficulties in reaching a common understanding, the students managed to achieve higher levels of social coordination during the collaborative task with IP. This result resonates with some existing findings that reported higher quality collaboration discourses (Gijlers et al., 2013; Authors, 2022; Authors, 2023; Farrokhnia et al., 2019).

There can be several reasons for this finding. The provision of IP gave students a collective knowledge structure that enabled them to maintain similar cognitive wavelengths for a longer period. As a result, this could have allowed them to find it easier to understand each other's perspectives, share information and coordinate their actions. In other words, it is likely that the IP activity provided in this study enabled students to capitalize on the advantages of information processing, notwithstanding the coordination challenges. This thus facilitated a comprehensive understanding of the task, where students could utilize their cognitive capacity fully while the influence of their peers' external expressions were mitigated (Cress & Kimmerle, 2008). Consequently, a better understanding of the task allowed students to be more "ready" for joining discussions. It is also worth noting that during the IP stage, both individuals worked on one shared working space, providing them with group awareness support. Group awareness support refers to students' knowledge about peers' knowledge, understanding, opinions (Janssen & Bodemer, 2013). When present during IP, group awareness was found to benefit students' collaboration and coordination (Engleman et al., 2009) as it reduced the transactive working memory (Wegner, 1986). Researchers who designed digital tools to promote group awareness have reported its benefits for the maintenance of students' shared understanding. Therefore, it is possible that with the provision of the shared digital workspace, students' coordination challenges were alleviated to some extent and information processing advantages were heightened by the interface. Building on this research, more empirical studies are expected to examine the role of group awareness support during IP on students' collaboration.

This study had its limitations. With a lab-based experiment setting, the CSCL environment in this study may be different from that in authentic classrooms. In addition, the various factors (e.g., specific tasks, learners'



profiles, genders) can play roles in the way students interacted and coordinated with each other. In addition, it remains an open question to what extent this finding informs the collaborative learning context with longer collaboration time duration and higher task complexity level. Future studies are expected to conduct well controlled experiments to validate the findings. Specifically, the relationship between IBS results and collaborative outcomes should be further explored.

## Conclusion

To summarize, this study uses neural evidence to highlight the importance of social and cognitive coordination challenges in managing collaborative outcomes. This potentially adds value to the existing analysis of students' discussions, collaboration outcomes and learning gains (Farrokhnia et al., 2019; Authors, 2023). While the lack of differences in IBS patterns between collaboration and individual learning were somewhat surprising, the significant IBS across collaboration conditions resonate with the prior IBS studies in social interaction settings (e.g., Hasson et al., 2012). These findings provide physiological evidence for the growing body of collaboration scripts literature, uncovering the benefits of IP for subsequent future collaboration. Further, the results can also aid in fine-tuning the design of collaboration scripts and optimizing the benefit of IP in CSCL context.

## References

- Arioli, M., & Canessa, N. (2019). Neural processing of social interaction: Coordinate-based meta-analytic evidence from human neuroimaging studies. *Human Brain Mapping, 40*(13), 3712-3737.
- Barron, B. (2003). When smart groups fail. *The journal of the learning sciences, 12*(3), 307-359.
- Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61-100.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences, 2*(2), 141-178.
- Chang, D., Hwang, G. J., Chang, S. C., & Wang, S. Y. (2021). Promoting students' cross-disciplinary performance and higher order thinking: A peer assessment-facilitated STEM approach in a mathematics course. *Educational Technology Research and Development, 69*, 3281-3306.
- Chang, S. C., Hsu, T. C., & Jong, M. S. Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers & Education, 146*, 103758.
- Chen, W., Looi, C.-K., & Tan, S. (2010). What do students do in a F2F CSCL classroom? The optimization of multiple communications modes. *Computers & Education, 55*(3), 1159-1170. <https://doi.org/10.1016/j.compedu.2010.05.013>
- Chen, W., Lyu, Q., & Su, J. (2022). The effect of individual ideation before discussion on computer supported collaborative argumentation in a primary classroom. In Sridhar Iyer, J.-L. Shih, W. Chen, & Mas Nida Md Khambari (Eds.), *Proceedings of the 30th International Conference on Computers in Education Conference* (Vol. 1, pp. 182- 191). Asia-Pacific Society for Computers in Education. <https://icce2022.apsce.net/proceedings/volume1>
- Chen, W., Tan, S.H. J., & Pi, Z. (2021). The spiral model of collaborative knowledge improvement: an exploratory study of a networked collaborative classroom. *International Journal of Computer-Supported Collaborative Learning, 16*(1), 7-35. <https://doi.org/10.1007/s11412-021-09338-6>
- Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning, 3*(2), 105-122.
- Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z.Z., Konig, P., (2020). Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci. 14*, 39. <https://doi.org/10.3389/fnhum.2020.00039>.
- Dennen, V. P., Hall, B. M., & Hedquist, A. (2023). A systematic review of research on intersubjectivity in online learning: Illuminating Opportunities for Cohesion and Mutual Understanding in the Research Conversation. *Online Learning, 27*(1).
- Dikker, Suzanne & Wan, Lu & Davidesco, Ido & Kaggen, Lisa & Oostrik, Matthias & McClintock, James & Rowland, Jess & Michalareas, Georgios & Van Bavel, Jay & Ding, Mingzhou & Poeppel, David. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology. 27*. 10.1016/j.cub.2017.04.002.

- Engelmann, T., Tergan, S.-O., & Hesse, F. W. (2009). Evoking knowledge and information awareness for enhancing computer-supported collaborative problem solving. *The Journal of Experimental Education*, 78(2), 268-290.
- Farrokhnia, M., Pijera-Díaz, H. J., Noroozi, O., & Hatami, J. (2019). Computer-supported collaborative concept mapping: The effects of different instructional designs on conceptual understanding and knowledge co-construction. *Computers & Education*, 142. <https://doi.org/10.1016/j.compedu.2019.103640>
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist*, 48(1), 56-66. <https://doi.org/10.1080/00461520.2012.748005>
- Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process Geophys.* 11, 561–566. doi: 10.5194/npg-11-561-2004
- Gijlers, H., Weinberger, A., van Dijk, A. M., Bollen, L., & van Joolingen, W. (2013). Collaborative drawing on a shared digital canvas in elementary science education: The effects of script and task awareness support. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 427-453.
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007
- Hoshi, Y. (2003). Functional near-infrared optical imaging: utility and limitations in human brain mapping. *Psychophysiology* 40, 511–520. doi: 10.1111/1469-8986.00053
- Hsiao, H. S., Chen, J. C., Chen, J. H., Zeng, Y. T., & Chung, G. H. (2022). An assessment of Junior High School students' knowledge, creativity, and hands-on performance using PBL via cognitive–affective interaction model to achieve STEAM. *Sustainability*, 14(9), 5582.
- Isöhätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, 81, 11-24. <https://doi.org/10.1016/j.ijer.2016.10.006>
- Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist*, 48(1), 40-55.
- Janssen, J., & Kirschner, P. A. (2020). Applying collaborative cognitive load theory to computer-supported collaborative learning: towards a research agenda. *Educational Technology Research and Development*, 68(2), 783-805. <https://doi.org/10.1007/s11423-019-09729-5>
- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, 32(45), 16064-16069
- Jonker, F. A., Jonker, C., Scheltens, P., & Scherder, E. J. (2015). The role of the orbitofrontal cortex in cognition and behavior. *Reviews in the Neurosciences*, 26(1), 1-11.
- Kawasaki, M., Yamada, Y., Ushiku, Y. et al. Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci Rep* 3, 1692 (2013). <https://doi.org/10.1038/srep01692>
- Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, 13(2), 213-233.
- Korde, R., & Paulus, P. B. (2017). Alternating individual and group idea generation: Finding the elusive synergy. *Journal of Experimental Social Psychology*, 70, 177-190. <https://doi.org/10.1016/j.jesp.2016.11.002>
- Li, R., Mayseless, N., Balters, S., & Reiss, A. L. (2021). Dynamic inter-brain synchrony in real-life inter-personal cooperation: A functional near-infrared spectroscopy hyperscanning study. *NeuroImage*, 238, 118263.
- Liu, T., and Pelowski, M. (2014). A new research trend in social neuroscience: towards an interactive-brain neuroscience. *Psych. J.* 3, 177–188. doi: 10.1002/pchj.56
- Lu, K., Xue, H., Nozawa, T., & Hao, N. (2019). Cooperation makes a group be more creative. *Cerebral Cortex*, 29(8), 3457-3470.
- Lyu, Q., Chen, W., Su, J., & Heng, K. H. (2023). Collaborate like expert designers: An exploratory study of the role of individual preparation activity on students' collaborative learning. *The Internet and Higher Education*, 59, 100920. <https://doi.org/https://doi.org/10.1016/j.iheduc.2023.100920>
- Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist*, 56(1), 29-53. <https://doi.org/10.1080/00461520.2020.1828086>
- Quedenfeld H. B. (2022). Physiological Response Data in Education: Could this be the future for analyzing social interactions and learning? A Systematic Literature Review. [Master's thesis, Tampere University] <https://trepo.tuni.fi/bitstream/handle/10024/138086/QuedenfeldHarry.pdf?sequence=2>

- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/s0140525x12000660
- Tan, S. J., Wong, J. N., & Teo, W. P. (2023). Is neuroimaging ready for the classroom? A systematic review of hyperscanning studies in learning. *NeuroImage*, 120367.
- Tsovaltzi, D., Judele, R., Puhl, T., & Weinberger, A. (2015). Scripts, individual preparation and group awareness support in the service of learning in Facebook: How does CSCL compare to social networking sites? *Computers in Human Behavior*, 53, 577-592. <https://doi.org/10.1016/j.chb.2015.04.067>
- Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185-208). Springer.
- Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *Neuroimage*, 172, 527-537. <https://doi.org/10.1016/j.neuroimage.2018.02.007>
- Van den Bossche, P., Gijssels, W., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. *Small Group Research*, 37, 490–521. <https://doi.org/10.1177/1046496406292938>.

### **Acknowledgement**

This study was funded by National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore (RS 1/22 CWL) and administered by NIE, NTU. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NIE, NTU. The study was approved by the NTU IRB (IRB approval number: IRB-2022-210).

# Identification of Different Patterns in Solving Collaborative Jigsaw Puzzle Tasks Using Hidden Markov Models

Qizhi Xu, University of Science and Technology of China, fred@ustc.edu.cn  
Zongting Ge, University of Science and Technology of China, gezongting@mail.ustc.edu.cn  
Mengxiao Zhu, University of Science and Technology of China, mxzhu@ustc.edu.cn  
Ziwei Dou, University of Science and Technology of China, douziwei@mail.ustc.edu.cn  
Jing Wang, Anhui Xiyue Educational Technology, jingwang@huohuaschool.com

**Abstract:** To assess learner's collaborative problem-solving (CPS) skills, it is necessary to identify their behavior patterns in completing collaborative tasks. These patterns are concealed within the collaboration and problem-solving processes, and can only be analyzed through modeling of the detailed process data. This study utilized Hidden Markov Models (HMMs) to analyze 15,767 instances of process data generated by 63 students in an online collaborative jigsaw puzzle task. With the goal of comparing low and high performing teams, we identified three hidden states in the low-scoring group, including *Warm-up*, *Action*, and *Strategy Execution and Validation* states, and four hidden states for the high-scoring group, including the three states for the low-scoring group plus a new *Strategy Optimization* state. Furthermore, the results showed that even though both groups initiated the task with the *Warm-up* state, the high-scoring group exhibited more complex transitions between states than the low-scoring group.

## Background

As the complexity of work increases and specialization continues to refine, the importance of Collaborative Problem-Solving (CPS) is considered one of the key competencies of 21st-century talents and is becoming more evident in the education and labor sectors (OECD, 2017b). Thanks to the advances in communication and artificial intelligence technologies, the prospects of automated intervention and evaluation of students' CPS are promising. This has become one of the most cutting-edge research areas in computer-supported collaborative learning and gained significant attention (Fiore et al., 2018).

To actualize the goal of automated intervention and assessment, computers must have the capacity to identify learner states and behavior patterns, especially those associated with CPS performance. Past research has found that if students cannot successfully address the difficulties and challenges encountered during collaborative problem-solving processes, they often exhibit poor team performance and CPS skills (Earle-Randell et al., 2023). Identifying these difficulties and designing personalized interventions (Zhang et al., 2020) is seen as a feasible method to help students improve their CPS skills. It is thus a crucial prerequisite to understand what behavior patterns are associated with poorer team performance before one can design personalized interventions.

Accomplishing this task cannot be achieved solely by analyzing the students' outcomes, because capabilities for CPS typically manifest themselves in the process of cooperation and problem-solving. Researchers equipped with advanced information technologies now have the opportunity to access process data with detailed student behaviors that go beyond the final responses (Zhu et al., 2016). Analysis of the process data makes it possible to get crucial insights into patterns of student behavior, and gains better understanding of their cognitive and collaborative process.

However, the data generated during the CPS process is particularly intricate, encompassing not only human-computer interactions but also inter-member exchanges within teams (Yuan et al., 2019). Moreover, this data exhibits a high level of granularity (Andrews-Todd & Kerr, 2019) in addition to its considerable volume (Kerr & Chung, 2012). Modeling these data to find meaningful behavior patterns has emerged as an intriguing problem. This study uses the Hidden Markov Model (HMM) to analyze process data of high-school student teams with different performance levels working on an online jigsaw puzzle task, and aims to identify and better understand patterns of student behavior in online CPS. The findings of this study extend our understanding of the differences in behavioral patterns of teams with different CPS skills and provide a basis for designing personalized interventions to enhance student CPS skills.

## Literature review

Process data refers to a series of actions recorded during students' learning processes, which can be gathered from computer system log files or from devices such as video recorders, audio recorders, or eye trackers. Process data analysis provides the opportunity to understand how students plan, evaluate, and make decisions to achieve certain problem-solving goals (Xiao et al., 2021). Researchers are continually exploring new methods for processing data generated from students' problem-solving processes, such as cluster analysis (Klingler et al., 2016), sequence mining (Hao & Shu, 2015), network analysis (Zhu et al., 2016)

During the analysis of data generated in the process of CPS, some researchers have utilized similar analytical methods to comprehend strategies adopted by students. Chang et al. (2017) used a lag sequential analysis technique to analyze multiple data sources including group discussions, problem-solving activities, etc. The results showed that underperforming groups could only apply an intuitive trial-and-error strategy and were unable to transform their discussion into an executable plan. Stadler et al. (2019) used N-Gram to analyze the sequence of behaviors and search for differences in the behavior observed within those students who applied the VOTAT strategy (Greiff et al., 2015) and successfully solved the task and those who applied the strategy but still failed to solve the task.

However, in contrast to the singular problem-solving strategies identified in these studies, students may employ composite strategies, suggesting a potential shift from one strategy to another or an application of multiple strategies concurrently to enhance their efficiency. The application of Hidden Markov Models (HMM) in such a field has shown good potential in identifying the states or stages of students and comparing students' learning strategies. HMMs are double stochastic processes, to observe a hidden stochastic process by another observable stochastic process (Rabiner & Juang, 1986). Pan et al. (2020) identified three emotional states from the interaction data of 8537 students via HMM, including flow, anxiety, and boredom. Lopez-Pernas and Saqr (2021) found meaningful behavior pattern differences between high- and low-performance students by using HMM to analyze the data from students' online programming tasks.

Based on this, HMM has the potential to analyze CPS process data to gain new insights into students' behavioral patterns and their relations to CPS performance. A few researchers have already attempted to analyze the conversational data and operational data using HMM. Earle-Randell et al. (2023) analyzed 7594 excerpts of discourse from 44 high school students in a paired programming task by using Zakaria's classification method and HMM. They identified six conversational states including socialization, confusion, exploratory talk, frustrated coordination, directive & disagreement, and disagree & self-explanation. Rodríguez and Boyer (2015) analyzed the operational data generated in a visual programming task by 30 students allocated to a collaborative work group and an independent work group. The application of HMM found that students working collaboratively are more likely to stick to a single problem-solving pattern than individual students.

However, focusing solely on either conversational data or operational data may result in the loss of a substantial amount of valuable information. Ouyang et al. (2023) attempted to bridge this gap by synthesizing two types of process data generated by 13 teams of 24 students in a concept map building task. Three types of transformation patterns were identified, described as behaviorally oriented transformations, communication-behavior-coordination transformations, and communicationally oriented transformations. Nevertheless, students' performance in the study heavily relied on participants' prior domain knowledge which has a possible impact on students' behavior and performance (Funke & Frensch, 2017).

Therefore, we developed a CPS task for high-school students, which can be considered domain knowledge independent, to collect their operational and conversational data. HMM was employed with the expectation of uncovering behavioral pattern differences among teams exhibiting disparate performances. We aimed to address the following two research questions:

RQ1: What are the hidden states present during students' online collaborative problem-solving processes, and how do these states reflect their problem-solving strategies?

RQ2: What differences in behavioral patterns can be observed among teams with different performances?

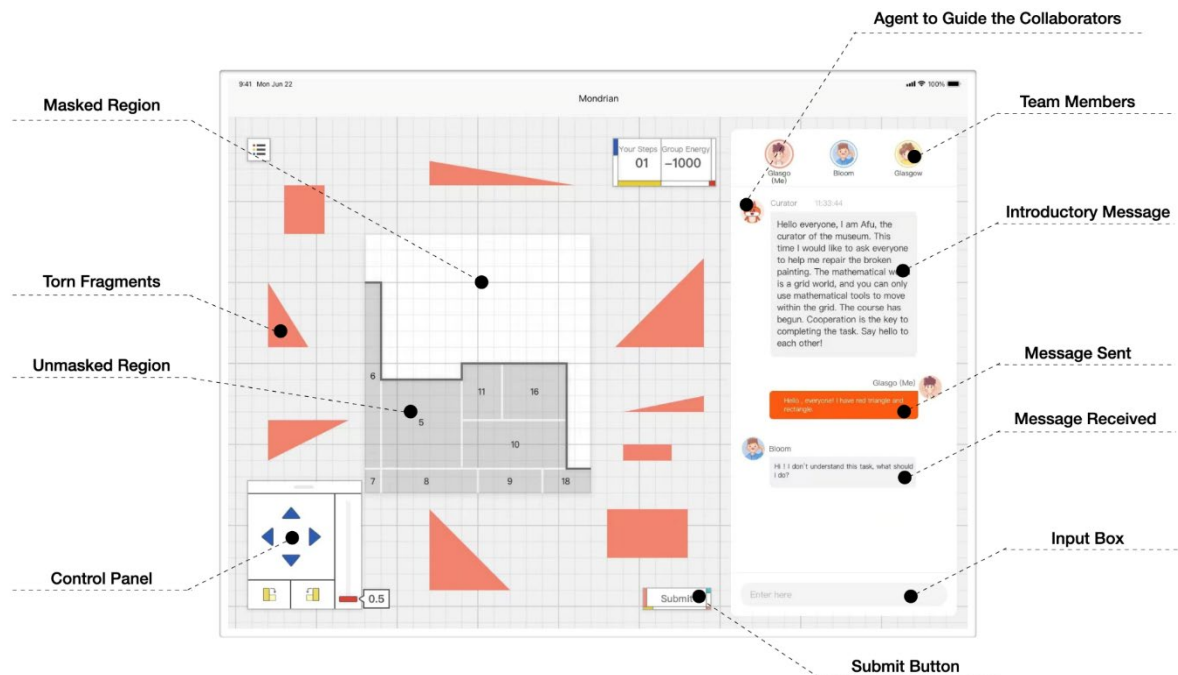
## Method

### Data collection

This research employed a geometrics problem-solving task named Mondrian's Geometric Fantasy (MGF), in which three participants were assigned randomly into a group to collaboratively restore two Mondrian paintings that were torn into triangle and rectangle fragments. These two paintings are divided into two levels with varying difficulty based on the number of fragments. The first painting consists of 13 fragments and the second 16. Collaborators are required to complete the first level before proceeding to the second and the total number of fragments assembled within a set time of 40 minutes was considered the final score.

The task interface is shown in Figure 1, in which participants need to complete the task by rotating and moving the fragments in the left frame and collaborating with teammates via the instant messenger in the right frame. The “painting” in the left frame is divided into various sub-regions, with the shape and number of these sub-regions corresponding to the geometric fragments held by the participants. However, each participant’s field of view is constrained with certain regions being masked, while his two counterparts can perceive the occluded areas. Thus, the participants are motivated to communicate with their collaborators to reveal the situation in those specific regions. Before entering this task interface, all participants will undergo a tutorial to familiarize themselves with the fundamental operation procedures.

**Figure 1**  
*The Task Interface of Mondrian’s Geometric Fantasy*



All participants were recruited students from a high school in China, ages ranging from 17 to 18, with written permission from their guardians and teachers. All students were informed that the task was not a competition; the final result would not be counted as an exam score; and they were free to quit the task at any time without punishment. Altogether 75 students, randomly assigned into 25 teams, participated in the task and cooperated with teammates anonymously. Four teams failed to pass the tutorial session and their data were excluded from the final analysis. The two most proficient teams completed the assembly of 13 fragments, whereas the three least efficient teams failed to assemble any fragment.

These participants’ mastery of the domain knowledge required by the MGF task is virtually identical because the concepts of translation and rotation of geometric shapes are taught in the 4<sup>th</sup> grade of primary school in China. Therefore, the CPS skills tested in the MGF task can be considered domain-free.

Leveraging the robust process data collection mechanism of the Collaborative Learning And Support System (CLASS), the MGF task is capable of capturing every participant’s operational and conversational data at a frequency of 120Hz (Xu et al., 2023) and recording them in log files. CLASS later translated the raw log data to final process data, including Team ID, User ID, Time of Creation, User Name, Content of Conversation, Content of Operation, Virtual Student Name, and other fields, for human coding. The exported process data consists of 16,109 rows.

## Coding process

To ascertain the CPS skills reflected in the process data, this study applied the theoretical framework adopted in the CPS assessment of PISA 2015 as the coding framework (see Table 1), in which CPS skills are considered a matrix composed of 3 socio-emotional dimensions and 4 problem-solving dimensions, thus forming 12 skills

codes (OECD, 2013, 2017a). Meaningless dialogues, usually some off-task social chats, were coded as U standing for “unrelated”.

Two trained coders coded the process data. Five sets of data were randomly selected for the coders to achieve a coding protocol. After establishing a consensus on coding, the remaining 20 sets were coded independently by the two coders who achieved a Cohen’s kappa of 0.74, a high level of agreement (Sun, 2011).

**Table 1**  
*Matrix of Collaborative Problem-Solving Skills for PISA 2015*

	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organization
(A) Exploring and Understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction required and establishing goals	(A3) Understanding roles to solve problem
(B) Representing and Formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describing roles and team organization (communication protocol/rules of engagement)
(C) Planning and Executing	(C1) Communicating with team members about the actions performed	(C2) Enacting plans	(C3) Following rules of engagement
(D) Monitoring and Reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organization and roles

### Hidden Markov modeling

In Hidden Markov Models (HMMs), there exists a sequence of hidden states that cannot be directly observed but can be inferred through a set of observable sequences. These hidden states adhere to the Markov property within the model, meaning each hidden state is dependent solely on its preceding state. HMMs are typically characterized by three sets of parameters: the initial state probability distribution, the state transition probability distribution, and the observation probability distribution (Rabiner & Juang, 1986).

The 21 teams were divided into a high-scoring group (HSG) and a low-scoring group (LSG) with a median score of 11. The two groups have 11 teams and 10 teams respectively. The encoded process data was divided into HSG and LSG and then used as observation sequences.

To establish the optimal number of hidden states in the HMM, we employed both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (see Table 2). The Baum-Welch Algorithm was applied for model parameter estimation.

**Table 2**  
*Model Fit of HMMs with Different Numbers of States*

Number of States	LSG		HSG	
	AIC	BIC	AIC	BIC
2	8324.37	8487.18	8875.33	9060.21
3	7998.19	8262.75	8724.45	9023.11
4	8002.29	8382.17	8574.97	9001.62
5	7995.04	8503.80	8733.11	9301.98
6	7982.97	8634.19	8659.27	9384.58
7	7959.13	8766.37	8634.40	9530.37

Based on the principle that “the lower the AIC or BIC value, the better the model fits” (Akaike, 1973; Konishi, 2004), the number of the hidden states of HSG is well defined, presenting only one possible value of four (AIC=8574.97, BIC=9001.62). For the LSG, we observed that the model has the lowest AIC when the number of hidden states is 7, and the lowest BIC when the number of hidden states is 3. After comparing the observation emission matrix and state transition matrix for both the 3-state and 7-state models, we found 4 states

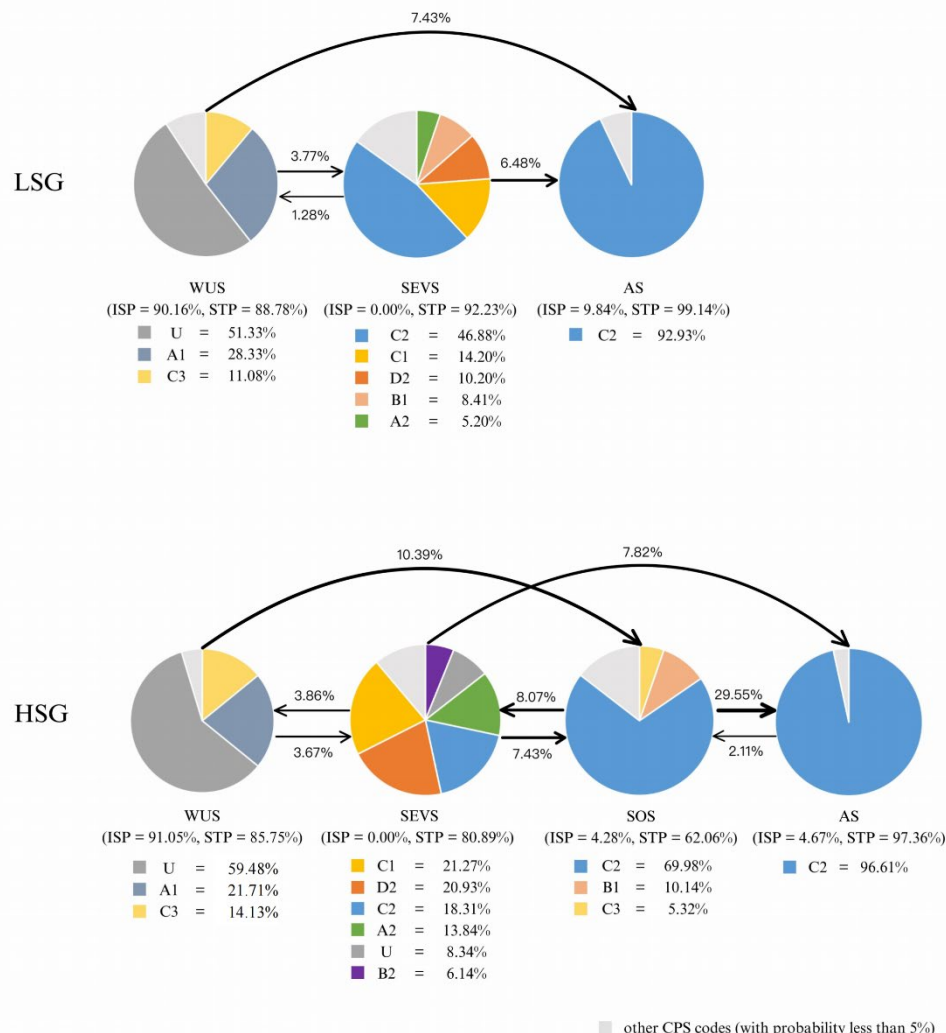
with similar high probabilities of emitting observations, and the probabilities of state transitions are above 90% in the 7-state model. Therefore, we ultimately selected an HMM with 3 states for the LSG.

All computations of HMM learning on observation sequences were performed using the *hmmlearn* package in Python, which is an open-source package for unsupervised learning and inference of Hidden Markov Models (hmmlearn developers, 2023).

## Results

We define the students' problem-solving strategies based on the structure of the observations emitted by each hidden state and interpret students' behavior patterns in conjunction with the state transition probability matrix. The matrix for LSG and HSG is visualized in Figure 2, in which each box represents a state, listing the observations and corresponding probability values that are emitted with a relatively higher probability. Arrows indicate transitions between states, with the numbers beside them reflecting the corresponding transition probabilities. For clarity, transitions with probabilities less than 1% are not displayed.

**Figure 2**  
Visualization of the Model of the Collaborative Problem-Solving Process



### Model of collaboration in LSG

The Hidden Markov Model identified three latent states, which are characterized by the most probable occurrence of CPS skills in each state and named respectively as the *Warm-up State* (WUS), the *Strategy Execution and Verification State* (SEVS), and the *Action State* (AS).



The initial state is WUS, with an initial state probability (ISP) as high as 90%. In this state, participants engage in casual conversation (U=51.33%), acquaint themselves with teammates (A1=28.33%), and encourage their peers to promptly adapt to the situation (C3=11.08%). There is a 7.43% probability that WUS will transition into AS and 3.77% into SEV.

The participants deploy a more extensive array of CPS skills in SEVS. They discover what teammates have (A2=5.20%), engage in communication and establish a shared understanding of the problem (B1=8.41%), direct their teammates' actions (C1=14.20%), execute planned actions according to teammates' instructions (C2=46.88%), and provide feedback on the success of the undertaken actions (D2=10.20%). The probability of a participant remaining in the SEVS state is assessed at 92.23%, while the corresponding transition probabilities to WUS and AS are evaluated at 1.28% and 6.48% respectively.

AS contains only one significant observation which is executing actions (C2=92.93%). It is noteworthy that a striking probability of 99.14% is estimated for the participants to persist in this state, indicating a considerable challenge for them to transition out.

## Model of collaboration in HSG

The model of HSG contains 4 hidden states, with three similar states to LSG, including WUS, SEVS, and AS, and one new state - *Strategy Optimization State* (SOS).

Bearing considerable resemblance to LSG, WUS in HSG is also constituted by the same CPS skills U, A1, and C3. The probabilities are strikingly similar as well, with respective percentages of 59.48%, 21.71% and 14.13%. This state is also the most likely initial state with a probability of 91.05%. The transitioning probabilities from this state to SEVS and SOS are 3.67% and 10.39% respectively.

The scenario for AS is similar. The state for HSG and LSG share a strong resemblance, predominantly composed of a single C2 skill, albeit the probability differs slightly. An important distinction is that participants in HSG display a 2.11% probability of transitioning from AS to SOS, in contrast to LSG who find it difficult to extricate themselves.

HSG also exhibits an extensive application of CPS skills in SEVS, however, there are nominal variations in the specific skills utilized and their probability. The skills of A2 (13.84%), C1 (21.27%), C2 (18.31%), and D2 (20.93%) also manifest in HSG, whereas B1 in LSG is absent with the addition of B2 (6.14%) and U (8.34%), representing instances where participants engage in casual conversation and describe the task to be accomplished to peers. The probabilities of transitioning from SEVS to WUS, SOS, and AS are 3.86%, 7.43%, and 7.82% respectively.

SOS is uniquely observed with HSG. Under this state, participants strive to optimize their strategy by establishing a consensus perspective on the problem (B1=10.14%), executing correspondingly actions (C2=69.98%), and mandating team members to maintain team organizations (C3=5.32%). The transitioning probabilities of this state to SEVS and AS are 8.07% and 29.55% respectively. Although B1 also appears in SEVS of LSG, it is observed that participants in LSG often limit their consensus-building discussions to the problem itself, such as task rules and communication methods. In contrast, participants in HSG are able to negotiate and discuss the understanding of the objective of the problem, "the top right corner is a rectangular area with a size of 1.5\*2". Both these two types of behaviors are coded as B1, but they have different focuses.

## Discussion

By qualitatively comparing the hidden states and transitions between states of the low-scoring and high-scoring groups, a few significant differences in behavior patterns between the two groups are identified.

Firstly, HSG has a unique *Strategy Optimization State*. Through the utilization of the replay function provided by CLASS, it is observable that the participants explicitly instruct their teammates to place a certain fragment into a particular area (CPS skill B1), and the teammates endeavor to act according to the guidance (CPS skill C2). Concurrently, team members mandate others to accomplish collaboration following the same set of rules (CPS skill C3). This suggests that after a period of participation in the collaborative task, high-scoring group students gain a deeper understanding of the task rules and develop better coordination with their teammates, thus spontaneously optimizing their strategies to improve problem-solving efficiency. This may indicate that they possess elevated levels of social and cognitive abilities. Previous research has shown that individuals endowed with these skills demonstrate productivity in CPS activities (Li et al., 2022). This could be a key reason for the performance difference between the two groups.

In addition, the *Strategy Execution and Verification State* differs between the two groups, with a noticeable difference in the emission probability of the "A2" observation. In our participatory observations, we found that in most cases, students in the low-scoring group do not explicitly reveal their fragment holdings. Instead, they directly move the fragment into the visible range and wait for their teammates to notice the fragment before

they provide guidance and feedback. In contrast, students in the high-scoring group tend to be more proactive in clarifying the actions required for problem-solving.

Finally, the state transitions in the high-scoring group are notably more complex than those in the low-scoring group, with a greater variety of transition types. This may suggest that the high-scoring group is adept at continuously exploring and improving during the problem-solving process. Meanwhile, the transitions between states in the LSG exhibit a sense of “disconnection” as if they are “trapped” in their current states and unable to transition to other states.

## Implications and limitations

The skill of CPS is crucial to team success, and understanding student behavior patterns is vital in fostering these skills. In this study, we used HMMs to analyze students’ conversational and operational data during the collaborative problem-solving process and identified four and three distinct states in high-scoring and low-scoring groups, respectively, revealing differences in problem-solving strategies and behavior patterns among teams with varying performances.

The most important finding is that a strategy optimizing state may be crucial to team performance. Both high-scoring and low-scoring groups, after becoming acquainted with each other, enter a stage of strategic discussion and ultimately conclude in a state dominated by plan execution. However, following the strategic discussion, high-scoring teams engage in the application of strategies to observe their effectiveness, making decisions between transitioning into an execution state and reverting to a discussion state. This process can be observed through state transitions. This indicates that educators should focus on students’ state of strategic optimization and intervene timely to assist them in entering the said state or transitioning to other states.

Another important finding pertains to the complexity of state transitions. In comparison to the low-scoring group, the high-scoring group exhibits a more intricate network of state transitions, with a higher probability of transitioning between different states. This may imply that HSG is more flexible in employing problem-solving strategies. Thus, it suggests that when cultivating students’ CPS skills, attention should be given to their decision-making processes regarding state transitions, that is, how they determine when and under what circumstances to enter another state.

However, due to the limitations in the size of the samples, other observations from the CPS assessment framework that were not included in the states did not manifest prominently. This could potentially result in overlooking some valuable problem-solving strategies. The future study will expand the scope of the evaluation to discover a more diverse range of student behavior patterns in the collaborative problem-solving process.

## References

- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. *Proceeding of the Second International Symposium on Information Theory*, 267–281.
- Andrews-Todd, J., & Kerr, D. (2019). Application of ontologies for assessing collaborative problem solving skills. *International Journal of Testing*, 19, 172–187. <https://doi.org/10.1080/15305058.2019.1573823>
- Chang, C. J., Chang, M. H., Chiu, B. C., Liu, C. C., Chiang, S. H. F., Wen, C. T., Hwang, F. K., Wu, Y. T., Chao, P. Y., Lai, C. H., Wu, S. W., Chang, C. K., & Chen, W. L. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education*, 114, 222–235. <https://doi.org/10.1016/j.compedu.2017.07.008>
- Earle-Randell, T. V., Wiggins, J. B., Ruiz, J. M., Celepkolu, M., Boyer, K. E., Lynch, C. F., Israel, M., & Wiebe, E. (2023). Confusion, Conflict, Consensus: Modeling Dialogue Processes During Collaborative Learning with Hidden Markov Models. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (Vol. 13916, pp. 615–626). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-36272-9\\_50](https://doi.org/10.1007/978-3-031-36272-9_50)
- Fiore, S. M., Graesser, A., & Greiff, S. (2018). Collaborative problem-solving education for the twenty-first-century workforce. *Nature Human Behaviour*, 2(6), 367–369.
- Funke, J., & Frensch, P. A. (2017). Complex problem solving: The European perspective—10 years after. In *Learning to solve complex scientific problems* (pp. 25–48). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315091938-2/complex-problem-solving-european-perspective%E2%80%9410-years-joachim-funke-peter-frensch>
- Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students’ minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, 91, 92–105. <https://doi.org/10.1016/j.compedu.2015.10.018>

- Hao, J., & Shu, Z. (2015). Analyzing Process Data from Game/Scenario- Based Tasks: An Edit Distance Approach. 7(1).
- hmmlearn developers. (2023). Tutorial—Hmmlearn documentation. <https://hmmlearn.readthedocs.io/en/latest/tutorial.html>
- Kerr, D., & Chung, G. K. (2012). Identifying key features of student performance in educational video games and simulations through cluster analysis. *Journal of Educational Data Mining*, 4(1), 144–182.
- Klingler, S., Käser, T., & Solenthaler, B. (2016). Temporally Coherent Clustering of Student Data. *Proceeding of the 9th International Conference on Educational Data Mining*.
- Konishi, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, 91(1), 27–43. <https://doi.org/10.1093/biomet/91.1.27>
- Li, S., Pöysä-Tarhonen, J., & Häkkinen, P. (2022). Patterns of action transitions in online collaborative problem solving: A network analysis approach. *International Journal of Computer-Supported Collaborative Learning*, 17(2), 191–223. <https://doi.org/10.1007/s11412-022-09369-7>
- Lopez-Pernas, S., & Saqr, M. (2021). Bringing Synchrony and Clarity to Complex Multi-Channel Data: A Learning Analytics Study in Programming Education. *IEEE Access*, 9, 166531–166541. <https://doi.org/10.1109/ACCESS.2021.3134844>
- OECD. (2013). PISA 2015 collaborative problem solving framework. OECD Publishing Paris, France.
- OECD. (2017a). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving (revised edition). OECD Publishing.
- OECD. (2017b). PISA 2015 Results Collaborative Problem Solving. PISA.
- Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning*, 18(1), 39–66. <https://doi.org/10.1007/s11412-023-09387-z>
- Pan, Z., Li, C., & Liu, M. (2020). Learning Analytics Dashboard for Problem-based Learning. *Proceedings of the Seventh ACM Conference on Learning @ Scale*, 393–396. <https://doi.org/10.1145/3386527.3406751>
- Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1), 4–16. <https://doi.org/10.1109/MASSP.1986.1165342>
- Rodríguez, F. J., & Boyer, K. E. (2015). Discovering Individual and Collaborative Problem- Solving Modes with Hidden Markov Models. *Artificial Intelligence in Education. AIED 2015*.
- Stadler, M., Fischer, F., & Greiff, S. (2019). Taking a Closer Look: An Exploratory Analysis of Successful and Unsuccessful Strategy Use in Complex Problems. *Frontiers in Psychology*, 10, 777. <https://doi.org/10.3389/fpsyg.2019.00777>
- Sun, S. (2011). Meta-analysis of Cohen’s kappa. *Health Services and Outcomes Research Methodology*, 11(3–4), 145–163. <https://doi.org/10.1007/s10742-011-0077-3>
- Xiao, Y., He, Q., Veldkamp, B., & Liu, H. (2021). Exploring latent states of problem-solving competence using hidden Markov model on process data. *Journal of Computer Assisted Learning*, 37(5), 1232–1247. <https://doi.org/10.1111/jcal.12559>
- Xu, Q., Zhang, B., Wang, J., Zhang, X., & Zhu, M. (2023). A Collaborative Learning and Support System for STEM Education and Learning Analytics. *13th IEEE Integrated STEM Education Conference (ISEC 2023)*, Laurel, Maryland.
- Yuan, J. L., Xiao, Y., & Liu, H. Y. (2019). Assessment of Collaborative Problem Solving Based on Process Stream Data: A New Paradigm for Extracting Indicators and Modeling Dyad Data. *Frontiers in Psychology*, 10, 14. <https://doi.org/10.3389/fpsyg.2019.00369>
- Zhang, J.-H., Zou, L., Miao, J., Zhang, Y.-X., Hwang, G.-J., & Zhu, Y. (2020). An individualized intervention approach to improving university students’ learning performance and interactive behaviors in a blended learning environment. *Interactive Learning Environments*, 28(2), 231–245. <https://doi.org/10.1080/10494820.2019.1636078>
- Zhu, M., Shu, Z., & von Davier, A. A. (2016). Using networks to visualize and analyze process data for educational assessment. *Journal of Educational Measurement*, 53(2), 190–211.

## Acknowledgments

This study is supported by the National Natural Science Foundation of China (62177044). The data used in this study is from the work supported by Huohuaschool.com.

# Interaction Forms of Collaborative VR Video Learning: An Exploratory Study

Qiao Jin, Ishan Sodhi, Angie Chen, Svetlana Yarosh  
jin00122@umn.edu, sodhi012@umn.edu, chen7313@umn.edu, lana@umn.edu  
Department of Computer Science and Engineering, University of Minnesota

**Abstract:** Collaborative and immersive environments both contribute to increasing engagement in video-based learning (VBL). This exploratory study delves into the dynamics of interaction forms within collaborative VR video-based learning and evaluates how shared and individual video control influence these interactions. Through a within-subject study involving 18 groups of three participants ( $N = 54$ ), we aim to address three research questions: 1) identifying interaction forms in collaborative VR video-based learning, 2) comparing interaction forms between shared and individual video control, and 3) examining the relationship between interaction forms and knowledge acquisition and satisfaction. Our findings provide valuable insights for technology developers, researchers, and educators looking to enhance VBL systems and instructional approaches, as well as gain a deeper understanding of how social interactions impact knowledge acquisition and satisfaction in VR video-based learning environments.

## Introduction

Video-based learning (VBL) has been shown to be an effective educational approach (Kay, 2012), leveraging visual and auditory elements for enhanced learning process. However, VBL for online classes faces challenges such as isolation and reduced motivation (Fang et al., 2022). In response to these challenges, collaborative video viewing, such as the Distributed Collaborative Video Viewing (DCVV) model, has emerged as a pedagogical strategy with broad-reaching benefits (Cadiz et al., 2000). These approaches have been demonstrated to enhance students' engagement and improve their overall learning experiences. Further, the infusion of Virtual Reality (VR) technology into VBL (e.g., VR videos<sup>(1)</sup>) has been shown to amplify students' intrinsic motivation, bolster engagement in learning activities (Fang et al., 2022). Additionally, VR environments naturally offers a shared digital space that allows students to immerse themselves in learning contexts, aiding students in constructing and retaining visual knowledge (Kallioniemi et al., 2017).

Previous research has connected the effectiveness of collaborative learning to the types of interactions taking place among learners (Vuopala et al., 2016). Additionally, research has shown that technology design can influence the types of interactions that occur in virtual learning environments. Inspired by this, our study seeks to examine the various forms of interaction within different collaborative modes and their impact in the context of VR video-based learning. In pursuit these objectives, we formulate the following research questions:

- *RQ1*: What forms of interaction occur during collaborative VR video-based learning?
- *RQ2*: How do forms of interaction vary between the shared control (Sync mode) and individual control (Non-sync mode) in collaborative VR video viewing systems?
- *RQ3*: How do forms of interaction relate to the measures of knowledge acquisition and satisfaction?

We conducted a within-subject experiment with 18 groups and utilized video analysis to answer our research questions. The study reveals how collaborative interactions unfold and impact on knowledge acquisition and satisfaction during VR video-based field trips, shedding light on the influence of shared and individual video control techniques on interaction patterns. Our findings advance the understanding of collaborative learning within VR videos, offering preliminary insights that can inform both technology design and practical instructional strategies in the field of computer-supported collaborative learning (CSCL).

## Related work

Video-based learning (VBL) can be an effective pedagogical approach (Kay, 2012) as it promotes greater sensory engagement and interactivity. However, some limitations of VBL have been proposed in prior work: students risk distraction, loss of motivation, or isolation without communication with an instructor or peers, potentially hindering learning engagement (Fang et al., 2022). Watching videos together positively increases the attention span and engagement of students (Li et al., 2014), and fosters conversation, allowing students to share perspectives and generate conceptual diversity (Goldman, 2014). These conversations are vital for knowledge building, and some literature suggests that they may be as or more important than the videos themselves (Pea & Lindgren, 2008). One collaborative VBL model is the Distributed Collaborative Video Viewing (DCVV) model proposed

by Cadiz et al. (2000) wherein small groups of students watch educational videos without an instructor, periodically pausing to discuss, which was shown to have positive effects on student engagement and learning.

Alternatively, problems found with VBL such as poor intrinsic motivation and engagement (Torff & Tirota, 2010), may be solved by learning in an immersive environment. One possible immersive environment is VR video which is more cost-effective and accessible (Jin et al. 2022) compared to other computer-generated graphical immersive environments. VR videos have been shown to increase students' intrinsic motivation, engagement in learning activities (Fang et al., 2022), allows students to access otherwise inaccessible learning contexts, and helps students build and retain visual knowledge (Kallioniemi et al. 2017). VR video can be applied successfully in many learning scenarios (Pirker & Dengel, 2021) include virtual tours, recorded processes and procedures, recorded situations, recorded experiences, etc. Although some literature exists on collaboration within VR video-based learning (Jackson & Fagan, 2000; Jin, et al. 2023), there has been little investigation into the forms of interaction between students in collaborative VR video-based learning. It has been shown that the forms of interaction which take place during a collaborative learning session can have a significant effect on its success (Vuopala et al., 2016). Certain forms of interaction significantly contribute to the quality of collaborative learning (e.g., interaction was more group-related than task-related) (Lebie et al., 1996).

Therefore, we specify our research to VR video-based learning environment, seeking to investigate the influence of technology design on various forms of interaction and their impact on learning outcomes. More specifically, our study aims to explore interaction forms, their effects on knowledge acquisition and satisfaction, and how technology influences these interactions. This is important for educators, engineers, and researchers in designing learning environments that encourage these positive interactions.

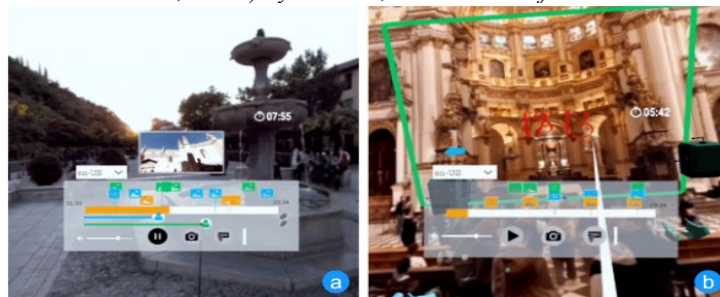
## Collaborative VR video viewing systems

We implement two collaborative VR video viewing systems (Figure 1) to explore how different video controls influence the forms of interaction in collaborative VR video-based learning. The Non-sync mode in collaborative VR video watching offers individual timeline control with collaborative tools enhancing in-VR communication. It includes *awareness* tools like spatialized voice chat, activity visualization displaying user progress and notes, and viewport visualization showing gaze directions (only under following function). *View sharing* tools comprise peek and full window options, and a follow function for synchronized viewing. *Note-taking* tools feature drawing, speech-to-text notes, and screenshot capabilities, with visibility adjustments for different timelines.

In contrast, the Sync mode, based on the DCVV model, allows shared control over the video playback. This mode enhances *awareness* through embodied visualization with 3D avatars and spatialized voice chat using avatar positions. It simplifies activity visualization by omitting individual progress displays but maintains note indicators. The visualized gaze is constant due to shared timeline progress, and the ray cursor of teammates is color-coded for better interaction. Unlike the Non-sync mode, view sharing functions are not required due to shared progress, and *note-taking* tools such as screenshots and speech-to-text notes are consistent, with drawings visible to all users at all times. The design considerations and technical details of systems are described in our previous work (Jin et al., 2023).

**Figure 1**

*Collaborative VR video viewing systems: a) Non-sync mode, which enables individual timeline control, and b) Sync mode, which allows for shared video.*



## Methods

### Participants and learning settings

This study is a part of a larger research project investigating collaborative VR video tools. 54 participants in 18 groups of three from a Midwest U.S. university involved in this study. Teams were formed based on availability

and participants' choices for teammates. Participants were aged 19 to 31 ( $M = 22.5$ ,  $SD = 2.65$ ). There were 26 females, 27 males, and one undisclosed gender. Of the participants, 26 were unfamiliar with their teammates, 18 knew one teammate, and ten knew two teammates. 14 had significant VR experience, 26 had limited experience, and 14 were VR newcomers. 23 had watched 360 videos using VR headsets, 12 watched them on non-VR devices, and 19 had no 360-video exposure. All had normal or corrected-to-normal vision and received a \$50 gift card for their participation.

The study took place in four adjacent university rooms—three for participants and one for an in-VR tech facilitator. Each room included a swivel chair, desk, laptop, VR headset, and a tripod-mounted camera capturing the process. In this study, we chose a virtual field trip as the learning context to leverage VR's capacity for accessing remote locations. A group would be watching VR videos together and discussing the city's attractions, history, and architecture. The videos, from the “One day in” 360° travel collection<sup>(2)</sup>, depicted various global destinations. We used two narrated city-tour videos, edited for equal length and difficulty, and narrated by an AI-generated voice. Each video, highlighting historical and architectural aspects, lasted about 3.5 minutes and contained 14 clips, with resolutions of 1920 x 1024 pixels.

## Procedure

The study is primarily divided into four parts: ice-breaking and introduction, training, and learning units (each featuring one mode). Within each learning unit, there are components for pre-knowledge assessment, a video session, and post-knowledge acquisition assessment, along with a satisfaction questionnaire. After a brief ice-breaking activity (~5 minutes), where three participants introduced themselves and shared VR-related impressions or experiences, an introduction (~5 minutes) by researchers outlined the study's objectives, procedures, learning goal, with an overview of system features. Following this, participants wore headsets for a system tutorial. Training duration ranged from 20 to 30 minutes based on participants' prior VR experience. After training, participants took a 5-minute break to ensure they were free from cybersickness. Once all participants were ready to enter the learning unit, the knowledge assessment as a pre-learning session began. Then, participants entered the system to watch the video. In this session, participants watched a 3.5-minute video and could pause, replay, and relocate it as needed within 10 minutes. Post-study, knowledge acquisition was assessed, followed by self-reported questionnaires of satisfaction. Two learning units (Sync and Non-sync mode) were conducted in a counterbalanced order, separated by a 5-minute break based on the motion sickness participants experienced.

## Measures

The knowledge assessment was designed to assess students' knowledge acquisition as result of experiencing two technologies. The test consisted of 10 multi-option questions regarding the information presented in the videos according to Bloom's taxonomy (Anderson & Krathwohl, 2001) and prior educational VR video research (Radia et al., 20018), including four questions based on auditory information, three visual information, and three conceptual questions. The test was piloted two times and adjusted in order to avoid a ceiling effect.

Participants took an independent self-report questionnaire about their satisfaction after watching the video. We selected satisfaction as a metric since it is a significant indicator reflected “the degree of learner reaction to values and quality of learning, and motivation for learning” (Saffo et al., 2021). There were 11 questions in this validated questionnaire (So & Brush, 2008), aimed at assessing participant satisfaction with both the learning activities and the extent to which the system design aligned with their learning expectations. Because those measures were all collected individually, and we used an average of three participants as the final score of the measure for this group.

## Data analysis

We chose a qualitative approach to gain insights into the qualities and variations in forms of interaction across different modes. A total of 36 videos from 18 groups were collected during the study. Two researchers watched and coded five videos together based on a prior coding scheme and process (Vuopala et al., 2016). Throughout the analysis, there are three main categories: Task-related, Group-related, and Off-task interactions and related subcategories and codes. The coding scheme was iteratively updated to better fit the study's learning content, rather than the original's. We used the new coding scheme to code another five videos and conducted the inter-rater reliability test. The final agreements from the Cohen's Kappa reached around 70%. Finally, we coded the rest of the videos individually, as well as counted the *occurrence* and the *proportion* of each item in coding scheme within each group to answer the RQ1. For RQ2, a one-way repeated measures ANOVA was conducted to check for significant effects of video control on the occurrence of coding scheme items in different levels. For RQ3, we used multiple regression analyze (Mason & Perreault, 1991) to identify which forms of interaction have a significant impact on the selected measures, and how much of an impact they have. We added “video technology

mode” as one feature because it might influence the selected measures and used Dummy Coding to process categorical variables in multiple regression. We did the feature selection using correlation matrix with all codes in different levels to determine which features were dependent to improve the accuracy of our model and avoid overfitting. After selecting the features based on the correlation matrix, we then performed a multiple regression model to analyze the relationship between the selected features and the response measures.

## Results

### RQ1: What forms of interaction occur during collaborative VR video-based learning?

We collected a total of 1350 codes from 36 videos to analyze the forms of interaction and their proportions during collaborative VR video-based learning. Our coding scheme (Table 1) was adapted from a previous study (Vuopala et al., 2016). We found that *task-related interactions* were predominant, accounting for more than half of the interactions (50.89%). *Group-related interactions* (48.74%) were nearly as prevalent, highlighting the social and organizational aspects of collaboration. *Off-Task interactions* were minimal (0.37%), indicating occasional deviations from the primary focus. In terms of subcategories, *socio-emotional expressions* made up 25.49% of interactions, emphasizing the importance of emotional aspects in group dynamics. *Coordination of group activities* accounted for 23.25% of interactions, showcasing a structured approach to collaboration. Then, *answer or comment* (19.63%), *question* (19.04%), and *new knowledge* (12.22%) had relatively lower frequencies during VBL. At the code level, *declaratory comments* (12.59%), *organizing ongoing activities* (10.44%), and *expressing cohesion* (10.3%) were the most prevalent interaction forms.

**Table 1**

*Coding Scheme for Forms of Interaction and Their Proportions during Video-based Learning*

Main category	Sub category	Code	Coding rule	Example
Task-related interaction (50.89%)	Answer or comment (19.63%)	Declaratory comment (12.59%)	Agrees, states, repeats	“Okay. That sounds good.”
		Comment with explanation (7.04%)	Explains, justifies, clarifies	“Maybe just rewatch, because we can discuss a little bit when revisit the content...”
	Question (19.04%)	Clarifying question (7.04%)	Clarifies previous question or asks for clarification	“What did you say about this building?”
		New question (6.44%)	Brings new question into the discussion	“What’s the year of this museum?”
		Suggestion (5.56%)	States or suggests and waits for comments.	“You guys wanna watch the last few?”
	New knowledge (12.22%)	Content-based (10.15%)	Brings new topic based on the video content	“The Portuguese tiles are the bits of reference to the walls.”
		Experience-based (2.07%)	Brings new topic based on experience or opinion	“This was on the pre-questionnaire.”
Group-related interaction (48.74%)	Socio-emotional expressions (25.49%)	Expressing cohesion (10.3%)	Helping, rewarding, acknowledging	“That’s a good point. Very cool.”
		Accompanying (9.04%)	Expressing presence, mumbling	“...Moorish... fort... (repeat the video content)”
		Decreasing tension (6.15%)	Laughing, joking	“Ah, Gothic oyster, that’s close enough but I cannot eat. (laugh)”
	Coordination of group	Organizing ongoing activities (10.44%)	Planning and organizing current group activities	“Ok, I’m gonna pause it and you take notes of the things she said about this.”

	activities (23.25%)	Organizing upcoming activities (4.74%)	Planning and organizing future group activities	“I don't know exactly where anybody wants to go but just skip back to someplace you need to review.”
		Technological issues (4.15%)	Technological challenges, use of technology	“I was trying to delete my notes because I realized that it did not interpret what I said at all.”
		Evaluating group work (2.59%)	Evaluating group work	“Our notes so far probably missed a good chunk of information”
		Reporting current activities (1.33%)	Reporting, acknowledging current self-activities to group	“Um, I'm writing it down.”
Off-task interaction (0.37%)	Off-task (0.37%)	Off-task (0.37%)	Topics that are not related to course content or group work.	“Wee I'm floating!”

## RQ2: How do forms of interaction vary between the shared control (Sync) and individual control (Non-sync) in collaborative VR video viewing systems?

We investigated the difference of occurrences and proportion of interaction forms under two collaborative conditions (Sync and Non-sync mode) in different levels of coding scheme.

### Occurrence of interaction forms

The occurrence reflects the number of times an interaction takes place. The results showed that the mean of *task-related interaction* occurrences in Non-sync mode was statistically significantly lower than that of the Sync mode ( $p < 0.01$ ,  $d = 1.09$ ). Within *task-related interactions*, the subcategory of *new knowledge* occurrences had a statistically significant difference ( $p = 0.011$ ,  $d = 0.84$ ), with the sync mode having a higher mean than the Non-sync mode. When further broken down, the code *content-based* has a statistically significant ( $p = 0.01$ ,  $d = 0.89$ ) difference between the frequency of occurrence of the two modes, with Sync having a higher frequency. Likewise for the subcategory *question*, with the Sync mode mean with a higher statistically significant ( $p < 0.01$ ,  $d = 1.00$ ) mean number of occurrences than the Non-sync mode. The codes within this subcategory followed suit, with *new question* ( $p < 0.01$ ,  $d = 1.14$ ) and *suggestion* ( $p < 0.01$ ,  $d = 1.28$ ) both having a statistically significant higher mean number of occurrences in the Sync mode over the Non-sync mode. Another code within this subcategory was *clarifying question*. While this code did not have a strong statistically significant difference ( $p = 0.06$ ,  $d = 0.71$ ), we still saw that it had a higher mean number of occurrences in Sync as compared to Non-sync mode. The subcategory *answer or comment* shared similar values ( $p < 0.01$ ,  $d = 1.02$ ), with the mean number of occurrences for Sync mode statistically significantly higher than the Non-sync mode. Within this subcategory, we saw that both codes *declaratory comment* ( $p = 0.016$ ,  $d = 0.98$ ) and *comment with explanation* ( $p = 0.04$ ,  $d = 0.61$ ) had a statistically significant difference between the two modes. Again, we saw that both codes had a higher mean frequency of occurrence in the Sync mode when compared to the Non-sync mode. No other subcategories nor codes within task-related interactions saw any other significant differences.

Although we didn't find the statistical significance in *group-related interaction* between the two modes, the results show that Sync mode brings a higher frequency *group-related interaction* ( $p = 0.056$ ,  $d = 0.65$ ). Within the category, the subcategory *coordination of group activities* has a significantly significant ( $p = 0.022$ ,  $d = 0.83$ ) higher mean number of occurrences for the Sync compared to the Non-sync mean. Similarly, the code *organizing ongoing activities* from this subcategory had a statistically significant difference ( $p = 0.012$ ,  $d = 1.00$ ) in the mean frequency of occurrence between the two groups, with Sync mode having the higher mean. While the subcategory *socio-emotional expressions* did not see a statistically significant difference between the two modes, the code *expressing cohesion* within that subcategory did. There was a statistically sufficient difference ( $p = 0.026$ ,  $d = 0.80$ ) between the means. Once again, the Sync mode saw a higher mean than the Non-sync mode. Within this category, no other codes nor subcategories had any significant differences. We did not see any significant differences in the *off-task* category.

### Proportion of interaction forms

Proportion refers to the fraction or ratio of the number of times an interaction occurs relative to the total number of interactions. The results also demonstrate that the proportion of *task-related interactions* in Non-sync mode was statistically significantly lower than that of the Sync mode ( $p = 0.022$ ,  $d = 0.62$ ). Breaking this down, the



subcategory *answer or comment* has a significant difference ( $p = 0.021$ ,  $d = 0.88$ ) between the Sync and Non-sync mode, with Sync being higher. Looking at the individual codes within this subcategory, *declaratory comment* is the only one with a significant difference ( $p = 0.017$ ,  $d = 0.85$ ), Sync having a higher mean proportion than Non-sync. Whilst the subcategory *questions* show no significant difference in mean proportion, the code *suggestion* also shows a significant difference ( $p = 0.052$ ,  $d = 0.75$ ), with Non-sync having lower mean proportion of occurrences than Sync mode. The categories *group-related interaction* and *off-task interaction* did not show any significant changes. No other subcategories nor codes show any significant differences.

### RQ3: How do forms of interaction relate to the measures of knowledge acquisition and satisfaction?

#### Knowledge acquisition

For *auditory knowledge*, we didn't find statistically significant independent variables in main categories and subcategories for auditory knowledge acquisition over occurrence. For the occurrence of codes, it was found that occurrence of *content-based* ( $\beta = 0.0195$ ,  $p = 0.024$ ) and *reporting current activities* ( $\beta = 0.1229$ ,  $p = 0.012$ ) significantly predicted audio-based knowledge acquisition with positive effects. This results came from the regression model with weak statistical significance,  $F(11, 22) = 1.906$ ,  $p = 0.095$ ,  $R^2 = 0.232$ . For the proportion, the multiple regression analysis did not reveal any statistically significant relationships between the audio-based knowledge acquisition and the proportions of main categories and subcategories. For the proportion in code level, it was found that *organizing upcoming activities* ( $\beta = 5.948$ ,  $p = 0.008$ ) and Sync mode ( $\beta = 0.1736$ ,  $p = 0.042$ ) significantly predicted audio-based knowledge acquisition, suggesting a positive impact. These results came from a non-statistical significance regression model,  $F(17, 16) = 1.615$ ,  $p = 0.172$ ,  $R^2 = 0.632$ .

For *conceptual knowledge*, we didn't find statistically significant independent variables in main categories and subcategories for these conceptual knowledge acquisition over occurrence. For the occurrence of all codes, it was found that the *experience-based* ( $\beta = -0.08$ ,  $p = 0.012$ ) significantly predicted conceptual knowledge acquisition, suggesting negative effects. This is resulted from a non-statistical significance regression model,  $F(11, 22) = 1.228$ ,  $p = 0.326$ ,  $R^2 = 0.38$ . For the proportion, we didn't find statistically significant independent variables in main categories and subcategories. It's worth noting that in subcategories over proportion, *question* ( $\beta = 0.7941$ ,  $p = 0.035$ ) had a positive effect, and *coordination of group activities* ( $\beta = -0.4593$ ,  $p = 0.040$ ) had a negative effect, both statistical significantly influencing conceptual knowledge acquisition. When considering all codes, the model did not reach statistical significance,  $F(17, 16) = 1.795$ ,  $p = 0.124$ ,  $R^2 = 0.656$ . However, it was found that *clarifying question* ( $\beta = 1.3765$ ,  $p = 0.040$ ), *declaratory comment* ( $\beta = 1.3235$ ,  $p = 0.033$ ), *organizing ongoing activities* ( $\beta = -1.3235$ ,  $p = 0.011$ ) significantly predicted conceptual knowledge acquisition for the proportions of all codes.

For *visual knowledge*, we didn't find statistically significant for the regression model nor independent variables in three levels influencing the occurrence or proportion of virtual knowledge acquisition.

#### Satisfaction

There is no statistically significant difference in regression models on the occurrence of main category and sub category for predicting satisfaction. We found that the regression for satisfaction over occurrence of all code was statistically significant,  $F(11, 22) = 4.455$ ,  $p = 0.0014$ ,  $R^2 = 0.69$ . It was found that the occurrence of codes *content-based* ( $\beta = 0.0441$ ,  $p = 0.004$ ), *decreasing tension* ( $\beta = 0.0591$ ,  $p = 0.046$ ), *accompanying* ( $\beta = 0.0754$ ,  $p = 0.005$ ) significantly predicted satisfaction, suggesting a positive effect on satisfaction. However, *technological issues* ( $\beta = -0.0591$ ,  $p = 0.046$ ), *reporting current activities* ( $\beta = -0.3927$ ,  $p < 0.001$ ) and Non-sync mode ( $\beta = -0.2572$ ,  $p = 0.040$ ) indicated a negative impact on satisfaction. No independent variables were found to significantly predict satisfaction in any of the proportion models.

### **Conclusion and discussion**

By designing and evaluating two collaborative VR video viewing systems — one with shared video control (Sync mode) and the other with individual control (Non-sync mode) — we have explored the dynamics of collaborative interactions in VR video-based learning and interaction forms' impact on knowledge acquisition and learner satisfaction. This exploratory study provides the foundation for additional research of understanding forms of interaction in CSCL environments in VR and optimizing the design the systems and pedagogical strategies.

Our findings from RQ1 reveal that task-related interactions played a pivotal role in the collaborative VR VBL environment, constituting the majority at 50.89%. This dominance underscores the fundamental nature of task-related discussions within the learning process. Interestingly, group-related interactions closely followed,

accounting for 48.74% of the interactions. These interactions emphasize the significance of the social and organizational dimensions inherent in collaborative learning, indicating learners' active engagement in group dynamics, which is similar to prior CSCL interaction forms research (Vuopala et al., 2016). When we delved into the subcategories and codes of interactions, the results underscore the crucial role of emotional elements (e.g., *socio-emotional expressions*) and structured and organized group activities (e.g., *coordination of group activities*) in group dynamics the overall learning experience. This result implies that when designing collaborative learning environments, supporting task-related interactions and group-related interactions from both instructional and technical perspectives are very important.

RQ2 examines the variations in occurrence and proportion of interaction forms within two collaborative conditions. Based on our results, Sync mode appears to offer several advantages over Non-sync mode in the context of CSCL: it suggests more dynamic knowledge exchange, active questioning and discussion, and better coordination among learners compared to Non-sync mode. The major reason is that Sync mode provides the shared context for participants, so they're more comfortable communicating with each other and spending more time on discussion. When came to group conversation structure (aka. proportion). Sync mode has a higher proportion of task-related interactions compared to Non-sync mode. However, it is interesting that the group-related interaction and off-task interaction proportion of the whole discussion did not show any significant changes between modes. The result implies that when designing collaborative VR learning experiences, emphasizing synchronous communication and a shared context can lead to more knowledge exchange related to task.

From the results of RQ3, fostering content-based interactions, especially those involving new knowledge, appeared to enhance auditory knowledge acquisition. Additionally, effective group coordination through activities like organizing upcoming tasks and reporting current activities positively influenced auditory knowledge acquisition. Moreover, prioritizing the proportion of organizing upcoming activities during discussions was linked to increased audio-based knowledge acquisition. Surprisingly, Non-sync mode, which allows asynchronous interactions, seemed more conducive to auditory knowledge acquisition compared to Sync mode. When it comes to conceptual knowledge, encouraging students to ask questions, particularly clarifying ones, positively impacted learning. However, an excessive focus on group coordination activities appeared to hinder conceptual knowledge acquisition, suggesting that balance is crucial. Regarding satisfaction, content-based interactions were found to have a positive influence, emphasizing the importance of substantive discussions. Social interactions, such as accompanying and reducing tension significantly enhanced satisfaction. However, the technological issues and an overemphasis on reporting current activities negatively affected satisfaction. Interestingly, Sync mode was associated with higher satisfaction levels than Non-sync mode, possibly due to real-time interactions.

This exploratory research still requires confirmation from field studies and other subjects or educational scenarios. The most significant limitation of this study is its small sample size. Due to this limitation, our regression models about knowledge acquisition and satisfaction may not be strong enough to reach statistical significance as a group. While the models may not be significant, we still reported some individual independent variables which were found to be significantly explanatory for a dependent variable. This suggests that while those independent variables may be related to the dependent variable, the entire set of independent variables may not be jointly related. Additionally, it is possible that there may be interaction effects or nonlinear relationships between the independent variables and the dependent variable that are not captured by the current model. It is also possible that the sample size is too small to detect a significant effect of the independent variables on the dependent variable. So, due to the limited sample size, it was difficult to form a convincing conclusion. This work was also subject to some limitations regarding contextual factors which may influence forms of interaction between its participants. Specifically, relationships between team members, the nature of the learning content or tasks, and the duration of the learning session vary across different learning contexts and may play a role in forms of interaction which occur. The findings of this study may not be directly generalizable to situations with distinct team dynamics, content domains, task complexities, or study durations. Further, the within-subject design of this study introduces a potential limitation through the risk of carryover effects, as participants' experiences with one control mode may influence subsequent ones, potentially leading to order effects and impacting the validity of results. Furthermore, the study's controlled laboratory environment may limit its generalizability to real-world educational contexts. Factors such as individual infrastructure, access to technology, and classroom dynamics, which play significant roles in authentic educational settings, are not fully captured in this controlled environment.

Therefore, we suggest future research to address these limitations. Specifically, conducting larger-scale studies with diverse participant groups and considering various educational scenarios will help validate and generalize the findings. Additionally, future research should include a thorough power analysis to determine the necessary sample size to achieve statistical significance in regression models. This will help ensure that studies are adequately powered to detect meaningful effects and draw more robust conclusions. Researchers could also explore potential interaction effects and nonlinear relationships among variables and employ more advanced

statistical techniques to account for these complexities. To enhance the external validity of the findings, future research should also consider contextual factors (e.g., content domains, task complexities, and study durations). Comparative studies across different educational settings and learning contexts will provide a more comprehensive understanding of how forms of interaction vary. Moreover, researchers should adopt a mixed-methods approach, combining quantitative data with qualitative insights to gain a deeper understanding of the dynamics of interaction in VR-based collaborative learning. This approach can capture nuances that quantitative measures may overlook. Lastly, future studies should aim to replicate collaborative technologies in more authentic educational environments. This will ensure that the results are applicable to real-world educational scenarios, providing valuable insights for educators and instructional designers.

## Endnotes

- (1) In this work, “VR videos” specifically refer to monoscopic 360-degree videos experienced through a VR headset.  
 (2) “One day in” 360° travel videos: <https://www.youtube.com/playlist?list=PLHiCdB8YTO76Gv843e8rdcza-FJbYCYSj>

## References

- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.
- Fang, J., Wang, Y., Yang, C. L., Liu, C., & Wang, H. C. (2022). Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1-21.
- Goldman, R. (2014). ORION™, An Online Digital Video Data Analysis Tool: Changing Our Perspectives as an Interpretive Community. In *Video research in the learning sciences* (pp. 507-520). Routledge.
- Jackson, R. L., & Fagan, E. (2000, September). Collaboration and learning within immersive virtual reality. In *Proceedings of the third international conference on Collaborative virtual environments* (pp. 83-92)
- Jin, Q., Liu, Y., Yarosh, S., Han, B., & Qian, F. (2022). How will VR enter university classrooms? multi-stakeholders investigation of vr in higher education. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).
- Jin, Q., Liu, Y., Sun, R., Chen, C., Zhou, P., Han, B., ... & Yarosh, S. (2023). Collaborative Online Learning with VR Video: Roles of Collaborative Tools and Shared Video Control. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).
- Kallioniemi, P., Keskinen, T., Hakulinen, J., Turunen, M., Karhu, J., & Ronkainen, K. (2017, November). Effect of gender on immersion in collaborative iodv applications. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (pp. 199-207).
- Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3), 820-831
- Lebie, L., Rhoades, J. A., McGrath, J. E. (1996). Interaction process in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work*, 4(2-3), 127-152.
- Li, N., Verma, H., Skevi, A., Zufferey, G., & Dillenbourg, P. (2014). MOOC learning in spontaneous study groups: Does synchronously watching videos make a difference? (No. CONF, pp. 88-94). *PAU Education*.
- Mason, C. H., & Perreault, W. D. (1991). Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research*, 28(3), 268-280.
- Pea, R. & Lindgren, R. (2008). Video laboratories for research and education: An analysis of collaboration design patterns. *IEEE Transactions on Learning Technologies*, 1(4), 235-247.
- Pirker, J., & Dengel, A. (2021). The Potential of 360-Degree Virtual Reality Videos and Real VR for Education-A Literature.
- Radia, M., Arunakiranthan, M., & Sibley, D. (2018). A guide to eyes: ophthalmic simulators. *The Bulletin of the Royal College of Surgeons of England* 100, 4 (2018), 169-171.
- Saffo, D., Bartolomeo, S. D., Yildirim, C., & Dunne, C. 2021. Remote and Collaborative Virtual Reality Experiments via Social VR Platforms. Association for Computing Machinery, New York, NY, USA.
- So, H.-J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, 51(1), 318-336.
- Torff, B., & Tirota, R. (2010). Interactive whiteboards produce small gains in elementary students' self-reported motivation in mathematics. *Computers & Education*, 54(2), 379-383.
- Vuopala, E., Hyvönen, P., & Järvelä, S. (2016). Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education*, 17(1), 25-38.

# Do Different Goals Affect the Configuration of University Students' Internal Collaboration Scripts? Results of an Epistemic Network Analysis Study

Tugce Özbek, Martin Greisel, Christina Wekerle, Ingo Kollar  
tugce.oezbek@uni-a.de, martin.greisel@uni-a.de, christina.wekerle@uni-a.de, ingo.kollar@uni-a.de  
University of Augsburg

**Abstract:** Computer-supported collaborative learning (CSCL) offers a great potential for student learning. However, its successfulness is influenced by learners' internal collaboration scripts. Drawing from Script Theory of Guidance (SToG), which posits that internal collaboration scripts are dynamically configured based on learners' goals, this study explores how an induction of learning and performance goals influences the selection and sequence of activities learners perform during collaboration.  $N=233$  pre-service teachers were asked to collaboratively analyze a classroom situation. The instructions included information on the importance of this task either for students' competence development (learning goal condition), or for their performance (performance goal condition). While we found no significant differences regarding the change of selected scriptlets, Epistemic Network Analysis revealed distinct configurations in the sequence of scriptlets for learning and performance goal conditions. The results partly support SToG's configuration principle, emphasizing the role of situational goals in shaping internal collaboration scripts.

## Aims

Computer-supported collaborative learning (CSCL) serves as a powerful instructional method to support student engagement in high-level socio-cognitive processes (e.g., Chen et al., 2018). The way learners act in CSCL environments can be considered as being influenced by two forces that mutually interact with each other: (a) the design of the learning environment, and (b) the learners' learning prerequisites. With respect to (a), CSCL research has accumulated numerous insights, for example on the effects of different kinds of scaffolds on collaborative learning processes and outcomes. With respect to (b), there is considerably less empirical research. One prerequisite that has particularly been suggested in the context of research on CSCL scripts (i.e., scaffolds that structure the collaboration process through the provision, specification and distribution of learning activities and roles within small groups) is the learners' internal collaboration scripts. According to the Script Theory of Guidance (SToG; Fischer et al., 2013), internal collaboration scripts are cognitive structures of individuals that guide them in the way they understand and act in collaborative learning situations. According to the SToG, internal collaboration scripts consist of different knowledge components (play, scenes, roles, and scriptlets) that are dynamically configured in learners' memory. One central principle of the SToG refers to the assumption that this configuration is influenced by the current goals of the learners (internal script configuration principle).

Yet, it is striking that this principle has hardly been tested empirically so far. In fact, learners can have very different goals when collaborating and it is not clear exactly how these affect the configuration of their internal collaboration script (Pintrich, 2000a). A prolific model to conceptualize goals in the context of teaching and learning is Achievement Goal Theory (Dweck & Leggett, 1988; Elliot, 2005). However, so far there is only limited research on how goals are related to the configuration of the learners' internal collaboration scripts. For this reason, in this paper we investigate how different goals, respectively their induction, are related to learners' internal script configuration in the context of CSCL.

## How internal collaboration scripts shape within-group collaboration – the Script Theory of Guidance

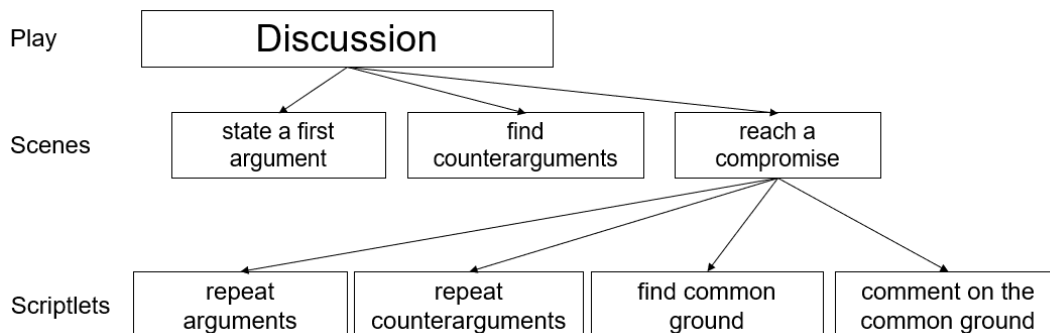
Based on recent meta-analyses, CSCL offers a vast potential to support learners' academic achievement (e.g., Chen et al., 2018). From a theoretical point of view, there are many potential benefits of collaborative learning, even without the support of digital technologies, which include academic (e.g., fostering critical thinking), social (e.g., developing social skills) or affective-motivational aspects (e.g., reduced anxiety; see Johnson & Johnson, 1989; Laal & Ghodsi, 2012). However, there may also occur problems during collaborative learning, such as individual learners not participating in collaboration or learners not actually working together but rather splitting the tasks among themselves (Roberts, 2005; Salomon, 1992). Especially in such circumstances, digital technologies may support the collaborative learning process, for example by providing tools to organize learners'

ideas and contributions, to provide resources or guidance to structure the collaboration process (Stahl et al., 2006; Suthers & Seel, 2012).

A theoretical approach that conceptualizes such guidance is the Script Theory of Guidance (SToG; Fischer et al., 2013), referring to such guidance as “external scripts”. External collaboration scripts provide group learners with guidance on the kinds and sequence of activities and roles they are supposed to take over during collaboration, often supporting their execution via prompts or other kinds of scaffolds. SToG however assumes that what actually happens during collaboration is not only influenced by such external scripts, but also by the learners’ internal collaboration scripts. According to the SToG, internal collaboration scripts are configurations of knowledge components in the learner’s cognitive system that determines how they understand and act during collaboration. Based on Schank’s (1999) dynamic memory approach, a basic tenet of the SToG is that internal scripts consist of configurations of four different script components (see Fig. 1): (a) The “play” component includes knowledge about the kind of the situation an individual experiences, e.g., a discussion held in a chat forum or the joint writing of a blog post. Once a learner has (consciously or unconsciously) selected a specific “play”, this “play” then binds together a set of (b) “scenes”. These include the person’s knowledge about the different situations that typically make up the play. In a discussion, for example, a learner’s “discussion” play might for example include a scene in which the group collects information to develop arguments, while another scene might be to exchange arguments. Once a certain scene is activated, the person also has expectations on what kinds of activities are typically part of that scene. Knowledge about these activities as well as the sequence of different activities that are likely to occur during that scene, according to the SToG, is represented in so-called (c) “scriptlets”. In the scene “reaching a compromise”, for example, the first scriptlet might be the summary of the most important arguments at the beginning. Finally, learners hold knowledge about different kinds of (d) “roles”, i.e., about the question what kinds of activities are likely to be taken over by what person in a collaborative situation. Similar to a theatre play, roles can extend over several scenes and include several activities.

**Figure 1**

*Example of an Internal Collaboration Script of a Learner During a Discussion*



*Note: Only a few examples of possible scriptlets are shown. Theoretically, every scene is connected to a specific set of scriptlets.*

## The Internal Script Configuration Principle within the Script Theory of Guidance

A central assumption of the SToG is that through experience, learners acquire a range of plays, scenes, scriptlets, and roles, and that these knowledge components are dynamically stitched together in each new situation. This idea lays the foundation of the so-called configuration principle of the theory. It states that “How an internal collaboration script is dynamically configured by a learner from the available components to guide the processing of a given situation, is influenced by the learner’s set of goals and by perceived situational characteristics” (Fischer et al., 2013, p. 57–58).

As described, the internal collaboration script consists of different knowledge components of collaborative learning that are considered to be very flexible in the way they are combined with each other. This means that in any situation, a learner has different plays, scenes, scriptlets, and roles available in memory that they can apply to make sense of the current situation. Even small changes in the situation and changing requirements can result in a quick (and very often subconscious) adaptation and new configuration of the internal collaboration script components. For example, certain tool features such as a flashing cursor might indicate an opportunity to enter a text, making entries by learners more likely than if there was no such flashing cursor (Fischer et al., 2013; Kirschner et al., 2008; Schank, 1999).

Yet, not only situational characteristics (i.e., external factors) may influence a learner's script configuration, but also factors that lie within the person matter in this respect. This refers in particular to learners' goals. According to the configuration principle, a learner's goal can, on the one hand, influence the *selection* of script components, i.e., plays, scenes, scriptlets, and roles. This means that the learner is likely to choose or act out the script components that are most useful to pursue their current goal and may yield activities that are conducive to this goal and inhibit activities that are not. On the other hand, these very flexible configurations of knowledge components may also influence how these different activities are *sequenced* (e.g., prepare arguments, debate, compromise). Yet, also such sequences can be dynamically reconfigured (i.e., the order of the components can be changed) depending on perceived changes in the situation (including the learning partners' activities) or in the learner's goals. For example, if a learner notices that their learning partner does not seem to exert effort during collaboration, and if they have the goal to get the task done anyway, they may de-activate scriptlets that would guide them to ask their learning partner for input, and replace this scriptlet by a scriptlet "solve the task alone".

The authors provide evidence for the influence of goals on the configuration of learners' internal collaboration scripts by referring to a study by Pfister and Oehl (2009). Their study addressed the question of how goal focus, task type, and group size influence synchronous net-based collaborative learning discussions. For this, they varied the goal focus of the learners in so far as one group should follow an individual focus (i.e., they received rewards based on their individual performance), or a group focus (i.e., they received rewards based on their group's performance). Results indicated that learners with the group focus used more supportive features of the tool than learners in the individual focus. Fischer et al. (2013) interpret this finding in a way that the different focus of the learners has led to a (re-)configuration in their internal scripts, as represented in learners' use of different tool functions.

However, a couple of limitations of this study and of Fischer et al.'s (2013) interpretation need to be noted here. First, the (re-)configuration of the internal script is only inferred indirectly (from the use of a certain feature by the learners), rather than measured directly. To do that, it would be necessary to apply methods that lay the kinds of internal script components as well as their sequence open. Second, the authors of the study did not examine how the (initial) internal script of the learners was structured. Therefore, it is also not possible to assess to what extent this script and its components actually reconfigured. Third, "goal focus" was defined and varied by the authors as a distinction between individual-oriented and group-oriented performance. While this is certainly one way to think about goals, there is a vast amount of research on achievement goals that can be drawn on to differentiate different kinds of goals that have been shown to matter for learning. This research will be discussed in more detail in the next section.

## Learners' goals during CSCL

From a theoretical perspective, goals describe a standard by which learners can assess their learning progress and initiate regulatory processes accordingly (Pintrich, 2000b). In particular, achievement goals refer to the goals that a learner pursues in learning and performance contexts. They describe their purpose for engaging in competence-related behavior (Elliot, 2005; Elliot & Fryer, 2008). At a global level, the theory distinguishes between two types of goals. Firstly, learners may have so-called *learning goals*, which means that they are particularly motivated to engage in learning because they focus on improving their competence. Secondly, they may also display *performance goals*; for learners with these goals, it is particularly important to engage in learning in order to demonstrate their performance or outperform others (Heyman & Dweck, 1992). In accordance with this, empirical research has often shown positive effects of learning goals on various learning processes and outcomes, whereas performance goals show more of a mixed picture. Some studies indicate positive, some negative effects on learning processes and outcomes and some do not reveal a clear pattern (Daumiller, 2023; Harackiewicz et al., 1998; Payne et al., 2007).

Although these findings already indicate that different achievement goals may lead to different learning processes and outcomes, research that looks at their direct impact on the configurations of learners' internal collaboration scripts seems to be lacking. Nevertheless, one might expect different achievement goals to have a particular impact on learners' scriptlets, i.e., on the knowledge they activate regarding the kinds and sequences of activities that are likely to occur resp. to be acted out during collaboration. In terms of the activities performed during CSCL, a pronounced learning goal might encourage learners in a collaboration to ask more questions or make more explanations, for example. In contrast, learners with pronounced performance goals, could be more likely to use impression management techniques and perhaps only want to appear competent, for example by using subject-specific language very deliberately (Greisel et al., 2023). Yet, empirical evidence on these issues seems to be missing so far.

## Research question and hypotheses

In sum, various evidence from the field of CSCL, but also from research on achievement goals, indicates that different goals of learners may influence how learners act in collaborative learning. On this basis, the SToG also argues that learners' goals in CSCL lead to an activation of specific script components. However, so far there has been little research investigating this configuration and the actual change in the internal script as a function of the presence or absence of different achievement goals. Therefore, in the present study, we actively manipulated learners' goals either in the direction of an actualization of *learning goals* or of *performance goals* and investigated the effects of this manipulation on their internal collaboration scripts (more precisely, the scriptlets) when working on a CSCL task.

Our research question was: Do different kinds of achievement goals (learning goals vs. performance goals) influence the configuration of the internal collaboration script? We hypothesized (H1) that learners in the learning goal condition would *select* different scriptlets to guide their collaboration than learners in the performance goal condition. Furthermore, we assumed that not only the selection, but also the *sequence* of scriptlets would differ depending on the kind of achievement goal that is induced (H2).

## Method

### Participants and design

A total of  $N = 233$  pre-service teachers participated in the study, who were on average 22 years old ( $M_{Age} = 22.30$ ,  $SD_{Age} = 2.30$ ), mostly female (72%) and in their fifth semester ( $M_{Sem} = 4.99$ ,  $SD_{Sem} = 1.14$ ). They were enrolled in a teacher education program for elementary school teachers, middle school teachers, high school teachers, and secondary school teachers of various subjects. The study was embedded as a compulsory part of a course in educational psychology for pre-service teachers. However, the students were free to decide whether they wanted to participate in the scientific data collection. They received no reward for their participation. Their task was to analyze an authentic, written case that described a problematic classroom situation and a teacher's efforts to solve those problems.

For collaboration, the students used the collaboration tool "coLearn!". This tool serves to structure collaborative learning with external collaboration scripts by assigning roles, specifying prompts and providing materials. To investigate how different achievement goals impact internal collaboration script configuration, we established a  $1 \times 2$  between-subjects design with the conditions "induction of learning goal" and "induction of performance goal". The conditions differed in that their external scripts (that is, the instructions within the tool) included prompts that contained elements of the respective goal (e.g., for learning goal condition, that completing a respective task would be very important in order to expand one's skills, or, for performance goal condition, that their performance will be evaluated by their instructor). Participants were randomly assigned to dyads and to one of the two conditions mentioned before.

### Procedure

During pretest, we measured students' initial internal collaboration scripts. The students then worked for three weeks using the collaboration tool "coLearn!". Then, they were grouped into pairs. First (Week 1), they were instructed to analyze a case vignette individually that described a teacher who faced different kinds of problems during her lesson, using one of two scientific educational theories (Cognitive Load Theory by Sweller, 2011, or ICAP framework by Chi & Wylie, 2014). Afterwards, these analyses were swapped between the students within a dyad, and the students were instructed to evaluate the analysis of their respective partner and to expand on it with the help of the respective other theory (Week 2). Afterwards, the students received this evaluation and elaboration from their peer and were asked to revise their original analysis on this basis (Week 3). After the collaboration phase, the students' internal collaboration scripts were measured again.

### Operationalization of the independent variable

As described, the conditions differed from each other with respect to the presentation of statements in the external script that were integrated into the collaboration tool. Each week, a new page with instructions and entry fields was displayed in the tool. In addition to specific instructions regarding the case analysis, the prompts contained a specific goal induction. In the *learning goal* condition, after instruction, the task was labeled as "important in order to improve one's own competencies and to successfully cope with problems in later professional life". In addition, working with the tool was explicitly framed as a learning opportunity in this condition. In contrast, in the *performance goal condition*, it was stated that the task was "important in order to achieve good grades". In

addition, working with the tool was framed as an “important opportunity to prepare for exams”, and it was stated that tutors would check the assignments later.

### Assessment of learners’ internal collaboration scripts

In order to assess the learners' internal collaboration scripts, prior to using the collaboration tool, in the pre-test participants described how they would generally carry out a collaborative analysis of a problematic classroom situation. They were given a total list of 40 activities (e.g., "asking questions", "reflecting on the theory" etc.) to choose from, representing different scriptlets that could be carried out as part of such a collaboration. The participants were instructed to drag and drop the activities they would perform from the list and put them in the order in which they would perform them. In this way, we captured their internal collaboration scripts regarding a collaborative case analysis considering the specific sequence of scriptlets. In the posttest, participants were instructed to select and arrange activities from the same set of activities using drag-and-drop and arrange them according to how they actually carried them out during collaboration.

To measure the extent of reconfiguration of participants’ internal collaboration scripts, we checked what activities (scriptlets) a participant added or omitted at post-test compared to the ones they selected at pre-test. Each activity was coded with a 1 if it was omitted or added from pre to posttest and with a 0 if it remained the same (i.e., both times selected or not selected). We summed these changes separately for three factors to ease interpretation of pre-to-post changes. We identified these factors with an exploratory factor analysis of tetrachoric correlations between the dummy-coded pre-test activities using weighted least squares and an oblimin rotation. The least frequent activities ( $n < 10$  in pre- or post-test) were dropped before the analysis. The number of factors was determined using the post-test activities based on a scree plot, a map test and the VSS complexity 2 criterion. The resulting three factors were: task-related activities (e.g., “read case”), cognitive learning strategies (e.g., “imagine practical applications of new concepts”), and social learning strategies (e.g., “help peer”).

### Statistical analyses

To test whether the reconfiguration of the internal collaboration scripts differed between the change of selected scriptlets (H1), we conducted separate ANOVAs for scriptlets referring to task-related activities, cognitive learning strategies, and social learning strategies. Regarding the sequence of scriptlets (H2), we conducted an Epistemic Network Analysis (ENA; Shaffer et al., 2016) with the activities reported in the post-test. We used a moving stanza window which was set to span across seven activities. Before the analysis, we dropped the four least frequent activities ( $n < 10$  in pre- or post-test). Two further activities were dropped after the first analysis which were not connected to the rest of the network and represented outliers. To test H2, we report the subtraction network which compares the networks of learning and performance goal conditions and the corresponding  $t$ -Tests which compare the centroid values of both groups.

### Results

To test H1 regarding the selection of the scriptlets in the pre- and post-test, we conducted three ANOVAs, one for each category of activities we had identified based on the factor analysis described above. There was no significant effect of condition with regard to cognitive learning strategies ( $F(2, 231) = 2.04, p = .13; \eta^2 = 0.02$ ), task-related activities ( $F(2, 231) = 0.73, p = .48; \eta^2 = .0063$ ), or social learning strategies, ( $F(2, 231) = 0.64, p = .52; \eta^2 = .0056$ ). Thus, there were no significant effects of the type of achievement goal that was induced on the kinds of changes of activities (scriptlets) participants mentioned to have used during collaboration.

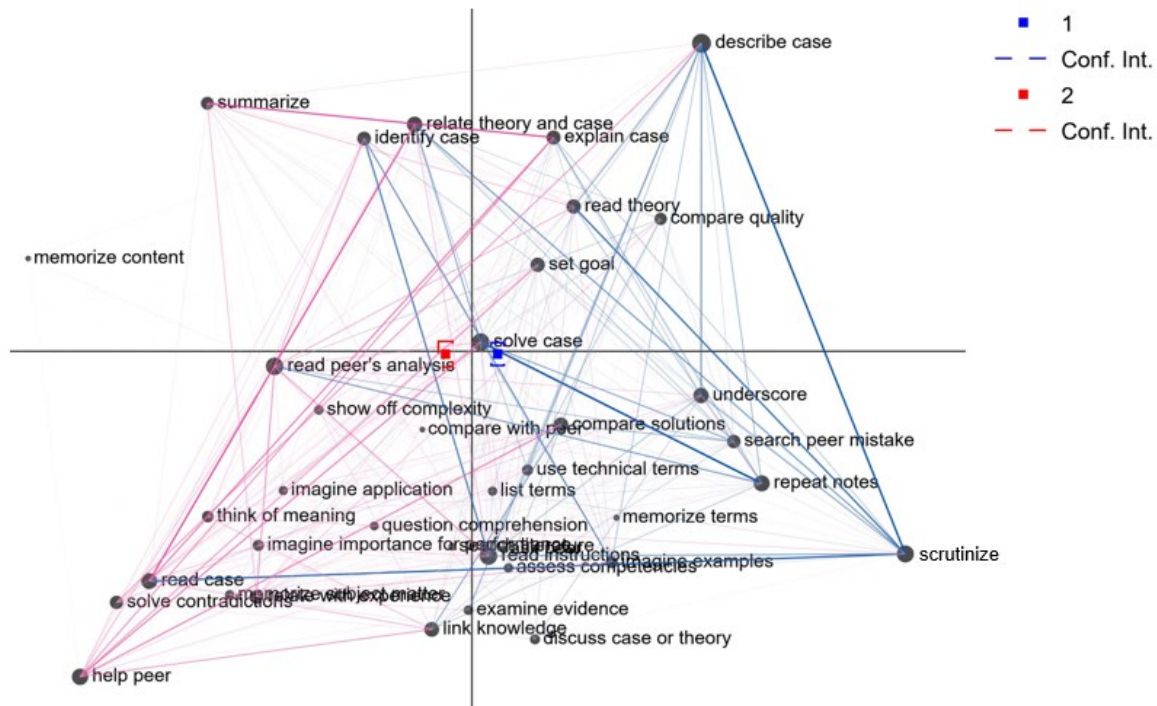
To answer RQ2, we conducted an ENA to compare the networks of scriptlets in each condition in the posttest (learning vs. performance goals induced). The mean centroid value for scriptlets in the epistemic network of the learning goal condition was significantly different from the mean centroid value in the network of the performance goal condition,  $t(148.57) = 10.76, p < .001$ , Cohen’s  $d = 1.77$ , indicating that the true difference in means is not equal to 0. Consequently, the results indicate that there were differences with respect to the sequence of the scriptlets selected between the conditions in which learning and performance goals were induced.

Subtracting the networks (Fig. 2) revealed that the network of the performance goal condition (red), in comparison to the learning goal condition (blue), displayed stronger connections between the scriptlets “read case”, “read peer’s analysis”, “relate theory and case”, “summarize” and “help peer”. In contrast, the learning goal condition showed a much stronger connection particularly between the “solve case” and “repeat notes” scriptlets. There were also comparatively stronger connections between the scriptlets “describe case”, “scrutinize” and “read case”.



**Figure 2**

*Epistemic Network Analysis of the Difference Between the Networks of Scriptlets in Learning Goal Condition (1/blue) and Performance Goal Condition (2/red)*



## Discussion

The SToG (Fischer et al., 2013) assumes that learners' goals affect the configuration of learners' internal collaboration scripts. However, this principle has not yet been tested directly empirically. Therefore, we investigated whether the induction of different achievement goals (Dweck & Leggett, 1988) affects the selection/change (H1) and sequence (H2) of the scriptlets that students select during a collaborative task.

Regarding H1, we assumed that the induction of different kinds of achievement goals would affect the kinds of scriptlets learners select during collaboration. Based on a factor analysis, we were able to identify three factors according to which the scriptlets could be clustered: scriptlets that refer to cognitive learning strategies, scriptlets that refer to task-related activities, and scriptlets that refer to social learning strategies. On this basis, we distinguished the change in the learners' initial internal collaboration script and the actual activities reported in the posttest. However, none of these ANOVAs indicated significant differences with respect to the change in the selected activities between the learning goal and performance goal conditions. Consequently, the hypothesis that different kinds of achievement goals would have an impact on the selection of scriptlets must be rejected, at least on the basis of the evidence regarding H1. This may mean that the change regarding the kinds of selected scriptlets is quite resistant to induced goals (at least with regard to learning and performance goals). Moreover, this may again confirm rather mixed results with regard to performance goals or evidence that learning and performance goals could even correlate positively (Daumiller, 2023). However, a meta-analysis on goal induction (Noordzij et al., 2021) shows that to induce learning goals, it is important to relate this goal to a specific task (e.g., "While performing this task, it is your goal to... by ..."). In comparison to this, the prompts used in our study may have been too vague. Thus, the goal induction may have been too weak at this point to have caused a change in the selection of scriptlets.

With respect to H2, however, and in contrast to the selection of scriptlets, the results of the ENA showed significant differences in the configuration of the scriptlets between the learning and performance goal conditions in the posttest. This means that the participants in the different conditions specified significantly different sequences of scriptlets. This partially supports our hypothesis and can be seen as evidence in favor of the configuration principle, at least regarding the differing sequence of scriptlets. Through ENA, we can see that particular activities are mentioned more frequently in a specific order by the groups. As described, learners with

pronounced learning goals are particularly motivated to engage in learning because they focus on improving their competence. In contrast, for learners with pronounced performance goals it is particularly important to demonstrate their performance or outperform others (Heyman & Dweck, 1992). Considering the scriptlets embedded in the network from this point of view, one explanation could be that some cognitive learning strategies (e.g., "scrutinize", "solve case", or "repeat notes") are more strongly integrated in line with the prompted goal condition, which also implies the acquisition of competence as the primary goal of the learners. In contrast, in the performance goal condition (which particularly emphasizes performance in comparison to others), many activities associated with a social context also play an important role (e.g., "help peer" or "read peer's analysis"). This might suggest very goal-specific configurations and seems to be in line with research on achievement goals pointing to their context specificity (Daumiller, 2023). However, this requires further analyses, which could possibly also account for the previously identified clusters of activities.

## Limitations and conclusions

Of course, this study has limitations. First, it is important to note that students selected from a range of activities, which means that they were not free in their choice of scriptlets, as certain activities were already suggested to them. In contrast, however, it is quite conceivable that the students would also name other or further activities, possibly even more so if none were specified to them beforehand (Csanadi et al., 2021). Future studies could therefore include interviews, for example, to more validly capture script components and elicit their (re-) configuration more adequately (März et al., 2021). In this context, it is also important to emphasize that in the ENA, the initial internal collaboration script of the learners was not taken into account and therefore only the differences in the subsequently reported activities can be determined, but not in comparison to the initial internal collaboration script. Thus, it would also be worthwhile considering a more process-oriented approach and, for example, monitoring activities in real time in order to record the activities carried out as validly as possible. Furthermore, only scriptlets were examined as internal script components in this study. Thus, our data did not allow separating for different script levels. It is conceivable that learners already have had very heterogeneously elaborated scripts and therefore also responded differently to the external script (Kollar et al., 2007; Vogel et al., 2017). In this regard, the interaction of learners' internal script levels and/or prior collaboration skills and goals might be an interesting research gap to look at in future studies.

Another limitation is that we only manipulated two (achievement) goals in this study. In fact, there are many different kinds of achievement goals learners may have that could also be taken into account (e. g., avoidance goals; Daumiller, 2023). On top of that, research on achievement goals not only suggests that goals can be very situation-specific, but learners can also pursue multiple goals. This might also indicate that different and multiple goals might be particularly important in different collaboration scenarios (e.g., relational goals). Future research should therefore also include or control for further goals of the learners.

Nevertheless, the results of the ENA indicate that learners' internal collaboration scripts are configured differently depending on the induction of learning or performance goals. These results support the SToG's configuration principle that learners' internal scripts are configured depending on (situational) goals. This constitutes an important step in the empirical validation of the model and contributes to our understanding of internal collaboration scripts. A closer look at the scriptlets also provided further insight into how certain activities are configured depending on specific goals. The fact that there are significant differences in the sequence, but not in the change of the selected scriptlets, might indicate that the learners' internal script should be examined on a rather fine-grained level. The manipulation of goals resulting in differences in learners' internal collaboration scripts highlights the importance of integrating motivational prompts in CSCL. For example, teachers may want to integrate prompts that target specific goals into the design of CSCL environments, possibly leading to script configurations and activities that are particularly conducive to student learning. Especially in the context of CSCL, the adaptability of digital technologies should be utilized, for example, by offering different goal settings to the instructors, for instance, specific competence-related goals. The study thus holds important implications regarding the design of CSCL-environments and provides stronger evidence for the SToG configuration principle.

## References

- Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research*, 88(6), 799–843.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.

- Csanadi, A., Kollar, I., & Fischer, F. (2021). Pre-service teachers' evidence-based reasoning during pedagogical problem-solving: Better together? *European Journal of Psychology of Education, 36*, 147–168.
- Daumiller, M. (2023). *Achievement Goals: The Past Present, and Possible Future of Achievement Goal Research in the Context of Learning and Teaching*. PsyArXiv. <https://doi.org/10.31234/osf.io/xn2v5>
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256.
- Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52–72). Guilford Press.
- Elliot, A. J., & Fryer, J. W. (2008). The goal construct in psychology. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science* (Vol. 18, pp. 235–250). Guilford Press.
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational psychologist, 48*(1), 56–66.
- Greisel, M., Melzner, N., Kollar, I., & Dresel, M. (2023). How are achievement goals associated with self-, co-, and socially shared regulation in collaborative learning?. *Educational Psychology, 43*(4), 384–402.
- Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. *Motivation and Emotion, 16*, 231–247.
- Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin, 136*(3), 422.
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Interaction Book Company.
- Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708–721.
- Kirschner, P. A., Beers, P. J., Boshuizen, H. P., & Gijsselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior, 24*(2), 403–420.
- Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia-Social and Behavioral Sciences, 31*, 486–490.
- März, E., Wessels, I., Kollar, I., & Fischer, M. R. (2021). “I just stand around and look friendly”–Comparing medical students' and physicians' ward round scripts. *Medical Teacher, 43*(5), 560–566.
- Noordzij, G., Giel, L., & van Mierlo, H. (2021). A meta-analysis of induced achievement goals: The moderating effects of goal standard and goal framing. *Social Psychology of Education, 24*, 195–245.
- Pfister, H.-R., & Oehl, M. (2009). The impact of goal focus, task type and group size on synchronous net-based collaborative learning discourses: Impact of goal focus, task type and group size. *Journal of Computer Assisted Learning, 25*(2), 161–176. <https://doi.org/10.1111/j.1365-2729.2008.00287.x>
- Pintrich, P. R. (2000a). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544.
- Pintrich, P. R. (2000b). The role of goal orientation in self-regulated learning. In M. Boekaerts, P.R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Elsevier.
- Roberts, T. S. (Ed.). (2005). *Computer-supported collaborative learning in higher education*. Idea Group Pub.
- Salomon, G. (1992). What does the design of effective CSCL require and how do we study its effects? *Acm Sigcse Outlook, 21*(3), 62–68.
- Schank, R. C. (1999). *Dynamic memory revisited*. Cambridge University Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45.
- Stahl G., Koschmann T., Suthers D. (2006). Computer-supported collaborative learning: An historical perspective. In Sawyer R. K. (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, England: Cambridge University Press.
- Suthers, D. D., & Seel, N. M. (2012). Computer-supported collaborative learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning*, (pp. 719–722). Springer, New York.
- Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Elsevier.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*, 477–511.

# Investigating Collaborative Problem Solving Behaviors during STEM+C Learning in Groups with Different Prior Knowledge Distributions

Caitlin Snyder, Vanderbilt University, caitlin.r.snyder@vanderbilt.edu  
Cai-Ting Wen, Soochow University, leafhere@gmail.com  
Nicole M. Hutchins, University of Florida, nicole.hutchins@ufl.edu  
Caleb Vatal, Vanderbilt University, caleb.m.vatal@vanderbilt.edu  
Chen-Chung Liu, National Central University, celiu@cl.ncu.edu.tw  
Gautam Biswas, Vanderbilt University, gautam.biswas@vanderbilt.edu

**Abstract:** In collaborative problem-solving (CPS), students work together to solve problems using their collective knowledge and social interactions to understand the problem and progress towards a solution. This study focuses on how students engage in CPS while working in pairs in a STEM+C (Science, Technology, Engineering, Mathematics, and Computing) environment that involves open-ended computational modeling tasks. Specifically, we study how groups with different prior knowledge in physics and computing concepts differ in their information pooling and consensus building behaviors. In addition, we examine how these differences impact the development of their shared understanding and learning. Our study consisted of a high school kinematics curriculum with 1D and 2D modeling tasks. Using an exploratory approach, we performed in-depth case studies to analyze the behaviors of groups with different prior knowledge distributions across these tasks. We identify effective information pooling and consensus building behaviors in addition to difficulties students faced when developing a shared understanding of physics and computing concepts.

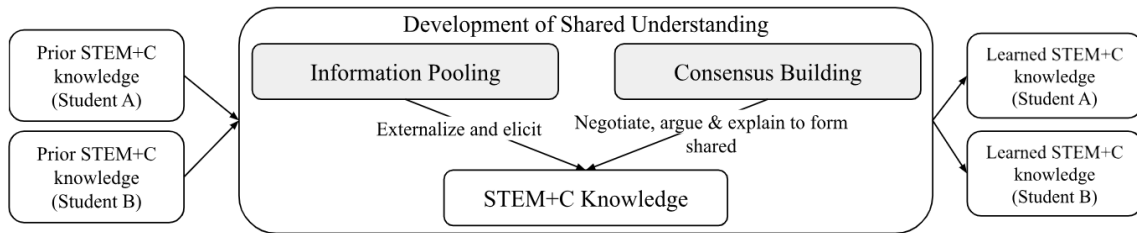
## Introduction

In collaborative problem solving, students jointly construct knowledge through conversations to reach a *shared understanding* and apply it to problem-solving tasks (OECD, 2015). Our study focuses on groups working collaboratively to build computational models in an open-ended learning environment (OELE). Effective computational modeling and problem solving necessitate integrating the STEM and computing domains (NRC, 2012). Building on prior research that highlights the influence of prior knowledge on learning in single domains (e.g., Zambrano et al., 2019), we investigate how groups' prior knowledge in science and computing influenced their Collaborative Problem Solving (CPS) behaviors as they constructed computational models in kinematics.

Our analysis of students' CPS builds on research that has identified shared understanding as a key component of effective collaborative problem solving (e.g., Baker, 2015; OECD, 2015). Specifically, we study how students develop their shared understanding through: (1) *information pooling*, where students externalize and elicit domain-specific knowledge from members in the group; and (2) *consensus building*, where students use arguments and explanations to negotiate and create shared knowledge and apply it to their problem solving tasks (Meier et al., 2007). Clearly, these conversations combine domain-specific information and social interactions (e.g., elicit knowledge from their partners by asking questions and negotiate differences to form a consensus; Weinberger & Fischer, 2006) to discuss STEM and computing concepts and construct their computational models (Snyder et al., 2019). We predict that these interactions will vary depending on students' prior knowledge within groups, as their existing understanding of the problem impacts the information they seek and share during collaborative efforts. Similarly, research indicates that differences in initial knowledge significantly affect argumentation skills and consensus building (Yang et al., 2015). We utilize this framework (Figure 1) to examine students' collaborative problem-solving behaviors in STEM+C learning.

Our computer based OELE targets synergistic learning in the science and computing domains. In this work, we used a 1D and 2D high school kinematics curriculum that combines inquiry activities, instructional tasks, formative assessments, and model building activities. Instructional and inquiry activities, along with formative assessments, help students learn the primary physics and computing concepts and relations between these concepts. At the end of each unit, students are given a challenge task, which requires them to build a comprehensive computational model. By combining students' conversations and their model building activities in the environment, we adopt an exploratory case study approach to analyze students' information pooling and consensus building behaviors in groups with different types of prior knowledge distributions. Our data comes from students' work in the 1D and 2D challenge tasks and pre-post assessments in science and computing.

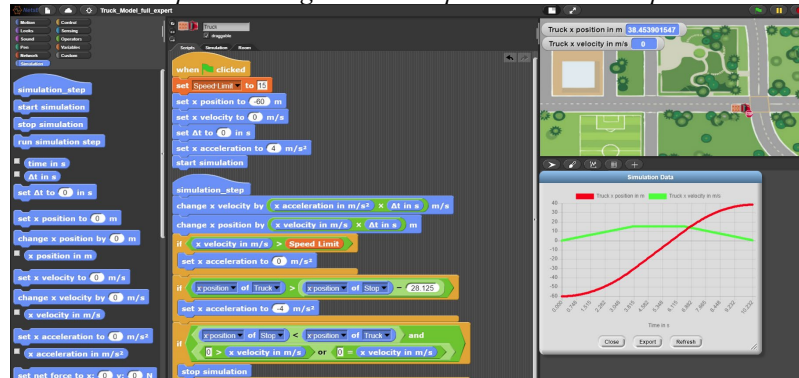
**Figure 1**  
CPS Framing: Group's Development of a Shared Understanding of STEM+C Knowledge



## STEM+C Learning Environment

Students work collaboratively in Collaborative Computation STEM (C2STEM; see Figure 2), our open-ended computational modeling environment that adopts a modular approach to help students progressively learn complex science and computing concepts in specific curricular domains, e.g., kinematics (Hutchins et al., 2020). Within C2STEM, students create partial or complete models to study the movement of the objects. Along with animation and variable inspection functions that are displayed on the simulation stage, students have access to graphing and table tools that are updated dynamically at each simulation step to help them debug their evolving models. Students create these models by developing and leveraging their understanding of kinematics (e.g. relationship between position, velocity, and acceleration) and computing knowledge (e.g., initializing and updating variables and applying conditional constructs). *Synergistic learning*, i.e., the simultaneous learning of science and computing, has been shown to be effective in developing successful solutions (Hutchins et al., 2020) but students may also have difficulties, such as transferring their STEM knowledge to computing constructs to build their computational model (Basu et al., 2016) that can be mitigated through collaboration.

**Figure 2**  
C2STEM Example Challenge Task Group Solution and Graph Tool



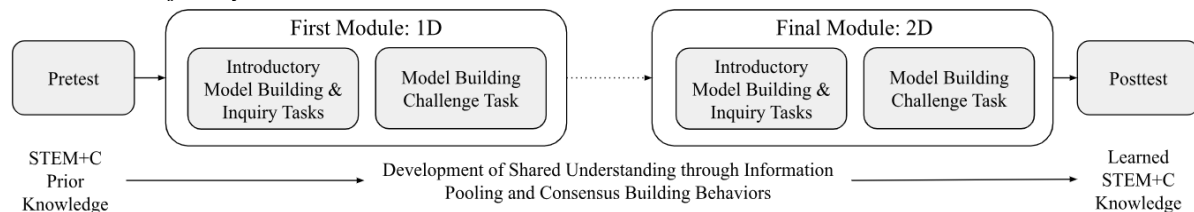
## Methods

Each module in our curriculum (Figure 3) comprises inquiry, computational modeling, and formative assessment tasks, developed using a systematic evidence-centered design (ECD) approach (Mislevy & Haertel, 2006). Inquiry tasks utilize the CoSci platform (<https://cosci.tw>), that provides students a scaffolded simulation-based learning environment to explore physics variables and dynamic processes (Wen et al., 2018). In the 1D challenge, students model a truck's motion speeding up to a speed limit, cruising at that speed limit, and then slowing to a stop at a STOP sign. To construct the model, students had to translate their physics understanding of the relations between position, velocity and acceleration to a computational form that included initializing the necessary variables and modeling the dynamics of the truck movement by updating variables under different conditions. In the final 2D challenge, they simulate a drone dropping packages onto specific targets, applying physics principles like gravity to construct computational models.

Our analysis targets the research question: *How do groups prior knowledge distributions affect their information pooling and consensus building behaviors during computational modeling tasks and how do these behaviors relate to students' STEM+C learning?* We answer this research question adopting an exploratory case study approach where we analyze the differences between groups that were (1) *balanced*, i.e., one student in the

pair had high prior knowledge in physics and low prior knowledge in computing, whereas the second student had high prior knowledge in computing and low prior knowledge in physics; and (2) *unbalanced* where one student in the pair had high prior knowledge in physics and computing, and the second student had low prior knowledge in both domains. Specifically, we studied each groups' information pooling and consensus building behaviors across two model-building tasks (1D and 2D acceleration) to understand how their behaviors evolved over the course of the curriculum. We also related these behaviors to students' STEM+C learning gains measured by their pre- to post-test learning gains.

**Figure 3**  
*Curriculum Trajectory*



## Students and Their Data

Our research team conducted a two-month-long study, working with 10<sup>th</sup> grade high school students, aged 14-15, for two hours a week in a classroom in the United States. None of the students had taken a high school physics course, but some had been introduced to basic kinematics in introductory science classes. Their background in computing varied. 27 students were divided into 13 groups (12 dyads and one triad) assigned based on their pretest scores. The student with the highest total pretest score (i.e., the sum of their pretest scores in kinematics and computing) was matched with the student with the lowest pretest score and so on. The study was approved by the university Institutional Review Board. This included analyzing summative assessment data and computational models, and video and audio data collected using the OBS software on each group's shared laptop. Student conversations were transcribed using Otter.ai and then edited by two researchers. One student in the triad did not consent to data collection so we did not analyze the data for that group.

Each student was categorized as having high or low prior knowledge in physics and computing based on their pretest scores relative to the median. The pre- and post-tests contained four physics questions (17 points) and three computing questions (16 points) in multiple choice and constructed response formats. Students' pretest scores in physics ranged between [5,15] with a median score of 11 (SD = 2.37). Their computing scores ranged between [3,13] with a median score of 10.5 (SD = 2.80). When we looked at group scores, we had a total of 6 *unbalanced* groups, 3 *balanced* groups and 3 groups in which both students had low prior knowledge in physics. In this paper, we adopt an exploratory contrasting case study approach to compare the behaviors of *unbalanced* and *balanced* dyads as these groups had similar overall prior knowledge in physics and computing but the distributions of knowledge across the students differed. We chose two *unbalanced* and two *balanced* groups for in-depth analysis by considering the quality of the collected video and audio data.

## Analysis

To evaluate groups' social interactions during information pooling and consensus building, we coded students' utterances using the Weinberger & Fischer (2006) social modes framework (see Table 1). Additionally, we included an off-task label to code discourse that was not related to the computational modeling task. Two members of the research team coded the dialogue with good agreement (Cohen's inter-rater reliability kappa value of 0.77). From these codes, we extracted information pooling and consensus building segments of student dialog. To evaluate groups' domain-specific STEM+C knowledge integration during information pooling and consensus building, we coded each utterance based on the physics or computing concepts discussed in the utterance (Cohen's kappa value of 0.83). The physics concept codes included conversations about position, velocity, acceleration, displacement, and time, and the computing concept codes included  $\Delta t$ , control structures, initializing variables, updating variables, and conditional structures. Leveraging these coded utterances, we calculated a *synergistic score* for each information pooling and consensus building segment using following formula:  $SYN = \# Utterances_C - \# Utterances_{PHY}$ . This computed value was then normalized to the range [-1,1]; a value closer to 0 indicated high synergistic discourse (i.e., conversations in this segment included concepts in both domains), a value closer to -1 indicated more physics-focused conversation and a value closer to 1 indicated more computing-focused conversation. To evaluate groups' application of STEM+C knowledge, we scored each

group's final computational models using a predefined rubric. Learning gains were calculated based on a summative post-test in science and computing that was identical to the pre-test.

**Table 1**  
*Information Pooling and Consensus Building Coding Scheme*

Social Mode	Description
<b>Information Pooling:</b> “ <i>eliciting information and giving appropriate explanations</i> ” (Meier et al., 2007)	
Elicitation	One student is questioning another about information relating to the task
Externalization	Student(s) are articulating to the other by stating facts, observations and/or narrating their actions in the system
<b>Consensus Building:</b> “ <i>discussing and critically evaluating information in order to make a joint decision</i> ” (Meier et al., 2007)	
Conflict-oriented	During discussion, the students are disagreeing over their interpretation of a concept, model component, or what to do next.
Integration-oriented	During discussion, one student adds a new component to the discussion, integrates a concept/perspective, and/or applies the perspective proposed by the other student
Quick	During discussion, one student makes a suggestion, and their partner accepts it with no further discussion

## Results

Table 2 lists the groups' task scores, STEM+C synergistic integration scores, and social dimension metrics for information pooling and consensus building behaviors for the 1D and 2D challenge tasks. Note that most of the groups scored lower in the 2D task and there was a drop in computing performance in G1, G2, and G3. We argue this may be due to the increased computational complexity of this task as G2 and G3's physics performance stayed the same across the tasks while G4 raised their physics score, and subsequently their total score, in the final task. However, as discussed below, G4 received help from another group in the physics component of the final task. In the following subsections, we first contrast groups' CPS behaviors and conclude by analyzing individual students' STEM+C learning in the context of these behaviors and groups' different knowledge distributions. Students' prior knowledge categories were based on their pretest scores in Table 3.

## Contrasting Information Pooling and Consensus Building Behaviors

### Balanced Groups (G1 and G2)

The information pooling behaviors in G1 and G2 are relatively synergistic ( $|Avg\ SYN| \leq 0.08$ ) for both tasks. We hypothesize that the *balanced* groups leveraged synergistic information pooling because each partner externalized their knowledge in their high prior knowledge domain. For example, in the 1D task, S3 often leveraged his physics knowledge to externalize how he believed the model (the motion of the truck) should behave, while S4 often leveraged her computing knowledge to suggest the use of specific computing blocks to effectively simulate the behavior.

However, there were differences in G1 and G2's consensus building behaviors. In G2, consensus building differed between tasks. In the 1D task, conversations were more integrated-oriented (18%) compared to conflict-oriented (12%). Students tended to defer to their partner's ideas rather than challenge them. For example, in the 1D modeling task, S4 (who had high prior knowledge in computing) took the lead by identifying when they needed to use a *change block* (to update a variable based on the previous simulation value) or a *set block* (to set a variable to a specific value) and S3 (who had low prior knowledge in computing) followed along. This reliance on their partner's knowledge did not negatively impact task performance, since the group had the highest score on the 1D model (0.95). In the 2D task, conflict-oriented consensus building increased (22%), with instances of both students challenging each other, such as a disagreement over using a change block, where S3 (who had low prior knowledge in computing) correctly, disagreed saying “*no we just need to do set. Set y velocity*”, resulting in a discussion that concluded with S4 agreeing, “*Oh it has to be at 0, yeah you're right.*”

**Table 2**  
*Students' Task Scores, STEM+C Knowledge Synergistic Scores and Percentage of Social Interactions*

Groups		G1 (Balanced)		G2 (Balanced)		G3 (Unbalanced)		G4 (Unbalanced)	
Challenge Task		1D	2D	1D	2D	1D	2D	1D	2D
<b>Total Task Score</b>		0.95	0.63	0.97	0.84	0.89	0.79	0.95	0.97
<b>PHY Task Score</b>		0.94	0.67	1	1	0.94	0.94	0.94	1
<b>C Task Score</b>		0.95	0.6	0.95	0.7	0.85	0.65	0.95	0.95
<b>STEM+C Knowledge during Information Pooling and Consensus Building</b>									
<b>Avg SYN (SD) - Information Pooling</b>		0.02 (0.12)	-0.05 (0.16)	-0.08 (0.15)	-0.08 (0.16)	-0.20 (0.25)	-0.19 (0.30)	-0.10 (0.19)	-0.11 (0.16)
<b>Avg SYN (SD) - Consensus Building</b>		-0.06 (0.23)	0.07 (0.17)	-0.06 (0.17)	0.00 (0.16)	-0.05 (0.10)	-0.01 (0.32)	0.04 (0.15)	-0.09 (0.09)
<b>Social Interactions during Information Pooling and Consensus Building</b>									
<b>Information Pooling</b>	<b>Elicitation</b>	11%	12%	12%	22%	11%	11%	16%	5%
	<b>Externalization</b>	51%	44%	58%	45%	66%	67%	49%	77%
	<b>Total</b>	62%	57%	70%	67%	77%	78%	65%	82%
<b>Consensus Building</b>	<b>Conflict-oriented</b>	5%	9%	12%	22%	11%	11%	16%	5%
	<b>Integration-oriented</b>	25%	24%	18%	10%	10%	10%	19%	12%
	<b>Quick</b>	8%	10%	1%	1%	2%	1%	0%	0%
	<b>Total</b>	38%	43%	30%	33%	23%	22%	35%	18%

G1 also had an increase in conflict-oriented consensus building (5% to 9%) from the 1D to 2D task but they still favored integration-oriented consensus building behaviors in both tasks (25% and 29%). During these consensus-building segments, S1 often led with new ideas while S2 contributed. For example, when they were initializing the position of the packages, with S2 controlling the laptop, S1 made a suggestion about changing the value saying, “A little bit less, like 4.5 for now... we want it on top of the other packages” and S2 agreed, saying “Yeah we can move the other packages to put them [below]”. Interestingly, although their discussions remained synergistic, they shifted slightly from a physics focus in the 1D task to a computing focus in the 2D task (Avg SYN = -0.06 to Avg SYN = 0.07). We hypothesize that this may be attributed to the group struggling with the increased computational complexity in the 2D task. They had the lowest score out of all the groups (0.63), and their consensus building centered more around the computing component of the model. Overall, both G1 and G2 had synergistic information pooling and consensus building behaviors in both tasks. They primarily used integration-oriented consensus building behaviors on the 1D task. However, while G1 continued to have similar consensus building behaviors in the 2D task, G2’s behaviors evolved into more conflict-oriented discussions.

#### Unbalanced Groups (G3 and G4)

The information pooling behaviors for G3 and G4 were less synergistic and more physics-focused ( $-0.19 \leq \text{Avg SYN} \leq -0.10$ ). G3’s discourse was primarily information pooling focused on both tasks (77% and 78%, respectively) and was characterized by the high prior knowledge student (S6) narrating actions. For example, in G3, S6 primarily narrated the model construction actions with very little contributions made by S5 (low prior knowledge



student). During one segment, while debugging the truck slowing down segment, possibly to elicit collaboration from her partner, S6 expressed a lack of understanding, saying “*I don’t know what to do next because I’m confused*” but S5 did not respond and S6 resorted to an ineffective trial and error strategy. This lack of collaboration increased in the 2D task, where there was a 14-minute segment in which S6 tinkered with the model and made only five utterances whereas S5 did not make any. In fact, S6 attempted to collaborate with another group in the 2D task when her partner would not engage. Despite S6 having higher prior knowledge, the complexity of the model building tasks necessitated collaboration (Kirschner, et al., 2011). We hypothesize this lack of collaboration contributed to G3’s poor consensus building behaviors and low task performance, with the group scoring the lowest in the 1D task and second lowest in the 2D task.

G4 had comparable amounts of consensus building behaviors as G1 and G2 in the 1D task (35%). At the beginning of the task the low prior knowledge student, S7, took the lead in controlling the laptop mouse while the high prior knowledge student, S8, gave suggestions on what actions to take. They switched between information pooling behaviors (where S8 was narrating) and synergistic consensus building behaviors (Avg SYN = 0.04) as the two students often discussed specific suggestions made by S8. When the group switched control of the laptop, S7 stayed involved and elicited information from S8 as they performed actions. For example, when modeling the truck motion transitioning from cruising to slowing down, S7 asked, “*Question... What are we looking for here?*” with S8 clarifying the current goal, “*looking for how long we need to get this to cruise for because... you don’t know where to start decelerating...*”. During the 2D task their behaviors changed considerably as they had more information pooling behaviors (82%) with much less collaboration between the partners. Their struggles with a physics component of the 2D task, resulted in physics-focused information pooling and consensus building behaviors (Avg SYN = -.11 and -0.09, respectively). Eventually, they got another group to give them the answer.

In summary, across the two tasks, students exhibited three information pooling behavior types: (1) synergistic information pooling (G1 and G2, both students externalized knowledge); (2) information pooling externalized by one student after prompting from the other (G4-1D); and (3) non-collaborative information pooling in which one student primarily narrated actions (G3, G4-2D). There were also three consensus building behaviors: (1) primarily integration-oriented in which groups formed a consensus primarily through deferring to their partner but also added different ideas (G1, G2-1D, G4-1D); (2) primarily conflict-oriented consensus building in which groups developed enough shared understanding and individual knowledge to more easily challenge their partner (G1-2D); and (3) minimal consensus building overall (G3, G4-2D).

**Table 3**  
*Summative Physics (PHY) and Computing (C) Learning by Group*

Group	Student	Pretest			Posttest			Learning Gains (LG)		
		TOTAL	PHY	C	TOTAL	PHY	C	TOTAL	PHY	C
G1 ( <i>Balanced</i> )	S1	0.67	0.65	0.69	0.77	0.76	0.78	0.10	0.11	0.09
	S2	0.61	0.71	0.50	0.64	0.71	0.56	0.03	0.00	0.06
G2 ( <i>Balanced</i> )	S3	0.61	0.71	0.50	0.71	0.65	0.78	0.10	-0.06	0.28
	S4	0.65	0.59	0.72	0.76	0.71	0.81	0.11	0.12	0.09
G3 ( <i>Unbalanced</i> )	S5	0.58	0.65	0.50	0.59	0.65	0.53	0.01	0.00	0.03
	S6	0.74	0.71	0.78	0.83	0.88	0.78	0.09	0.17	0.00
G4 ( <i>Unbalanced</i> )	S7	0.39	0.53	0.25	0.42	0.35	0.50	0.03	-0.18	0.25
	S8	0.80	0.88	0.72	0.72	0.82	0.63	-0.08	-0.06	-0.09

## STEM+C Learning

Table 3 presents the summative assessment results. Students in the *balanced* groups, G1 and G2, had higher overall learning gains (Avg = 0.09) compared to the students in the *unbalanced* groups, G3 and G4 (Avg = 0.01). This result holds across all the *balanced* and *unbalanced* groups in the study as all the *balanced* groups had an average overall learning gains of 0.11 (SD = 0.04, n=3) and all the *unbalanced* groups had an average overall learning gains of 0.04 (SD = 0.12, n=6).

When considering the *balanced* groups’ STEM+C learning, the students had higher learning gains in the domains they started with low prior knowledge (i.e., low physics prior knowledge students S1 and S4 had learning gains of 0.11 and 0.12 in physics, respectively, while low computing prior knowledge students S2 and S3 had

learning gains of 0.06 and 0.28, respectively) suggesting that these groups successfully leveraged their partners' knowledge to develop their own individual knowledge. Interestingly, when considering the physics domain, the students with low prior physics knowledge (S1 and S4) ended up surpassing their partners' physics knowledge by the end of the study: in G1, S1 had a final PHY score of 0.76 while S2's score was 0.71 and in G2, S4 had a final PHY score of 0.71 while S3 had a final PHY score of 0.65. This is partly because S2 and S3 had minimal, if any, learning gains in physics. When investigating the pre-posttests in more detail, S2 in G1 was partially incorrect on a physics-focused 2D question on the pre- and post-test (S1's answer was incorrect on the pre but was partially correct on the post). Since this group struggled with the 2D challenge task, we hypothesize this may be why S2 had no positive learning gains in physics. In G2, S3 correctly answered a 1D graph question in the pre but incorrectly in the post, and S4 answered the same question incorrectly in the pre and post, suggesting that the group had a misunderstanding about graphs. In contrast, all groups improved in computing. Overall, these results suggest that while the balanced groups had overall learning gains, and particularly in their low prior knowledge domains, all students generally gained more computing knowledge.

When considering the *unbalanced* groups' STEM+C learning, the groups had difficulties developing knowledge in both domains. In the computing domain S5, S6, and S8 had minimal (0.03), no, and negative learning gains (-0.09), respectively. The low prior knowledge student in G4, S7, is the only one who showed learning gains in computing (LG.C = 0.25) but they also had the lowest pretest score (0.25) and ended with the lowest posttest score (0.50) in computing. Similarly, the high prior knowledge student in G3, S6, is the only student who gained physics knowledge (LG.PHY = 0.17). Overall, in the *unbalanced* groups knowledge development at the end of the curriculum was also unbalanced as in G3, the high prior knowledge student (S6) had learning gains in physics and their partner's learning gains were minimal, while in G4, the low prior knowledge student (S7) had learning gains in computing and S8 had negative overall learning gains (-0.08).

The *balanced* groups' summative assessments showed consistency in the pre-post answers (the same question incorrectly on the pre and post-test, suggesting a knowledge gap that was not addressed by the intervention or a question that was answered incorrectly on the pre but correctly answered on the post, suggesting that knowledge was gained through the intervention). But the *unbalanced* group (e.g., S8) showed mixed results (i.e., both correct to incorrect and incorrect to correct answers). Overall, the *unbalanced* pairing seems to have helped the low prior knowledge student gain knowledge in computing but negatively impacted the high prior knowledge student through the introduction of new knowledge misunderstandings.

Finally, when considering these STEM+C learning results in the context of their information pooling and consensus building behaviors, the results imply that the collaborative, synergistic information pooling and consensus building behaviors G1 and G2 exhibited led to effective learning overall. Like prior research that has identified disagreements are an important component to individual learning during CPS activities (e.g., Roschelle & Teasley, 1995), we hypothesize that the transition from integrated-oriented consensus building to more conflict-oriented consensus building from the 1D to 2D task helped G2's individual STEM+C learning (as they had more overall knowledge development than G1 who consistently had integration-oriented consensus building behaviors across both tasks). In addition, the integration-oriented consensus building behaviors G4 exhibited in the 1D task seem to be partially responsible for S7's gaining computing knowledge through model construction with suggestions from their partner. However, the results suggest that the less collaborative information pooling and minimal consensus building behaviors that G3 exhibited in both tasks, and G4 in the 2D task, negatively impacted STEM+C learning overall as both G3 and G4 had more difficulty developing knowledge.

## Discussion and conclusions

We found that when both partners participated in information pooling, they developed better shared understanding and STEM+C learning. Information pooling with contributions by both partners provides a base of shared facts that students can leverage to develop a shared understanding through consensus building (Baker, 2015). Our results imply that an increase in consensus building behaviors over time linked to pooled information leads to increased shared understanding through critical analyses of relevant STEM+C concepts. While integration-oriented consensus building had a positive impact, our results suggest that conflict-oriented consensus building is a key indicator of increased shared understanding by both partners, resulting in higher individual STEM+C knowledge gains. We also identified difficulties groups had in developing shared understanding, such as a lack of collaboration during information pooling and consensus building, and this negatively impacted STEM+C learning. Sometimes, groups' development of a shared understanding may cause new misunderstandings if students develop and integrate incorrect knowledge. When detected, a teacher (or agent) can intervene to suggest using more effective CPS behaviors.

Our results are consistent with previous research on prior knowledge distribution in groups. For example, Deiglmayr & Schalk (2015) found that knowledge interdependence among individuals with complementary prior

knowledge increases interactive engagement and fosters rich constructive discourse. We found that *balanced* groups overall exhibited more collaborative behaviors compared to the *unbalanced* groups. While previous research has shown that students with low prior knowledge in a domain perform better when working collaboratively as compared to when they work individually (Zambrano, et al. 2019), results on the impact of such unbalanced pairings on the high prior knowledge student are conflicted (Gijlers & De Jong, 2005; Zhang, et al. 2015). Our study confirms the conflicting STEM+C learning we see in the *unbalanced* groups. This exploratory analysis is limited due to its small sample size. In future work, we will extend such analysis to more dyads and include groups who lack prior knowledge overall (e.g., groups with low prior knowledge in physics). Future work will also leverage these results to develop supports to help students employ more effective CPS behaviors and combine their STEM and computing knowledge to construct computational models.

## References

- Baker, M. J. (2015). Collaboration in collaborative learning. *Interaction studies*, 16(3), 451-473.
- Basu, S., Biswas, G., Sengupta, P., Dickes, A., Kinnebrew, J. S., & Clark, D. (2016). Identifying middle school students' challenges in computational thinking-based science learning. *Research and Practice in Technology Enhanced Learning*, 11(1), 1-35.
- Deiglmayr, A., & Schalk, L. (2015). Weak versus strong knowledge interdependence: A comparison of two rationales for distributing information among learners in collaborative learning settings. *Learning and Instruction*, 40, 69-78.
- Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 42(3), 264-282.
- Hutchins, N. M., Biswas, G., Maróti, M., Lédeczi, Á., Grover, S., Wolf, R., ... & McElhaney, K. (2020). C2STEM: A system for synergistic learning of physics and computational thinking. *Journal of Science Education and Technology*, 29(1), 83-100.
- Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, 25(4), 615-624.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*.
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6-20.
- NRC - National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- OECD (2015). PISA 2015. Collaborative problem solving.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*. Springer, Berlin, Heidelberg.
- Snyder, C., Hutchins, N., Biswas, G., Emara, M., Grover, S., & Conlin, L. (2019). Analyzing students' synergistic learning processes in physics and CT by collaborative discourse analysis. In *Computer-supported collaborative learning*.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & education*, 46(1), 71-95.
- Wen, C. T., Chang, C. J., Chang, M. H., Fan Chiang, S. H., Liu, C. C., Hwang, F. K., & Tsai, C. C. (2018). The learning analytics of model-based learning facilitated by a problem-solving simulation game. *Instructional Science*, 46(6), 847-867.
- Yang, W. T., Lin, Y. R., She, H. C., & Huang, K. Y. (2015). The effects of prior-knowledge and online learning approaches on students' inquiry and argumentation abilities. *International Journal of Science Education*, 37(10), 1564-1589.
- Zambrano, J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2019). Effects of prior knowledge on collaborative and individual learning. *Learning and Instruction*, 63, 101214.
- Zhang, L., Kalyuga, S., Lee, C. H., Lei, C., & Jiao, J. (2015). Effectiveness of collaborative learning with complex tasks under different learning group formations: A cognitive load perspective. In *International Conference on Hybrid Learning and Continuing Education* (pp. 149-159). Springer, Cham.

# Improving Hybrid Brainstorming Outcomes with Scripting and Group Awareness Support

Mohammadreza Farrokhnia, University of Twente, the Netherlands, M.farrokhnia@utwente.nl

Omid Noroozi, Harm Biemans, Yvette Baggen  
Omid.noroozi@wur.nl, Harm.biemans@wur.nl, Yvette.baggen@wur.nl  
Wageningen University & Research, the Netherlands

**Abstract:** Previous research has shown that hybrid brainstorming, which combines individual and group methods, generates more ideas than either approach alone. However, the quality of these ideas remains similar across different methods. This study, guided by the dual-pathway to creativity model, tested two computer-supported scaffolds – scripting and group awareness support – for enhancing idea quality in hybrid brainstorming. 94 higher education students, grouped into triads, were tasked with generating ideas in three conditions. The Control condition used standard hybrid brainstorming without extra support. In Experimental 1 condition, students received scripting support during individual brainstorming, and students in the Experimental 2 condition were provided with group awareness support during the group phase in addition. While the quantity of ideas was similar across all conditions, the Experimental 2 condition produced ideas of higher quality, and the Experimental 1 condition also showed improved idea quality in the individual phase compared to the Control condition.

## Introduction

Brainstorming is a widely used technique for stimulating group creativity in various fields (Al-Samarraie & Hurmuzan, 2018), including entrepreneurship (Farrokhnia et al., 2022). This technique typically consists of two stages aligned with the divergent-convergent continuum (Al-Samarraie & Hurmuzan, 2018). The primary goal of the divergent (or idea generation) stage is to generate as many ideas as possible without immediate judgment (Ritter & Mostert, 2018). This is based on the premise that generating a larger quantity (i.e., number) of ideas increases the likelihood of producing higher-quality ones (Danes et al., 2020). Research has focused on developing methods to enhance idea quantity during the divergence stage of brainstorming sessions (Maaravi et al., 2021). Among these methods, brainstorming in hybrid settings, commonly known as "Hybrid brainstorming" (Korde & Paulus, 2017), is regarded as one of the most effective techniques for organizing brainstorming sessions (Paulus et al., 2018). Such a setting would allow both unconstrained ideation in individual brainstorming and the stimulation of additional ideas by exposure to the ideas of others (Korde & Paulus, 2017). In this method, participants are required to alternate between individual and group idea generation throughout the brainstorming session (Brown & Paulus, 2002). This flexibility allows for various sequencing, including individual-to-group (IG) (Ritter & Mostert, 2018), group-to-individual (Baruah & Paulus, 2008), and a combination of both (Korde & Paulus, 2017). Overall, empirical findings suggest that hybrid brainstorming, regardless of the sequence of individual and group phases, yields superior outcomes compared to individual or group brainstorming, especially in terms of the quantity of ideas generated (Korde & Paulus, 2017; Paulus et al., 2015).

Nevertheless, research findings indicate no significant differences in the quality of ideas generated in hybrid brainstorming sessions compared to those generated in individual or group brainstorming, measured by the degree of originality (Korde & Paulus, 2017; Paulus et al., 2015). Such findings resonate with studies indicating that increased quantity does not necessarily translate to higher quality (e.g., Baruah & Paulus, 2008; Rietzschel et al., 2014). Improving quality requires additional support that can effectively assist individuals in breaking free from conventional thinking (Rietzschel, 2018). The influence of such support on brainstorming outcomes is well elucidated by the dual-pathway to creativity model (DPCM) proposed by Nijstad et al. (2010). The DPCM suggests that there are two distinct but non-exclusive cognitive pathways for generating high-quality ideas: *persistence* and *flexibility*. Activating these pathways requires individuals to be mentally stimulated and engaged, for which the provision of certain process constraints such as rules, task structure, and instructions plays a crucial role (Nijstad et al., 2021).

Guided by the DPCM, this study aims to identify effective support mechanisms to enhance the quality of outcomes in hybrid brainstorming sessions. It draws on research in the field of computer-supported collaborative learning (CSCL) and examines the impact of two common computer-supported scaffolds: *scripting* and *group awareness support*. The study investigates whether these supports can guide individuals through the proposed cognitive pathways in hybrid brainstorming, thus influencing the quality of ideas generated.

## Theoretical framework

### The DPCM: The persistence pathway

In research on cognition, brainstorming is considered a repeated search for ideas in associative memory (Stroebe et al., 2010). However, such a search does not always yield promising outcomes as individuals tend to generate ideas that come to mind easily and give up very fast when it becomes harder to generate ideas (Rietzschel et al., 2014). Falling into this so-called “path of least resistance” often results in conventional (rather than original) ideas, as these ideas originate from knowledge that is highly accessible in one’s memory (Stroebe et al., 2010). To stimulate more original outcomes, individuals should be encouraged to leave the least resistance path (Rietzschel et al., 2014). Based on the DPCM, one approach to achieve this is to increase individuals’ persistence and perseverance when searching in their associative memory. This encourages them to follow what is referred to as the “persistence pathway”, which manifests itself as a prolonged cognitive effort to generate a higher number of ideas within a few idea categories (Nijstad et al., 2002). The effectiveness of this approach is because each category contains only a limited number of conventional ideas (De Dreu et al., 2008). Therefore, deep and persistent exploration within each category could elevate the likelihood of generating more original ideas over time (Rietzschel et al., 2007).

A key strategy for improving the quality of a brainstorming task involves providing support mechanisms that encourage a deeper, more focused, and systematic exploration within a few idea categories, thereby activating the persistence pathway (Nijstad et al., 2021). Persistence pathway requires systematic thinking, which involves blocking out distracting thoughts from one’s working memory and maintaining full attention on the task at hand (Nijstad, Dreu, et al., 2010). Consequently, if the support mechanism aims to enhance the quality of outcomes of a hybrid brainstorming task, it may be more effective when utilized in the individual phase. This is because, in the group phase, there is a higher likelihood of disruptions to the train of thought due to exposure to others’ ideas which can diminish the effectiveness of any supporting mechanism provided to activate the persistence pathway (Korde & Paulus, 2017).

Inspired by CSCL research (see Kobbe et al., 2007), *scripting* appears to be an effective support mechanism for activating the persistence pathway. Such support can be applied in various forms, such as sentence starters or detailed task descriptions, and also through carefully chosen question prompts (see Latifi et al., 2023; Noroozi et al., 2012, 2013) that can engage individuals in a more in-depth and systematic exploration. These question prompts could be those suggested by the SCAMPER technique. This technique guides individuals through a systematic search for solutions using a set of questions, encouraging them to (1) Substitute parts, (2) Combine elements, (3) Adapt to improve, (4) Modify features, (5) Put to other uses, (6) Eliminate unnecessary aspects, and (7) Rearrange or reverse components of an existing idea (Eberle, 1972). According to Rahimi and Shute (2021), the SCAMPER technique helps individuals actively generate new ideas, rather than simply waiting for these to form. The first objective of this study is to address the following question:

**RQ1:** *What is the impact of scripting in the form of providing SCAMPER question prompts during individual brainstorming on the originality of the ideas generated, compared to brainstorming without any additional support?*

### The DPCM: The flexibility pathway

From another theoretical perspective, original outcomes can also be achieved through “cognitive flexibility”. This flexibility is characterized by a holistic (and not in-depth) processing of information (Förster, 2009) and an exploration of broad cognitive categories (Nijstad & Stroebe, 2006), which may lead to the formation of novel and remote associations in memory (De Dreu et al., 2008). Building on this perspective, Nijstad et al. (2010) proposed another pathway in their DPCM that can result in quality ideas, known as “the flexibility pathway”. The activation of such a pathway in idea-generation tasks, such as brainstorming, is manifested by the generation of ideas across a large number of idea categories (Nijstad et al., 2021). In their meta-analysis, Nijstad et al. (2010) demonstrated that an increase in exploration breadth, i.e., the number of idea categories, is associated with generating ideas of higher average originality.

However, following the flexibility pathway does not typically occur spontaneously. Individuals often rely on their readily accessible knowledge, thereby limiting themselves to exploring only a few familiar idea categories (van Hooijdonk et al., 2022). This behaviour can contribute to a significant obstacle to creative idea generation during brainstorming sessions, often referred to as “the fixation effect”. This may result in the generation of mostly identical ideas, which all fall within limited idea categories. As a result, scholars recommend implementing support mechanisms that can assist individuals in overcoming such mental fixation (van Hooijdonk et al., 2022), thereby stimulating them to think outside (rather than inside) the box through increasing flexibility

(Peterson & Pattie, 2022). Driven by findings from the CSCL field, one approach is the use of *group awareness support*. This involves using computer-supported scaffolds to inform group members about each other's knowledge, also known as “cognitive awareness” (Bodemer & Dehler, 2011). In brainstorming sessions, such cognitive awareness can be achieved by sharing ideas during group brainstorming (Clayphan et al., 2011). The ideas shared by group members reflect their unique understanding and knowledge of the problem. When these ideas are effectively disseminated among the group, they can aid in assimilating diverse knowledge, and fostering new connections within their knowledge structures (Nijstad & Stroebe, 2006), consequently, increasing cognitive flexibility.

To date, many studies have explored the role of computer-supported scaffolds, in facilitating idea-sharing in group brainstorming and their effect on brainstorming outcomes (for an overview, see Maaravi et al., 2021). However, most of these studies have primarily focused on promoting the sharing of ideas generated “*during*” group brainstorming (Maaravi et al., 2021). Considering the significant role that the number of shared ideas plays in enhancing brainstorming outcomes (Paulus et al., 2013), such a setup may not fully harness the potential of idea-sharing in group settings. To address this issue, a possible solution is to tap into the potential of hybrid settings and improve group brainstorming outcomes by facilitating the sharing of ideas that were individually generated “*before*” group work commenced. Thus, the second objective of this study is to address the following question:

**RQ2:** *What is the impact of enhancing group awareness by sharing individually generated ideas with group members during brainstorming on the originality of the generated ideas, compared to the condition where sharing individually generated ideas is not facilitated?*

## Method

### Participants

In the present research study, 94 students—comprising both undergraduate (BSc) and postgraduate (MSc) cohorts—were involved. These students were participating in a university course specifically designed to inform them about future career possibilities and improve their entrepreneurial capabilities. The demographic breakdown of the participants was as follows: 43.7% were female, 55.3% were male, 12.8% were pursuing their Bachelor's degree, and 81% were in their Master's program. The average age of the participants was 24.1 years, with a standard deviation of 2.7 years.

### Material

#### The Ideation hub

The Ideation Hub is an online brainstorming platform that was specifically designed for this study to guide participants through a structured brainstorming process and implement the proposed supports effectively. The platform could offer clear instructions on the tasks to be completed during the individual and group phases, enforce time limits for each phase, and provide the necessary rules before and during each phase. Additionally, it could provide tailored support for each condition. For instance, participants in Experiment 1 and Experiment 2 conditions had the option to view their previously generated ideas and were encouraged to employ the SCAMPER principles using built-in question prompts. Students in the Control condition did not have access to such an option. Additionally, during the group phase, the platform would allow participants to submit their collaboratively generated ideas. All group members could submit ideas, and the submitted ideas became available to all members.

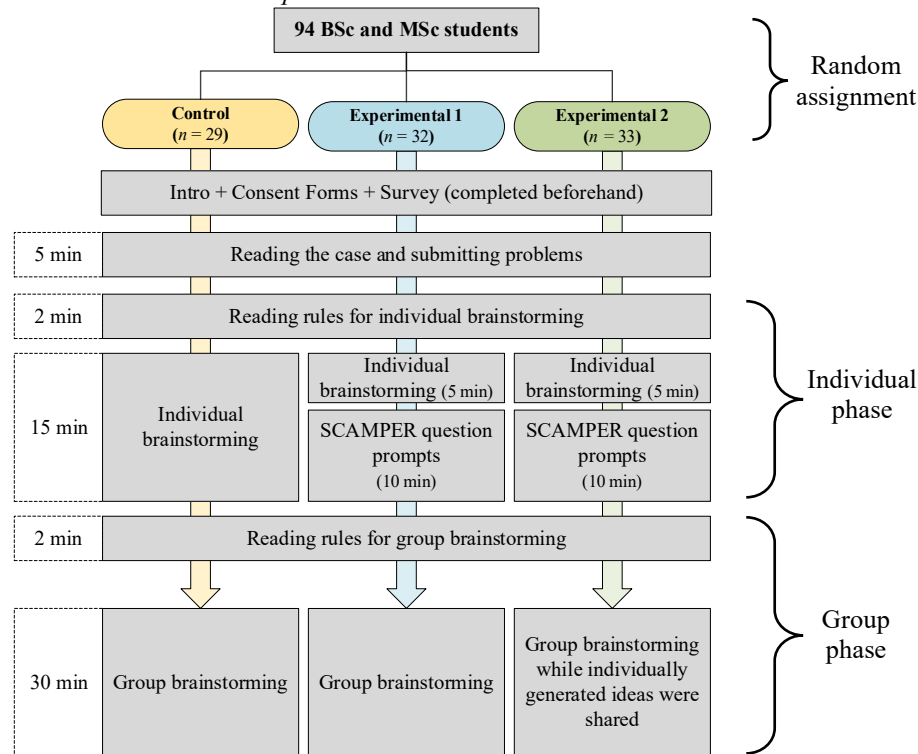
#### SCAMPER principles

The SCAMPER principles can be employed in various contexts. However, to make them more relevant to the task— as the goal was to generate business ideas to address sustainability issues – the first and third authors adapted the principles by refining the descriptions and providing real-world examples of successful businesses that have implemented these principles in practice. The inclusion of concrete examples was done to help participants better grasp the essence of each principle, foster greater understanding, and facilitate practical application.

### Procedure

The study design employed in this research was a between-subjects post-test design. Students were first randomly assigned to 31 groups. Subsequently, these groups were randomly assigned to one of three conditions: (1) Control Condition, (2) Experimental Condition 1, or (3) Experimental Condition 2. The overall procedure followed by all participants in the various conditions is outlined in Figure 1.

**Figure 1**  
*The Procedure Followed per Condition*



The main study was conducted in September 2022 as part of a university course, with prior permission from the lecturer. Ethical approval was received from the social sciences ethics committee of the associated university. Sustainable development was selected as the problem case to generate business ideas. It is defined as a form of development that “meets the needs of the present without compromising the ability of future generations to meet their own needs”. At the beginning of the workshop, an explanation was provided of what is meant by sustainable development and several specific examples regarding, for example, renewable energy, climate change, and sustainable education were given. The participants were then asked to imagine they were asked to provide input for business ideas for new start-ups in the area of sustainable development. These business ideas could concern people, the planet, and/or profit and may lead to social, environmental, and/or economic gains. What ideas for new start-ups come up in your mind?”. The participants in all conditions were given *five minutes* to read the case and contemplate various sustainability issues in their surroundings.

## Analysis

### Dependent measures

The dependent measures of students’ idea generation performance were the quantity and the quality of the generated business idea. In line with previous research on idea generation (Baruah & Paulus, 2008), the following criteria were employed to respectively evaluate the quantity and quality of ideas:

*Comprehensibility:* whether or not the idea can be qualified as an opportunity in terms of socially valued products or services in the context of sustainability (1 = comprehensible, 0 = incomprehensible). “Banning cars from cities to reduce air pollution is technically possible but does not constitute a product or service” (Eller et al., 2020). Ideas such as “wearing an extra sweater” or “turning down the heating” were scored as incomprehensible as they were more general recommendations to address sustainability-related issues than an idea for a start-up business. Incomprehensible ideas were excluded from further analysis. The number of comprehensible ideas is attributed to the quantity of generated ideas.

*Originality:* “the degree to which an idea is innovative” (Rietzschel et al., 2007, p. 934) which was determined using DeTienne and Chandler’s (2004) 6-point scale based upon the following categories: (1) no apparent innovation or not enough information to make a determination; (2) a product or service identical to an existing product/service offered to an underserved market; (3) a new application for an existing product/service, with

little/no modification or a minor change to an existing product; (4) a significant improvement to an existing product/service; (5) a combination of two or more existing products/services into one unique or new product/service; and (6) a new-to-the-world product/service, a pure invention or creation. The originality score of the individual and group phases was calculated by averaging the originality scores associated with ideas generated.

In addition, the cognitive flexibility in various phases was measured by the extent to which participants generated ideas in different categories. The categories were based on the examples of sustainable development in the problem case. Each idea was assigned to one category: (1) affordable and adequate food supply, (2) decent housing, (3) energy, (4) climate change, (5) education, and (6) personal health and safety. The flexibility score was calculated by counting the number of scored categories per participant. Moreover, cognitive persistence for each student was quantified by dividing the number of comprehensible ideas by flexibility.

### Coding of ideas

A codebook was created with two entrepreneurship scholars to evaluate the students' ideas. Initially, experts tested the codebook on 10% of the ideas, refining it after discussion until achieving high inter-rater reliability: Cohen's kappa of .91 for comprehensibility, weighted kappa of .73 for originality, and intraclass correlation coefficients of .89 and .73 for flexibility and originality, respectively. Then, two trained master's students assessed all ideas using the refined codebook.

### Unit of analysis and statistical tests

The unit of analysis, whether at the individual or group level, depended on the research question being addressed (see Table 1).

**Table 1**  
*Overview of Questions and Their Unit of Analysis*

Research questions	Unit of analysis	Data collection phase	Condition
RQ1	Individual	Individual phase	Control vs. Experimental 1 and 2
RQ2	Group	Group phase	Experimental 1 vs. Experimental 2

To answer the questions, multiple analyses of covariance (ANCOVA) were used for each facet of the brainstorming outcome (i.e., comprehensibility, originality, persistence, and flexibility) as the dependent variables, condition as the independent variable, and students' educational level and attitude toward entrepreneurship as covariates. Multiple ANCOVAs were conducted for each research question to examine whether there were significant overall differences between the various conditions in terms of the dependent variables. Subsequently, post hoc pairwise comparisons were performed to identify differences between the conditions for each dependent variable, employing Bonferroni corrections to account for multiple comparisons.

## Result

### RQ1

The difference among conditions was not significant for the number of comprehensible ideas ( $p = .32$ ) and their flexibility ( $p = .80$ ) but was significant for the average originality of ideas ( $p < .001$ , partial  $\eta^2 = .19$ ) and their persistence ( $p = .04$ , partial  $\eta^2 = .07$ ). As shown in Table 2, the results showed that the average originality of ideas generated in the individual phase of Experimental 1 (*Adjusted M* = 2.42, *SE* = .09) and Experimental 2 (*Adjusted M* = 2.54, *SE* = .09) conditions are significantly higher ( $p \leq .001$ ) than those generated in the control condition (*Adjusted M* = 1.98, *SE* = .09). Results from other comparisons also showed a significant difference at the 10 percent level ( $p = .08$ ) regarding the persistence of ideas generated in Experimental 1 (*Adjusted M* = 2.19, *SE* = .17) and 2 (*Adjusted M* = 2.24, *SE* = .17) conditions compared to those generated in the Control condition (*Adjusted M* = 1.67, *SE* = .17).

**Table 2**  
*Pairwise comparisons*

Dependent V.	Condition (I)	Condition (J)	MD (I-J)	SE	p*	95% CI for Difference <sup>b</sup>	
						LB	UB
<i>Originality</i>	Control	Experimental 1	-.44*	.12	.001	-.74	-.15
		Experimental 2	-.55*	.13	<.001	-.88	-.23
	Experimental 1	Control	.44*	.12	.001	.15	.74
		Experimental 2	-.11	.13	1.000	-.44	.21
<i>Persistence</i>	Control	Experimental 1	-.52	.23	.08	-1.08	.05
		Experimental 2	-.58	.26	.08	-1.20	.05



Experimental 1	Control	.52	.23	.08	-.05	1.08
	Experimental 2	-.06	.25	1.000	-.67	.56

Based on estimated marginal means

\*The mean difference is significant at the .05 level.

<sup>b</sup> Adjustment for multiple comparisons: Bonferroni. MD = Mean Difference.

## RQ2

The difference among conditions was not significant for the number of comprehensible ideas ( $p = .26$ ), their flexibility ( $p = .77$ ), and persistence ( $p = .13$ ) but was significant for the average originality of ideas ( $p = .02$ , partial  $\eta^2 = .25$ ). The average originality of ideas generated in the group phase of Experimental 2 condition (*Adjusted M* = 2.91, *SE* = .26) was shown to be significantly ( $p = .02$ ) higher than those generated in the group phase of Experimental 1 condition (*Adjusted M* = 2.62, *SE* = .23). Therefore, the sharing of individually generated ideas appeared to enhance the originality of ideas produced during group phase.

## Discussion and implications for practice

This study aimed to identify support mechanisms that enhance the quality of outcomes in hybrid brainstorming sessions. In doing so, it addresses an ongoing demand for empirical research on improving brainstorming quality (Maaravi et al., 2021) and a recent call to examine the DPCM empirically in authentic contexts, specifically when addressing real-life problems (Nijstad et al., 2021). To achieve these objectives, the study explored the effect of two computer-supported scaffolds: *scripting* and *group awareness support*, on the quality of ideas generated during hybrid brainstorming with the IG sequence addressing sustainability issues. The scripting support was delivered in the form of SCAMPER question prompts via the online platform. The findings revealed that participants supported by the SCAMPER question prompts during the individual brainstorming phase outperformed in terms of average originality and cognitive persistence, compared to those who did not use this technique. This suggests that incorporating the SCAMPER technique during individual brainstorming can positively influence the originality of outcomes when effectively scaffolded. This observation aligns with prior studies, indicating that utilizing the SCAMPER technique to improve a common product slightly increases originality compared to other methods like unsupported individual brainstorming (Gu et al., 2022). The improvement in idea originality becomes even more pronounced in studies that combined the SCAMPER techniques with brainstorming (Moreno et al., 2016; Rahimi & Shute, 2021), mirroring the approach in the present research.

The group awareness was enhanced during the group phase by sharing ideas generated individually. The findings suggested that this support could improve the average originality of the ideas generated in this phase. However, it did not successfully increase cognitive flexibility. This may imply that sharing individually generated ideas might not exert its effect on the quality of ideas through the flexibility pathway, but rather through another mechanism. During the group phase, sets of three students in each condition brainstormed for 30 minutes, aiming to collaboratively generate business ideas addressing sustainability issues. Participants in the Experimental 2 condition, however, had to additionally spend time reading, discussing, and understanding the individually shared generated ideas, which consequently left them with less time for collaborative ideation. While this setting offered them a chance to broaden their knowledge of the given problem and its possible solutions via shared ideas (Nijstad & Stroebe, 2006), the time constraints might have limited their ability to draw inspiration from them and generate a significantly diverse range of ideas. However, as was emphasized by the provided brainstorming rules, they could potentially collaborate and combine some of the shared ideas to generate new ones. This approach could lead to the generation of more original ideas than in the Experimental 1 condition, where participants did not have access to their peers' previously generated ideas and had to rely only on the information exchanged during the group phase.

The aforementioned findings have significant implications for practitioners aiming to organize effective hybrid brainstorming sessions. As with all types of brainstorming, additional support mechanisms must be still provided during various phases of a hybrid brainstorming session to guide individuals toward quality outcomes. A prime example of such mechanisms is the SCAMPER technique. This study showed that effectively prompting participants to utilize this technique during the individual phase of a hybrid brainstorming session significantly enhances the originality of their ideas. One critical consideration when using such prompts is to introduce them after participants have had a brief individual brainstorming period, allowing them to generate a pool of ideas to which the technique can be applied (Moreno et al., 2016; Rahimi & Shute, 2021). Moreover, it is advisable to adapt the technique to the context in use, providing concrete examples of how to apply each SCAMPER principle. This approach enables participants to better understand how to utilize the SCAMPER technique in the given context (Gu et al., 2022). Finally, it is essential to allocate sufficient time for the effective deployment of the

technique. The duration might vary based on the context and the nature of the problem at hand, be it an open-ended query or a complex real-life issue. Another effective support mechanism is promoting group awareness by sharing of previously generated ideas during group brainstorming. Consistent with prior studies (e.g., Paulus et al., 2013), the current research indicated that prompting participants to read, discuss, and comprehend their peers' ideas can significantly enhance the originality of ideas generated collaboratively.

## Limitations

This study has several limitations that warrant consideration. Firstly, the small sample size may limit the generalizability of the findings and potentially reduce the statistical power of the analyses, thereby affecting the ability to detect significant effects. Future research should aim to replicate these findings with larger, more diverse samples to enhance the external validity of the results. Secondly, the analysis did not employ statistical procedures designed to address the multilevel structure of the data, which could lead to an oversimplification of the relationships explored and possibly inflate type I error rates. Thirdly, the study did not account for the impact of individual emotional states on group brainstorming performance. Recognizing and understanding these emotional dynamics is crucial, as they significantly influence collaboration quality and outcomes. Future research could leverage emerging techniques, such as facial recognition technology, to monitor real-time emotional responses (see Sahraie et al., 2024), thereby enhancing our understanding and effectiveness of group interactions.

## References

- Al-Samarraie, H., & Hurmuzan, S. (2018). A review of brainstorming techniques in higher education. *Thinking Skills and Creativity*, 27, 78–91.
- Baruah, J., & Paulus, P. (2008). Effects of training on idea generation in groups. *Small Group Research*, 39(5), 523–541.
- Brown, V. R., & Paulus, P. B. (2002). Making group brainstorming more effective: Recommendations from an associative memory perspective. *Current Directions in Psychological Science*, 11(6), 208–212.
- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, 27(3), 1043–1045.
- Clayphan, A., Collins, A., Ackad, C., Kummerfeld, B., & Kay, J. (2011). Firestorm: a brainstorming application for collaborative group work at tabletops. In *Proceedings of the ACM international conference on interactive tabletops and surfaces* (pp. 162–171).
- Danes, J. E., Lindsey-Mullikin, J., & Lertwachara, K. (2020). The sequential order and quality of ideas in electronic brainstorming. *International Journal of Information Management*, 53, 102126.
- De Dreu, C. K. W., Baas, M., & Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: Toward a dual pathway to creativity model. *Journal of Personality and Social Psychology*, 94(5), 739–756.
- DeTienne, D. R., & Chandler, G. N. (2004). Opportunity identification and its Role in the entrepreneurial classroom: A pedagogical approach and empirical Test. *Academy of Management Learning & Education*, 3(3), 242–257.
- Eller, F. J., Gielnik, M. M., Wimmer, H., Thölke, C., Holzapfel, S., Tegtmeier, S., & Halberstadt, J. (2020). Identifying business opportunities for sustainable development: Longitudinal and experimental evidence contributing to the field of sustainable entrepreneurship. *Business Strategy and the Environment*, 29(3), 1387–1403.
- Farrokhnia, M., Baggen, Y., Biemans, H., & Noroozi, O. (2022). Bridging the fields of entrepreneurship and education: The role of philosophical perspectives in fostering opportunity identification. *The International Journal of Management Education*, 20(2), 100632.
- Farrokhnia, M., Noroozi, O., Baggen, Y., & Biemans, H. (2022). Sparking Creativity in Entrepreneurship Courses: Investigating the Effect of Hybrid Brainstorming Sessions on Business Opportunity Identification Outcomes. In *International Conference on Studies in Education and Social Sciences*, ICSES 2022.
- Förster, J. (2009). Relations between perceptual and conceptual scope: How global versus local processing fits a focus on similarity versus dissimilarity. *Journal of Experimental Psychology: General*, 138(1), 88–111.
- Gu, X., Ritter, S. M., Delfmann, L. R., & Dijksterhuis, A. (2022). Stimulating Creativity: Examining the Effectiveness of Four Cognitive-based Creativity Training Techniques. *The Journal of Creative Behavior*, 56(3), 312–327.
- Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, 2, 211–224.

- Latifi, S., Noroozi, O., & Talaei, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, 31(2), 655-669.
- Korde, R., & Paulus, P. (2017). Alternating individual and group idea generation: Finding the elusive synergy. *Journal of Experimental Social Psychology*, 70, 177-190.
- Maaravi, Y., Heller, B., Shoham, Y., Mohar, S., & Deutsch, B. (2021). Ideation in the digital age: Literature review and integrative model for electronic brainstorming. *Review of Managerial Science*, 15(6), 1431-1464.
- Moreno, D. P., Blessing, L. T., Yang, M. C., Hernández, A. A., & Wood, K. L. (2016). Overcoming design fixation: Design by analogy studies and nonintuitive findings. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 30(2), 185-199.
- Nijstad, B. A., Dreu, C. K. W. De, Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, 21(1), 34-77.
- Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, 10(3), 186-213.
- Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. M. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of Experimental Social Psychology*, 38(6), 535-544.
- Nijstad, B., Rietzschel, E. F., Baas, M., & De Dreu, C. K. W. (2021). The dual pathway to creativity model: Implications for workplace creativity. In *Handbook of Research on Creativity and Innovation*. Edward Elgar Publishing.
- Noroozi, O., Biemans, H.J.A., Weinberger, A., Mulder, M., & Chizari, M. (2013). Scripting for construction of a transactive memory system in multidisciplinary CSCL environments. *Learning and Instruction*, 25(1), 1-12.
- Noroozi, O., Weinberger, A., Biemans, H.J.A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCCL). A systematic review and synthesis of fifteen years of research. *Educational Research Review*, 7(2), 79-106.
- Paulus, P. B., Kohn, N. W., Arditti, L. E., & Korde, R. M. (2013). Understanding the group size effect in electronic brainstorming. *Small Group Research*, 44(3), 332-352.
- Paulus, P., Baruah, J., & Kenworthy, J. B. (2018). Enhancing collaborative ideation in organizations. *Frontiers in Psychology*, 0(OCT), 2024.
- Paulus, P., Korde, R., Dickson, J. J., Carmeli, A., & Cohen-Meitar, R. (2015). Asynchronous brainstorming in an industrial setting. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(6), 1076-1094.
- Peterson, D. R., & Pattie, M. W. (2022). Think outside and inside the box: The role of dual-pathway divergent thinking in creative idea generation. *Creativity Research Journal*, 1-19.
- Rahimi, S., & Shute, V. J. (2021). First inspire, then instruct to improve students' creativity. *Computers & Education*, 174, 104312.
- Rietzschel, E. F. (2018). Freedom, Structure, and Creativity. In R. Reiter-Palmon, V. L. Kennel, & J. C. Kaufman (Eds.), *Individual Creativity in the Workplace* (pp. 203-222). Elsevier Academic Press.
- Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, 43(6), 933-946.
- Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2014). Effects of problem scope and creativity instructions on idea generation and selection. *Creativity Research Journal*, 26(2), 185-191.
- Ritter, S. M., & Mostert, N. M. (2018). How to facilitate a brainstorming session: The effect of idea generation techniques and of group brainstorm after individual brainstorm. *Creative Industries Journal*, 11(3), 263-277.
- Sahraie, F., Rezvanfar, A., Movahedmohammadi, S. H., Ebner, M., & Farrokhnia, M. (2024). Analysis of learners' emotions in e-learning environments based on cognitive sciences. *International Journal of Interactive Mobile Technologies*, 18(7), 34-52.
- Stroebe, W., Nijstad, B. A., & Rietzschel, E. F. (2010). Beyond Productivity Loss in Brainstorming Groups. In *Advances in experimental social psychology* (Vol. 43, pp. 157-203). Academic Press.
- van Hooijdonk, M., Ritter, S. M., Linka, M., & Kroesbergen, E. (2022). Creativity and change of context: The influence of object-context (in)congruency on cognitive flexibility. *Thinking Skills and Creativity*, 45, 101044.

## Towards a Framework to Link Them All. A Provocation in the Ongoing Debate on Collaborative Learning

Sebastian Simon, LIUM, Le Mans Université, 72085 Le Mans, France, [sebastian.simon@univ-lemans.fr](mailto:sebastian.simon@univ-lemans.fr)  
Iza Marfisi-Schottman, LIUM, Le Mans Université, 72085 Le Mans, France, [iza.marfisi@univ-lemans.fr](mailto:iza.marfisi@univ-lemans.fr)  
Sébastien George, LIUM, Le Mans Université, 72085 Le Mans, France, [sebastien.george@univ-lemans.fr](mailto:sebastien.george@univ-lemans.fr)

**Abstract:** Common ground is a well-studied concept in the CSCL (Computer-Supported Collaborative Learning) community. Interestingly, while well-studied, its discovery did not seem to lead to common conceptual ground within the CSCL community. Since its beginnings 30 years ago, the community has produced a variety of frameworks, studies and theories around the concept of collaboration, collaborative learning and, in particular, collaborative learning tools. Missing common ground is a problem for analyzing collaboration itself and comparability across studies, which researchers such as Rummel, Wise and Schwarz highlight in a recent exchange on the subject. In this paper, we analyze existing frameworks and whether contradictions between different frameworks exist. We further propose an attempt to a joint definition and framework as a starting point and provocation for discussion in the community.

### Introduction

Since the first definition of collaboration in scholar settings by Roschelle et al., almost 30 years have gone by. While their definition still holds today, the variety of collaborative activities, tools and environments has given rise to a variety of definitions, frameworks and theories in different fields (Stahl, 2021).

The variety of theories and frameworks can be attributed to the multidisciplinary of Computer-Supported Collaborative Learning (CSCL). Indeed, CSCL is a domain at the crossroads of psychology, learning sciences, computer science, sociology, linguistics, anthropology and communications (Hmelo-Silver & Jeong, 2021). Another reason for the variety of models might also be the fact that, while studying collaborative learning (CL), the research community itself is a community of collaborative learners, building on the works from others to create new insights. New theories emerge based on existing work and as a result from conducted studies. For work to be taken into account however, it has to be written, published and made available to the community. Importantly, it then has to be read and taken into account by other researchers. More than a dozen domains involved in CSCL make it difficult to gain and maintain a holistic view.

The many coexisting frameworks perfectly fit different contexts and usages, but add to the complexity of a global vision. This is also a problem for comparability of results: The findings of one work cannot easily be compared to others if the underlying perception of what is experimented is not the same or known at all. Yet, comparability is key if CSCL is to advance (Griffiths et al. 2021): In a context of oftentimes small sample sizes, collaboration not being automatic and research frequently carried out on humans in environments with a large number of uncontrolled conditions (Bachour et al., 2010), it is in the interest of the community to share common ground on the conceptual stance of collaboration.

Efforts to reunite CSCL knowledge are underway (Cress, Rosé, Wise, & Oshima, 2021): the community's latest edition of a comprehensive handbook has been published in 2021. The urgency of conceptual consolidation has been recognized by leading researchers (Wise & Schwarz, 2017) questioning if "one framework to rule them all" is an adequate response to the problem. We argue instead for a duality of an overarching framework in conjunction with existing contextual frameworks. Knowledge representations, such as ontologies allow the flexible mapping of concepts (and also linking ontologies between domains) to provide a common understanding. We propose such an ontological, bridging approach as a starting point for a common conceptual vision, without sacrificing the many theories on collaboration but providing mapping abilities between the different perspectives.

To this end, the first part of this paper examines existing attempts on common CSCL frameworks. The second part details the method deployed to establish a common framework. We then present our findings and discuss its potential and future use.

### Previous work

Rummel's take on the "provocations" by Wise et al. (2017) is that it is very difficult, as a community, to reach common conceptual ground. There is also doubt on the extent to which common ground is required. Rummel (2018) for instance proposes a *taxonomy* to design collaborative activities and analyze collaboration in order to

contribute to the discussion of Wise et al. Objectively, the most urgent need in terms of consolidation is indeed an **analytics** and **design** framework due to the fast pace with which technology steadily increases possibilities of collaboration support and automated analysis. Meanwhile, it is clear that there is no *one* way of analyzing collaboration in the field, which makes comparison of studies and reproducibility difficult (Martinez-Maldonado et al., 2021). The same is true for design frameworks. The reason, in both cases, seems to be, that in order to analyze or design environments for collaboration or CL, the research community needs to agree on what those concepts mean in detail. Different ways of analyzing and designing collaborative environments are rooted in different visions of what needs to be observed and what needs to be supported by digital tools. Rummel’s taxonomy is no exception: It features for example “goals” and “delivery agents” which indicate a conceptual vision, for instance, for the role of an educator, and the goals of collaborative tools, for collaboration.

Frameworks on general aspects of collaboration (Kirschner & Erkens, 2013) exist alongside frameworks with a focus on specific aspects of collaboration, providing a lens through which collaboration as a whole is explained, such as the framework of Hesse et al. for assessing collaborative skills. Researchers from associated fields provide frameworks with a focus on cognitive, gamified, pedagogical or conditional aspects (Griffiths et al., 2021), through a variety of forms (ontologies, lists, graphs etc.).

## Method

In order to build a consolidating core framework within this heterogeneous landscape, two issues of conceptual uncertainty in CSCL have to be addressed: for one, collaboration itself has no clear definition. Similarly, there is a multitude of learning theories and the variety of frameworks in CSCL is, as noted Rummel, coherent with the variety of perspectives on both learning and collaboration. As will demonstrate this paper, existing frameworks and theories are not mutually exclusive but merely put forward different facets of a coherent set of phenomena.

**Table 1**  
*CSCL Framework Creation Process*

Nº	Stage	Method	Output
1	A joint definition of collaboration	Systematic Review	Definition
2	A core framework of collaboration	Systematic Review	Collaboration Framework
3	A collaborative Learning extension	Conceptual Integration	CL Framework
4	A CSCL mapping	Ontological Linking	CSCL framework mapping

Our framework creation process is outlined in table 1: Initially, we attempted a concise, compatible definition of collaboration (1). We compared different concept classes from systematic reviews of definitions and other sources (see table 2). Concept classes address fundamental entities. For example, collaboration may be seen as a state or a process. Both concepts are mutually exclusive but can be categorized as entities of a system. Categorizing definitions’ concepts this way revealed incompatible perspectives. After discussing contradicting properties, similar concepts were grouped into hierarchies or relationships, in which case the most inclusive concept was used in the definition (and related or lower level concepts were kept for the framework). Similarly, we focused the definition on mandatory collaborative properties, without which there is no collaboration. In contrast, properties enhancing collaboration are filtered and integrated in the framework.

Next, we created a common, abstract conceptual core of collaboration, extending on the definition (2), using contributions from CSCL, CSCW and research in other domains. Thereby, the core concepts remain compatible and the framework open to contributions from other domains. For each of the keyword queries in google scholar, Apa Psychinfo and ERIC “Conceptual collaborative learning framework”, “Conceptual collaboration framework for small groups”, “Conceptual framework for collaborative Problem solving”, “CSCL framework” and “CSCL ontology”, we examined the first 50 results. Complementary sources were the Journal of CSCL as well as the 2013 and 2021 edition of the handbook of international CSCL.

Jabareen et al. (2009) highlight the ambiguity on what a framework actually is among researchers. Thus, to filter the retained results and consequently attempt a joint version, we used his conceptual framework definition of “a network of linked concepts that together provide a comprehensive understanding of the phenomenon”. 42 papers matched this definition (other results contained all keywords, but would focus on frameworks on concepts or not provide insights about how the different concepts relate to each other). Frameworks were consequently categorized according to their perspective: skills, processes, general aspects, conditions, pedagogical approaches. We then compared *general* frameworks and integrated them into our conceptual core of collaboration (figure 1) before testing the core’s compatibility with different learning theories (figure 2). Then, we confronted our

conceptual stance with the remaining, more *specific* frameworks and other CSCL works (Figure 2Q4.4, Q4.5) to illustrate its compatibility and provide an artifact for a communitywide discussion.

## Results

### A definition of collaboration (1)

Establishing meaning through *evoking* and *linking* a concept to others in a short and concise manner provides the basis for a common understanding in a discussion. In table 2 we sum up presence and form of concepts from different definitions. The sources have been chosen to cover a variety of different perspectives, all concerned with providing a comprehensive vision: a dictionary definition, the perspective of transnational organizations such as the OECD, literature reviews and foundational works in collaborative learning. While in many cases, definitions overlap, this is not the case for the system entity type. While a *process* can be considered a series of related activities to reach a new state of a system, a state is a static snapshot of a system. Since research on collaboration has established the multifaceted nature of collaboration, an activity alone seems inadequate. Similarly, Vogel et al. (2017) point out the importance of transactive *actions* which contradicts the vision of a collaborative state. Roschelle et al. consider synchrony an attribute in their context of face-to-face collaboration. If synchrony was a distinctive, general feature of collaboration, asynchronous collaborations (e.g. meaning making in forums) would be excluded, even if the individual contributions still have the same goal and thus don't fit concepts such as cooperation or co-action (George & Leroux, 2001). Synchrony can thus not be considered an elementary feature of general collaboration. The elements of *joint effort* and *goals* are recurrent across definitions. The PISA 2015 definition does not explicit the common goal, but the "problem" actors try to solve. Definitions vary in terms of involved entities.

**Table 2**

*Definitions of Systemic Reviews and Frameworks (Red: Opposing Concepts, Yellow: Enhancing Properties, Green: Concept Present)*

Authors	1 System entity (s:state, p:process, a:activity)	2 Synchrony	3 Effort wt: working together je: joint effort	4 Objectives cg: common goal	5 Agency 2p: 2+ people 2a: 2+ agents	6 Common Ground	7 Pooling knowledge/skills	8 Outcome s: solution o: outcome	9 Open Communication	10 Trust	11 Mutual respect	12 Shared responsibility	13 Active Participation	14 Joint Decision Making	15 Problem Complexity	16 Horizontal Hierarchy	17 Complementary Expertise	18 Conflict Negotiation and Resolution	
Cambridge Dictionary 2023	s		wt	cg	2p														
PISA 2015	p		je		2a			s											
Roschelle et al.	p		je	cg															
Hesse et al 2015	a		je	cg				o											
Griffiths et al. 2021	p		wt	sg				s											
Alozie 2023	p		je	cg	2p			s											

While definitions by Roschelle, Hesse and Griffiths do not explicitly include the need for more than one entity, their works repeatedly evoke multiple participants of collaborative activities. To this extent and in the light of progress in artificial conversation agents (e.g. ChatGPT), it seems wise to keep an abstract notion of two or more entities. Less common attributes of collaboration (among definitions) are open communication, horizontal hierarchy and complementary expertise. Those attributes have proven important for improving collaboration but are not essential: Collaboration occurs in settings with strong hierarchies (companies, armies, etc. challenging 16). In CL settings, students often have a very similar expertise (challenging 17). Passive behaviors like social loafing impact collaboration negatively but may not lead to the collapse of collaboration among other members (challenging 13). Similarly, open communication may impact collaboration in the long run, but members may decide to not communicate situational concerns and still collaborate (challenging 9).

Joint decision and conflict management are necessary processes in collaboration since there is usually no external entity to guide or mediate the group's cognitive conflicts. The role of mediators (teachers etc.) is often cited as a counterexample. However, the essence of collaboration is to confront and negotiate a shared perception based on prior differing perceptions of group members. This is also the reason why trust, mutual respect and shared responsibility are implicit, essential properties of collaboration. If collaboration can be considered one or more processes, then the presence of an outcome is also implicit. Those considerations can be summed up as follows: **Collaboration** is a set of processes (1, 8, 15) in which 2 or more entities (5) engage in a joint effort (3) towards a common goal (4) managing conflicts and taking decisions by themselves (10, 11, 12, 14, 18).

## A collaborative core framework (2)

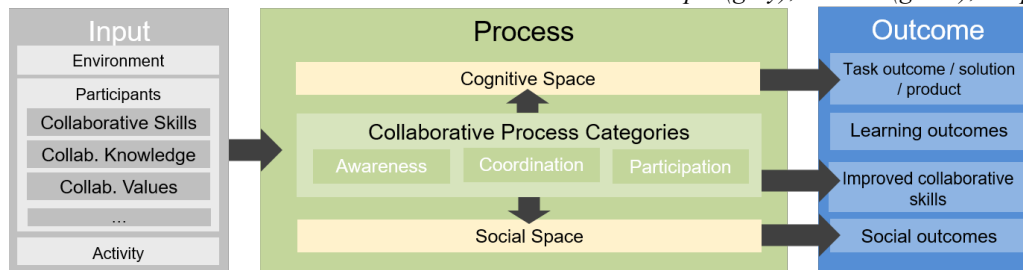
The previous definition of collaboration aligns with the **process** model used by different authors (Alozie et al., 2023; Mateescu et al., 2019; Simon et al., 2022). Alozie et al. (2023) group concepts into the categories “Input, Process” (sharing knowledge and resources) and “Output Process” (knowledge, artifacts). This process is then extended by norms (shared goals, reciprocal engagement, responsibility and accountability, mutual trust and respect, non-hierarchical shared power and voluntary participation) impacting the collaborative processes. Frameworks vary mostly in their focus. Hesse et al. analyze and assess individual collaborative skills. Kirschner et al. (2018) provide a theory of cognitive collaborative load for groups. Neumayr et al. (2018)’s framework focuses on collaborative coupling styles. The works of Laal (2013) and Johnson and Johnson (2004) focus on conditions favoring collaboration (social skills, promotive interactions, group processing, positive interdependence, individual/group accountability, etc.) whereas Vogel (2017) provides a scripting framework.

Those perspectives are not mutually exclusive: The problem, task or activity, the environment (tools, resources, constraints) and the individuals with their individual knowledge, collaborative and cognitive (task-related) skills, as well as their values, form the **input** of the overall collaborative process (see Figure 1).

The **output** of a collaboration can be categorized in task-related outcomes (products, artifacts, etc.), and learning outcomes: during collaboration, members are confronted with their peers’ task-related perspectives, solution strategies and values challenging their own. This turn on collaboration shows its interest for learning and work alike: Collaboration triggers learning processes but also allows to achieve an outcome related to a complex task. Consequently, CSCL will adapt the conditions to optimize learning outcomes whereas CSCW acts on more fixed circumstances. Participants engaging in a collaboration also have the opportunity to practice and improve their own collaborative skills, forming a third outcome of collaboration. Finally, relationships and individuals’ modified values beyond the collaboration represents a social outcome (Griffiths et al., 2021).

**Figure 1**

*Collaborative Core Framework Based on the Process Model: Input (grey), Process (green), Output (blue)*



Kirschner et al. (2018) developed a perspective of cognitive load on collaboration. Human cognitive resources (ability to concentrate, reflect or engage in creative activities and transactive interactions) are considered limited and thus, CSCL specifically strives to reduce cognitive load on particular aspects to channel cognitive resources on aspects tied to learning objectives. The theory also provides *raison d’être* for collaboration in complex problem settings: the group can pool individual knowledge and rely on a shared memory (transactive memory system) extending cognitive resources. In the context of a framework, it provides a general property of collaborative processes which is the cognitive load. There is an ongoing debate on the unit of analysis among CSCL researchers, but the collaborative load theory accounts for both group and individual considerations of collaborative processes.

Collaborative processes are dynamic in nature. In their fundamental CSCL contribution, Roschelle and Teasley (1995) defined collaboration as opposed to cooperation: collaboration differs to cooperation in that collaboration requires a continuous, common effort of a group or dyad to maintain common ground, whereas cooperation is the division of subtasks to a common goal on which group members work in parallel but without requiring mutual understanding. George & Leroux (2001) pointed out that collaboration and cooperation are not unrelated and generally occur in alternating patterns: Collaboration is cognitively intense and cannot be maintained over long periods of time (Kirschner et al., 2018). Inversely, cooperation requires a minimum of common ground. George & Leroux (2001) characterized collective activities as series of collaborative and cooperative phases. Indeed, when teams collaborate, collaboration may dynamically shift from collaboration between all members to collaboration in subgroups. Using the group as the unit of analysis, the group can be in a cooperative phase even if all members still collaborate (in subgroups). Engeström (2008) included cooperation as a key concept to describe collaboration in its framework.

In Simon et al. (2022), we analyzed common features from frameworks by Hesse, Mateescu and Meier to identify three main categories of collaborative processes: **Awareness, Coordination and Participation**. Participation is referred to as “an observable action of engagement in communication”, capturing both verbal and nonverbal exchanges. Awareness regroups internal processes that maintain awareness on social, cognitive and behavioral activities of peers. Coordination refers to task and social conflict resolution as well as decision making strategies.

Collaborative processes have been categorized differently by Kirschner et al. (2015), arguing that there are social processes, nourishing a **social space** (e.g. through encouraging peers) and task-related processes to establish common conceptual ground among group members. Their framework distinguishes cognitive from social performance. We thus integrate the notion of a social and **cognitive space** (Roschelle & Teasley, 1995) in the collaborative core to which members contribute through participative processes. Group members organize and structure both spaces through the use of coordinative processes to regulate contributions to the social and cognitive space. For those contributions to be constructive, members have to be aware of the social group dimension (Behavioral Awareness: “When can I contribute something?”, Social Awareness: “In which mental state are my peers?”) and the cognitive dimension (“Who knows what?”) (Ma et al., 2020). Processes of all three categories are thus implied in creating and maintaining both cognitive and social space. Verifying the coherence with Kirschner’s theory of collaborative load, we find that a cognitive load can be attached to all three process categories and processes for both cognitive and social contributions. Kreijns et al. (2013) define the social space as a place where trust, motivation, interpersonal relationships and the sense of community is developed and maintained.

The cognitive space refers to a common understanding of problems and solutions. The group actively engages in its construction (requiring active participation), management (upon arrival of new evidence) and repair.

Another main component of many frameworks (Hesse et al., 2015; Johnson & Johnson, 2004; Laal, 2013; Meier et al., 2007) includes conditions or process input. This aspect is based on the works of Johnson & Johnson (2004) who observed that for successful collaboration, social skills, promotive interaction and group processing abilities as well as positive interdependence and individual/group accountability have to be present among group members. Regarding the definition of collaboration, we can objectively regroup and extend the conditional aspect to participants, environment and the problem/task. Collaborative skills consequently are a property of the group’s members. This vision is still compatible with CSCL if computer support is considered a tool and tools a part of the environment in which the activity takes place.

Script theory is another perspective on collaboration and group interactions. It states that for a certain type of situation, humans have procedures and organize their internal knowledge about those situations in “internal scripts” (e.g. a script “restaurant visit” may include the process of waiting to be seated, being handed a menu, etc.). External scripts are explicit instructions for procedures to follow in a situation (e.g. flight preparation protocols). Internal collaborative scripts are part of the input group members bring to a collaborative setting, similar to collaborative skills. Collaborative scripts are a type of knowledge about collaboration (together with knowledge about collaborative processes). Finally, collaborative values refer to priorities on aspects of successful collaboration (such as equal access to resources). These three dimensions are negotiated and organized among a group through coordinative processes in the social space. Research on team compositions confirms the presence of those procedures and predispositions (Kreijns, Kirschner & Vermeulen, 2013).

The previously discussed, different aspects are combined in Figure 1. Conditions are the input of the global collaborative process that requires common social and cognitive spaces on which operate Awareness, Coordination and Participation processes. The potential outcome of this global process is the task outcome, the “lessons learned”, improved collaborative skills and social outcomes, such as relationships.

### A collaborative learning extension (3)

The previous, collaborative framework provides a common vision on collaboration. In order to establish common ground on CSCL, we examine this model’s compatibility and possible links with learning theories. Individuals have to acquire skills, knowledge and values to become functional parts of modern society. Wenger et al. (1991) describe learning as the legitimate peripheral participation in a community. Indeed, skills, knowledge and values are acquired by participating in the community’s activities. Teaching can be defined as the activity of actively steering and optimizing acquisition by those “new members” (Lave & Wenger, 1991). The appeal of collaboration as a vehicle for learning is thus immediate and fourfold, considering the output of the previous framework.

While still being actively researched, institutions have identified collaboration as the “super skill” of the 21st century and aim to teach students the necessary skillset (Praharaj, 2022). Coincidentally, engaging in collaboration also allows to improve collaborative skills, values and knowledge, provided collaborative skill/value/knowledge diversity among group members (Cress et al., 2021). CL can thus be contrasted with the



Learning of Collaboration. In CL, collaboration serves as a vehicle to convey skills, knowledge or values. Inversely, teaching/learning collaboration aims at conveying collaborative values, collaborative skills and collaborative knowledge (as defined previously). This distinction is crucial for the design and analysis of CSCL setups: if students are confronted with new study material, following collaborative load theory, it might be difficult for students to engage in meaningful learning without available collaborative affordances and inversely, collaborative skills might be better learned if the task content is already known to some extent.

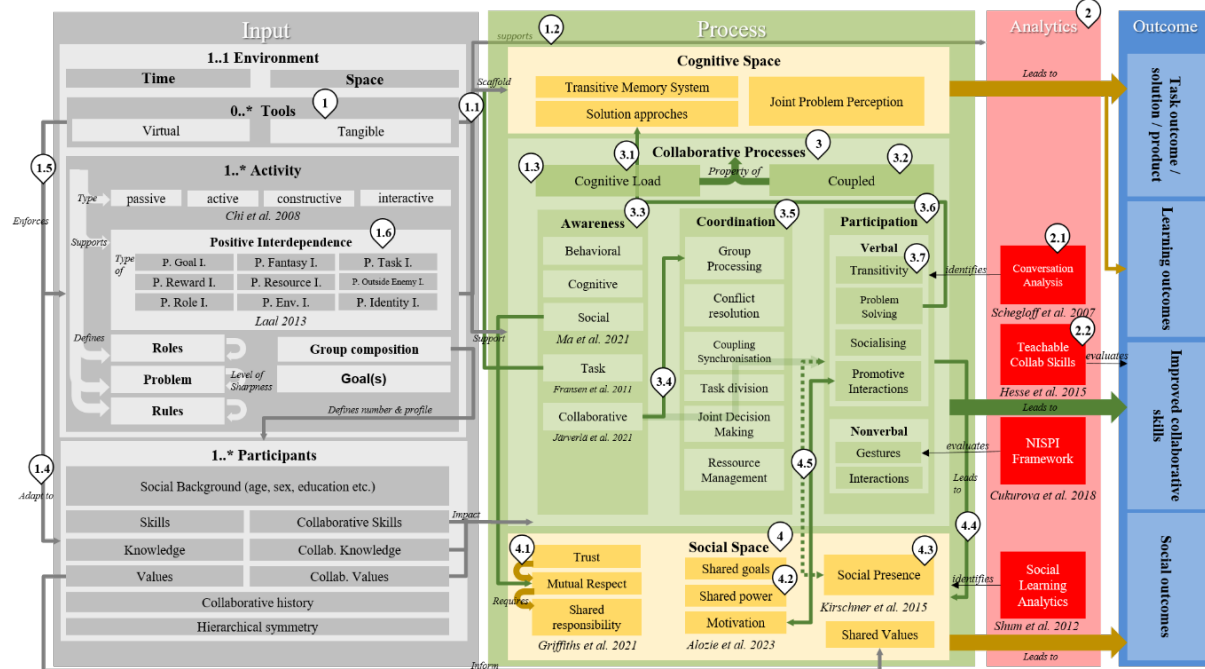
Challenging existing conceptions is central to the Piagetian model of learning, giving an active role to the learner in the process. Collaboration is also compatible with other learning theories, such as the activity theory of Engeström as demonstrated by Barros et al. (2002), having mapped the concepts of activity theory to a CSCL ontology. Short term and long term learning cycles also have an equivalent in collaboration: Every output type of collaboration can be its input in a cyclic pattern. Participation in such cycles can lead to virtuous or vicious CL cycles: Virtuous, in that it may improve collaborative performance and values of individuals over time and vicious if individuals experience repeated collaborative failure, leading to a negative attitude towards collaboration.

### A CSCL mapping (4)

Having established an abstract core framework of collaboration compatible with CSCL and CSCW, the following paragraphs will *illustrate* its mapping abilities for a subset of CSCL research (see Figure 2) and provide further details. CSCL intends to provide tools (◊ 1) to *enhance* (◊ 1.1) and *analyze* (◊ 1.2) CL, as well as the learning of collaboration. Given cognitive load theory, tools are designed to lower the task-related cognitive load and aim to increase the group cognitive load on specific, collaborative aspects (learning of collaboration) or the inverse (decreasing cognitive load on collaborative processes for CL) (◊ 1.3). Tools may adapt to participants (◊ 1.4) and contribute to maintain and enforce the activity’s rules and interdependence design (◊ 1.5). Links to Laal’s positive interdependence framework are provided at ◊ 1.6.

**Figure 2**

*Proposition for a Multilevel CSCL Framework. Arrow Types Depict Conceptual, Experimentally Confirmed or Hypothesized (dotted) Links. Numbers Show the Range of Quantity of Instances when Used in a Study Context.*



Collaborative analytics (◊ 2) is a recent field, due to the multimodal nature of data and evolving capabilities of analysis through the use of AI and new sensors (Martinez-Maldonado et al., 2021). We showcase their general compatibility with some of the framework’s concepts in Figure 2 to highlight the framework’s usability as a documentation tool for field studies. Analysis methods focus on outcomes (◊ 2.1) and processes alike (◊ 2.2). Hesse et al. have established an extensive list of collaborative skills and relevant indicators to assess those skills. Those indicators can be conveniently associated to the components of the presented framework. The

NISPI framework aims to detect collaboration based on non-verbal cues. Conversation analysis in turn can be considered a framework for verbal participative processes (Q 2.1).

Previously discussed features of collaborative processes (Q 3) are their cognitive load (Q 3.1) and dynamic nature (Q 3.2). Awareness processes (Q 3.3) are widely acknowledged to be fundamental for collaboration and CL. Beyond the already mentioned social, behavioral, cognitive and task awareness processes, awareness on collaborative processes, also known as metacognition, has been identified, allowing group members to reflect on their emotional and cognitive state and alter them accordingly (Cress et al., 2021) (Q 3.4).

Coordination (Q 3.5) can occur in the cognitive space in the form of strategies to solve a problem or accomplish a task or in the form of meta-strategies (group processing) monitoring and altering solution strategies depending on their performance. Coordinative processes are also required to channel and account for the social space. The notion of social presence is still debated (Weidlich, Kreijns, Rajagopal & Bastiaens, 2018) but broadly refers to the peer's perception of a member within a group through their communicative activities. Bachour et al. (2010) have shown the positive impact on mirroring tools showing the group participation of group members and highlighting the importance of a balanced participation for collaboration.

Participation (Q 3.6) can occur verbally and nonverbally. Verbally, researchers highlight the importance of transactive communication for successful collaboration. Transactive communication (Q 3.7) refers to group members mutually building on previous contributions. Communication can also focus on the social space to mediate conflicts or motivate each other (Johnson & Johnson, 2004; Vogel et al., 2017).

Griffiths et al. (2021) provide a hierarchical structure of collaborative processes for the social space (Q 4): the foundational relations are built on communication and trust (Q 4.1), which allows for the negotiation for shared values, which in turn encompasses shared goals and common understanding. Then, team members show active engagement by sharing responsibilities and active participation (Q 4.2). Finally, collaboration takes place when decisions are taken and negotiated collectively (Griffiths et al., 2021). While the serial nature of their perspective is problematic as the processes of active engagement have been observed to occur in parallel to the construction of shared values and relationship building (Kreijns et al., 2013), it provides structure to the social space and links between its properties. Kreijns et al. provide a framework for the integration of social space and social presence (Q 4.3), stating that through participation, members build their social visibility in a group (Q 4.4). Kreijns et al. hypothesize further that the social presence in turn impacts further participation of participants (Q 4.5). The aforementioned link demonstrates this framework's use to include ongoing research. The framework can be further used as a documentation of study setups (highlighting the type and quantity of each object).

## Conclusion

In this paper we presented an updated definition of collaboration and an attempt to structure part of the CSCL community's acquired conceptual knowledge in the form of a multilevel, modular framework. Its compatibility with two learning theories and other frameworks and CSCL findings was highlighted. Its design features align with a vision and a need for *collaborative research* in CSCL. As various authors have outlined before, the complexity of the phenomenon requires community wide collaboration. We hope for this work to trigger the implication of the different authors in a collaborative effort to discuss, criticize and consequently create a collaborative framework for collaboration (of which Figure 2 is but an initial, incomplete attempt).

Building common ground is no isolated, punctual effort and must be maintained throughout the collaborative process. In order to do so, the community must give itself appropriate *tools* and engage in group processing. We suggest the use and adaptation of knowledge graph building tools, frequently experimented in collaborative research (Scardamalia, 2002). The result can serve as a *map* for new and established CSCL members: to navigate the conceptual CSCL landscape, to quickly gain insights into research, but also for the community to build on and discuss concepts driven by their study results. Further perspectives include the integration of frameworks on collaborative computation, completing the list of e.g. the participative processes and their interaction with the cognitive space, detailing links (particularly for the frameworks on collaborative analytics).

In the tradition of collaborations of this kind, we hope to further consolidate a Joint Problem Space for CSCL (as well as a social space). Who if not the international CSCL community would be better suited to conduct and excel at such a collaborative endeavor?

## References

- Alozie, N., Yang, H., Rachmatullah, A., & Lopez-Prado, B. (2023). Toward A More Comprehensive Definition of Collaboration: Scholarly Literature vs. Practitioners. *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023*, 1246–1249. Montréal, Québec, Canada: ISLS.
- Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies*, 3, 203–213.

- Barros, B., Verdejo, M., Read, T., & Mizoguchi, R. (2002). Applications of a Collaborative Learning Ontology. *Proceedings of the 2nd MICAI*, 103–118. Yucatan, Mexico.
- Cress, U., Rosé, C., Wise, A. F., & Oshima, J. (Eds.). (2021). *International handbook of computer-supported collaborative learning*. Cham: Springer.
- George, S., Leroux, P., (2001). *Apprentissage collectif à distance. SPLACH: un environnement informatique support d'une pédagogie de projet*. Université du Maine, Le Mans.
- Griffiths, A.-J., Alsip, J., Hart, S. R., Round, R. L., & Brady, J. (2021). Together We Can Do So Much: A Systematic Review and Conceptual Framework of Collaboration in Schools. *Canadian Journal of School Psychology*, 36, 59–85.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A Framework for Teachable Collaborative Problem Solving Skills. In *Educational Assessment in an Information Age. Assessment and Teaching of 21st Century Skills: Methods and Approach* (pp. 37–56). Dordrecht: Springer Netherlands.
- Hmelo-Silver, C. E., & Jeong, H. (2021). Benefits and Challenges of Interdisciplinarity in CSCL Research: A View From the Literature. *Frontiers in Psychology*, 11, 579986.
- Johnson, D. W., & Johnson, R. T. (2004). Cooperation and the Use of Technology. In *Handbook of research on educational communications and technology, 2nd ed* (pp. 785–811). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Kirschner, P. A. (2013). Toward a Framework for CSCL Research. *Educational Psychologist*, 48, 1–8.
- Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning*, 13, 213–233.
- Kreijns, K., Kirschner, P. A., & Vermeulen, M. (2013). Social Aspects of CSCL Environments: A Research Framework. *Educational Psychologist*, 48, 229–242.
- Laal, M. (2013). Positive Interdependence in Collaborative Learning. *Procedia - Social and Behavioral Sciences*, 93, 1433–1437.
- Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Ma, X., Liu, J., Liang, J., & Fan, C. (2020). An empirical study on the effect of group awareness in CSCL environments. *Interactive Learning Environments*, 0, 1–16.
- Martinez-Maldonado, R., Gašević, D., Echeverria, V., Nieto, G. F., Swiecki. (2021). What Do You Mean by Collaboration Analytics? A Conceptual Model. *Journal of Learning Analytics*, 8, 126–153.
- Mateescu, M., Pimmer, C., Zahn, C., Klinkhammer, D., & Reiterer, H. (2019). Collaboration on large interactive displays: A systematic review. *Human-Computer Interaction*, 36, 1–35.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *Int. Journal of Computer-Supported Collaborative Learning*, 2, 63–86.
- Praharaj, S. (2022). *Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics* (Doctoral Thesis). Open Universiteit, Netherlands.
- Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 69–97). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Rummel, N (2018). One framework to rule them all? Carrying forward the conversation started by Wise and Schwarz. *International Journal of Computer-Supported Collaborative Learning* (pp. 123-129).
- Scardamalia, M. (2002). Collective Cognitive Responsibility for the Advancement of Knowledge. In *Liberal education in a knowledge society* (pp. 67–98). Chicago, Ill. : Berkeley, CA.: Open Court.
- Stahl, G. (Ed.). (2021). *Theoretical Investigations: Philosophical Foundations of Group Cognition*. Cham: Springer International Publishing.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: A Meta-Analysis. *Educational Psychology Review*, 29, 477–511.
- Weidlich, J., Kreijns, K., Rajagopal, K., & Bastiaens, T. (2018, June 26). *What Social Presence is, what it isn't, and how to measure it: A work in progress*.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12, 423–467.

## Acknowledgments

This work is part of the SituLearn project, supported by the French National Agency for Research, reference ANR-20-CE38-0012. The illustrations are the author's work.

## A Meta-Space for “Super Notes”: How Two Teachers Supported Cross-Classroom Collaboration for Knowledge Building

Yanqing Sun, Empire State University, SUNY, yanqing.sun@sunyempire.edu  
Jianwei Zhang, University at Albany, SUNY, jzhang1@albany.edu

**Abstract:** Existing research on teacher support for collaborative learning has focused on small groups in individual classrooms. The current study investigates how two teachers supported cross-classroom interaction between two Grade 5/6 classrooms. Students in the two classrooms studied human body systems for 10 weeks with the support of Knowledge Forum. While students in each classroom advanced their thinking in the discourse space of their home class, they had access to a shared meta-space for sharing knowledge advances across classrooms, presented in the format of “super notes.” Qualitative analyses of teacher interviews and classroom observations provided a detailed account of the teachers’ pedagogical approach to the meta-space and classroom support to guide students’ cross-classroom interaction.

### Introduction

Researchers have devoted efforts to understand teachers’ complex roles in computer-supported collaborative learning (CSCL) (van Leeuwen & Janssen, 2019). As students engage in collaborative learning, teachers co-engage with their students to understand their ideas, social/emotional interactions, and inquiry practices. Doing so help teachers identify emerging needs and opportunities and adopt responsive strategies to enhance students’ thinking and collaboration (Park & Zhang, 2023). Example strategies include promoting for clarification and explanations, asking questions, revoicing student ideas, facilitating connection-building among students, and reflecting on generative “big ideas” (Mitchell et al., 2017; Murphy et al., 2018; van Leeuwen & Janssen, 2019; Webb, 2009). As a guiding principle, teachers need to frame and position their input in a way that enhances students’ agency, so students take increasing control over what they do and how they collaborate (Zhang et al., 2018; Murphy et al., 2018). As a notable limitation, the existing research has focused on student collaboration in small groups or in individual classrooms. Researchers need to look for ways to sustain collaborative learning at larger social scales and higher social levels (Chen, Håklev, and Rosé, 2021; Stahl, 2013; Wise & Schwarz, 2017). The extended social interaction may give students the opportunity to conference with an expansive network of ideas, expertise, and people to support their knowledge building.

In an ongoing design-based research project, we work with teams of teachers to enable students’ collaboration across different classrooms and communities with pedagogical and technological support, including collaborating with real-world professionals (Zhang et al., 2020, 2024; Yuan et al., 2022). Drawing upon the existing literature (e.g., Laferriere et al., 2012), we developed a multi-layer interaction framework to support collaborative knowledge building across classrooms. While students of each classroom interact in their local discourse spaces (e.g., online forums), they have access to a cross-community space, or “meta-space”. The meta-space represents a higher-level of collaborative discourse shared among multiple communities of knowledge builders, who work together to solve complex problems drawing upon the knowledge built within each community. To contribute to the meta-space, members of each classroom need to reflect on what they have achieved through their inquiry and formulate accountable contributions of value to the broader communities. Students use the meta-space to share knowledge advances that have emerged from their home classroom. To make students’ knowledge advances sharable and accessible between classrooms, we tested having students compose “super notes” in the meta-space. While regular online posts (notes) tend focus on sharing specific questions and ideas, super notes offer a reflective view of student thinking in each line of inquiry, which may involve a series of online discourse entries (notes) in the home class space. A set of scaffolds (prompts) is designed to support super note writing, including (a) Our inquiry topic and problems, (b) We used to think...now we understand... and (c) We need deeper research. These scaffolds support students’ reflection on their inquiry directions and progress. The shared use of the scaffolds enables a common structure of super notes. Thus, the super notes may serve as boundary objects, with which members from different communities can understand and interact with one another’s work (Star & Griesemer, 1989).

A set of studies has been conducted in a network of classrooms (Zhang et al., 2020, Yuan et al., 2022). While members of each classroom work together to investigate various problems and deepen their understandings in their own discourse space, they post super notes in the meta-space to share knowledge advances. The findings suggest that the young students (10-to-12-year-olds) are able to compose reflective super notes to synthesize knowledge advances for cross-community sharing, capturing sophisticated scientific explanations and questions

developed in their home classroom. Social network analysis revealed intensive connections formed among the students within each classroom, between different classrooms, and across school years (student cohorts). Dialoguing with the ideas of different communities helped students to enrich and broaden their knowledge, engage in deeper reflection and inquiry, and further combine distributed expertise to investigate complex challenging issues. The ongoing sharing of super notes gave rise to “big ideas” and fruitful questions.

The current study aims to make a deeper dive in the teachers’ classroom practices to support cross-classroom collaboration. Specifically, we re-analyzed the data collected through a design experiment that tested cross-classroom collaboration for knowledge building across two Grade 5/6 classrooms (Zhang et al., 2020). The existing data analysis has examined students’ participation in cross-classroom collaboration. The current study views in the two teachers’ practices to support student cross-classroom interaction. Our research questions ask: How did the teachers frame the role of the meta-space for cross-community interaction? What support did the teachers offer to students as they work on the meta-space?

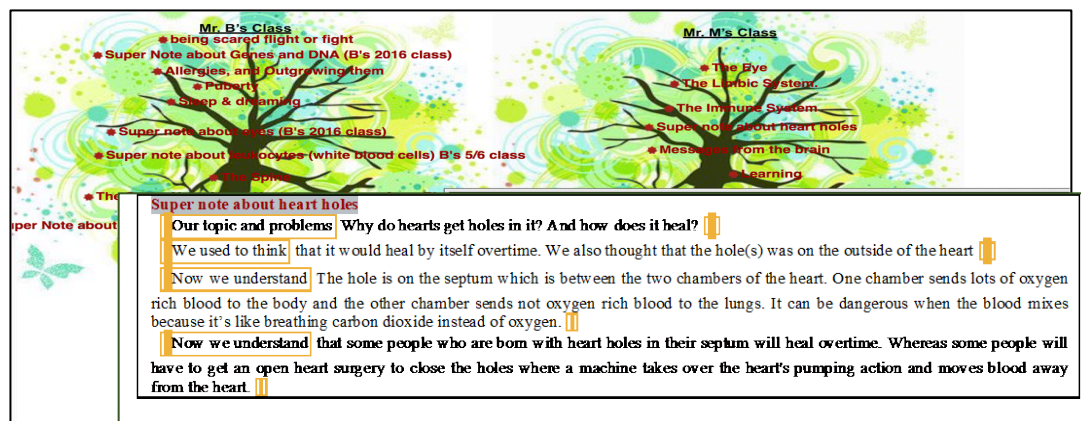
## Method

### Classroom contexts and participants

The participants were two teachers, Mr. B and Mr. M, who taught two Grade 5/6 classrooms, respectively, at the Dr. Eric Jackman Institute of Child Study laboratory school in Toronto. The school site has been implementing Knowledge Building (KB) pedagogy with the support of Knowledge Forum (KF) (Scardamalia & Bereiter, 2006) as a long-term innovation. Mr. B was a mid-career teacher with rich experience teaching science using the KB pedagogy and technology. Mr. M was an early-career teacher and was in the first year of teaching science using the KB pedagogy. There were 24 students in Mr. M’s classroom and 23 students in Mr. B’s classroom. The classrooms studied human body systems over a ten-week period. Students in each classroom generated questions about the human body, put forth initial ideas, and then subjected these ideas to testing through observations, experiments, and peer discussion to improve them. They further continued their discourse and interaction online in KF. Students worked in their home class views (workspaces) on KF, where they read and built on peers’ ideas in the online discourse. At the same time, they had access to the “Super View,” which served as a meta-space shared between the two classrooms. A visual background was added to the Super View consisting of two trees each with a number of branches where super notes about various inquiry topics could be placed (Figure 1). Prior to this study, several classrooms had studied human body systems. A set of super notes were created to summarize their knowledge advances and added to the Super View as a resource for the current classrooms.

**Figure 1**

*The “Super View” for Sharing Super Notes (Journey of Thinking syntheses) across Classrooms.*



Our research team and the teachers held monthly meetings to co-design the overarching process of cross-classroom collaboration, share and reflect on students’ inquiry progress, and identify ways to deepen students’ work. Each of the two teachers then worked out his own classroom arrangements with students’ input. The teacher in each classroom first introduced the Super View in the third week of the inquiry when their students had generated their own questions and conducted initial research. As students in each classroom conducted deeper research in the next two to three weeks, those working on various themes started to create super notes to summarize their knowledge progress for sharing with the other classroom. Students from the two classrooms read each other’s super notes and discussed insights gained.

## Data sources and analyses

Multiple data sources were used to analyze the teachers' role and practices, including classroom observations of each science lesson, video recordings of classroom discussions, students' super notes posted in the Super View, and teacher interviews. A researcher observed the science lessons in the two classrooms and took detailed observation notes. At the end of the science unit, a researcher conducted a semi-structured interview with each teacher using approximately 45 minutes, focusing on how they approached super note writing and sharing in the Super View as part of the knowledge building process.

To look at how the teachers understood the roles of the meta-space for cross-classroom sharing, two researchers analyzed the teacher interview data using a grounded theory approach (Strauss & Corbin, 1998). They individually read and re-read the transcriptions of the interviews, created raw codes, and clustered the raw codes into primary themes to capture compelling patterns. The researchers then congregated their codes through co-reviewing the raw codes, initial themes and examples and merging the common codes and themes. A set of themes was developed to capture teachers' framing of the various purposes the meta-space. These themes were cross validated by examining how the teachers described the meta-space (Super View) in the classroom recordings.

To investigate the classroom support provided by the teachers, we analyzed the classroom observation records and videos. Using a narrative approach to video analysis (Derry et al., 2010), a researcher first browsed the videos and transcriptions to develop an overall sense of the processes by which the teachers introduced the Super View, supported student writing of super notes, and facilitated the follow-up reading and discussions of the super notes. This was followed by the identification of classroom episodes that showcased major input or support from the teachers. These classroom episodes were transcribed for more detailed analysis.

## Results

### How did the teachers frame the role of the meta-space?

The analysis of the interview data identified the teachers' pedagogical expectations of incorporating a shared meta-space (Super View) in their classrooms. Their understandings and expectations of the meta-space were multifaceted, including leveraging the social drive for KB, synthesizing and benchmarking knowledge advancement, and supporting cross-boundary knowledge flow.

#### Meta-space leverages students' intentionality and responsibility for building knowledge

First, the teachers considered the meta-space as a way to give students an authentic purpose for their KB, which is to produce knowledge that benefits the broader communities. In the interviews, Mr. B and M both highlighted that the cross-community sharing served to enhance students' motivation and responsibility for contributing knowledge of help to other people in the same and different communities. Students from the two classrooms knew that they were studying the same science unit, so they were naturally curious about what the other class was doing and learning. It was motivating for students to come to the meta-space to see what other people were investigating in relation to their own works. As Mr. M said, "Students actually... discussed with each other after class on the topics that they didn't investigate but investigated by students in the other class. I didn't know that it's such a common thing for students. That's encouraging." Mr. B further emphasized the epistemic benefit of the Super View, which gave students a reason to deepen their inquiry and improve their ideas. "What you put out there... is somehow be of use." The Super View "adds a layer between the idea generator and idea receiver. You have to put it in a way that people can understand. I think it (the Super View) was helping in that role. Because it had that structure to it, it had the clarity to it, that makes it more accessible to other people, so it's more useful."

#### Meta-space supports metacognitive review and benchmarking of knowledge advancement

Second, the meta-space, together with the super note scaffolds, helped students reflect on and benchmark their progress in their collaborative inquiry. On an ongoing basis, students reflected on the "big ideas" learned as well as deeper problems to be further investigated, informing their plans for writing super notes. The teachers conveyed their expectation that students should think at a higher level and reflect about the "big ideas" in each theme of inquiry. As Mr. B said, "I try to use the super notes as a window, as a way of thinking. What I really want for them is to be thinking what the "big idea" within the human body is. Let's imagine you are adding a super note, you are interacting with students just beginning to study this. What would you tell them? What is important for someone to understand the human body. What is the most important part? How should they go about it? Super notes give us that rationale that asks them for deeper level or higher level thinking."

The teachers emphasized the value of the super note scaffolds in enhancing students' metacognitive review of their inquiry as a journey of thinking, being conscious about their previous thinking, improved understanding, and the knowledge gap. As Mr. M said, "What you used to think? Just discussing what their

previous thinking was... is a step so rich; but we often push forward and look at what else we can study instead of looking back. Just taking two minutes to look back is a huge huge learning opportunity. It's challenging for students. For example, EL, when I asked what he used to think, he said I don't know. I don't think about it. Stuff he learned he thought he knew already. But when he came down to it, and wrote down his previous thinking, he used to think, then something [new] learning, he learned different types of learning, short term memory, long term memory. It was more to it. He started to appreciate it how complicated it was. But even that, it was eye opening for him. So the scaffolding is definitely useful." Aligned with Mr. M's comments, Mr. B noted that super note writing helped students become more metacognitive. "I try to let them know what this sharing is about. It's not about sharing every detail you learned... It is sharing of the journey of your thinking. What is you used to think about, what is your understanding now, with an eye to helping someone who would benefit from that kind of thinking, not just telling them what it is." The teachers particularly emphasized that the super note was not to draw an end to the inquiry but to inform ever-deeper inquiry directions. As Mr. M said, "Because of the way the scaffolds were set up, they were asked where their future needs for research are, what else there are to be understood. That led students to go deeper... We try to move away from the thought that the super note being the end product. It's more part of the process but at different levels... If there's more time, they could write the super notes, revise, update, improve it."

### Meta-space supports knowledge flow across boundaries

The Super View was used to support student mutual learning between different communities. Students benefited from both contributing super notes and reading the super notes of others. According to the teachers, reading the super notes from different classrooms gave students the opportunity to see a diverse range of interests, ideas, and inquiry practices from different classrooms that worked on the same curriculum unit. As Mr. M said, students could look at the questions raised by the other classrooms, the ways used in their investigation, the "big ideas" they pulled out of it, and the ways of summarizing their ideas. "That gives them more information to work with. Just get them to think about things in different way." On the other hand, preparing content for writing their super notes helped students deepen their own inquiry and understanding. As Mr. B commented, the process of preparing stuff that they could include in their super notes pushed students to go deep with their inquiry, so they could put something useful on the super view. Writing super notes was "probably more valuable than what they got from reading from that view. At least, they are of equal value."

Beyond the online super note posting and reading, Mr. B and Mr. M further facilitated face-to-face discussions in their own classroom to reflect on what students had learned from different classrooms. Students reflected on new knowledge gained: "seeing the things we didn't see, haven't investigated yet, broadening our perspectives of human bodies, having a variety of topics to look at. It pushes students to think more about what topics to investigate. It's also broadening what we had looked at." (Mr. M)

### **What support did the teachers offer to students as they worked in the meta-space?**

The analysis of classroom videos and observations identified the specific classroom support provided by the teachers to support student work on the meta-space.

### Introducing the Super View as a new discourse space shared between classrooms

During the third week of the human body inquiry, the two teachers met to discuss the use of the Super View, and then each of them introduced the Super View to their own class. Both teachers described the Super View as a shared place where students from different classrooms could share their summaries of "big ideas".

Mr. B first shared his reflection on the existing Human Body view (space) used for within-class discussion. As he said, the discourse on this view had a huge amount of information, hard to be shared with other classrooms. Then he showed the Super View as a new space, noting that it could be used to share significant conversations with future classes, and with Mr. M's class in the current year. "We could share using those super notes, notes that sum up our thinking about what we are doing... It would be great if some people know they have made progress in one area, they want to get together and create one of these super notes. It can go in here and get shared with Mr. M's class. Future classes can look at these as well."

In Mr. M's classroom, the teacher introduced the Super View in a similar manner, while highlighting that the super notes were not about summarizing what they had researched already but to bring their thinking to the next level and synthesize important ideas for others to know. Students' writing of super notes was not limited to their own inquiry topic but could include any important points that the community should know. He said: "And there already is a ton of information that's on in the different (home class) views. But now our job is... to see what the "big ideas" are, and what are going to be the really important ideas for everybody to know... Once you've gotten to a point where you think you can synthesize and bring together a few different points that you've

researched and really make kind of an important discovery or synthesis of what you've been researching, that would be the time to come to this Super View.”

Both teachers talked about the background image of the Super View with two trees (see Figure 1), using the metaphor that the two classrooms were growing their trees of knowledge with some overlap, and the different branches of each tree would represent different “big ideas.” The teacher in each classroom showed the example super notes created based on the discussion of the previous classrooms studying the human body unit. The teacher read the super notes aloud while pointing out the scaffolds used. Doing so helped students get clear sense of what the Super View was for and how they could contribute to this new space.

#### Facilitating meta-talk about “juicy” inquiry areas and knowledge advances

The teacher in each classroom facilitated a meta-talk (metacognitive conversation about the ongoing discourse) about the various areas of inquiry represented in students’ classroom work and online discourse, focusing on identifying “juicy” areas of inquiry for future super note writing. On April 28, right after the introduction of the Super View, Mr. B asked students to reflect on what they had generated on KF or in personal notebooks. “So who knows right now that they they've done a fair amount of work in a particular area?” Students identified topics such as spine and spinal cord, eyes, brain, sleep, genes and DNA, and puberty. Mr. B recorded the topics on the blackboard. Then, he encouraged students to form groups according to their questions of interest. Members of each group further reflected on what they had learned and deeper questions they were still wondering about, informing their writing of super notes.

Mr. M also facilitated the conversation about “juicy” inquiry areas in the subsequent week, not in the same lesson that introduced the Super View. His plan was to give students more time to do research before writing super notes. On May 9, he asked: “I would like to know where you are on your investigation?” Students shared progress in their understanding of the various topics, including healing, memory, learning, brain and brain damage. Students in both classrooms were encouraged to identify additional “juicy” areas of inquiry as their work proceeded and form new groups to write super notes. While sharing a few common areas of interests, each classroom dived in a few unique topics that the other classroom did not explicitly discuss.

In each classroom, the teacher encouraged each student group to review their online posts and personal notes to identify knowledge advances as well as gaps, such as questions and issues that needed to be better understood. The teacher walked between groups to listen to students’ thoughts, offered input to help clarify their ideas and questions, suggested resources that they might use to do deeper research, and gave guidance on how they might collaborate on the super notes. For example, a girl in Mr. M’s room working on a super note about eyes attempted to figure out what role nerves play in the eye and what would happen if the nerves were disconnected from the brain. These issues pushed the students to conduct further research using books and websites. Some of the sources were beyond the students’ reading level. The teacher worked as a co-learner and helper to interpret the information, explain scientific terms, and model rephrasing ideas using simpler terms. The needs to contribute to the community’s knowledge motivated students to seek deeper thinking, read more closely across sources, and reflect on what they had figured out as well as issues to be better understood.

#### Facilitating super note writing

The teachers provided guidance to help students understand the purpose, structure and style of super note writing. Such support was important as super note writing was new to the students. Both teachers highlighted how to use the super note scaffolds: Start with “we used to think”, and continue with “now we understand”, and finish with “we need deeper research.” The teachers further facilitated reflective discussions on what might be worthwhile to include in a super note. For example, on April 28, Mr. B facilitated the following discussion.

Mr. B: When writing a (super note) summary, would you want to put in here every detail that you've researched? [A few students say “No.”] Why not? ... Why is that not useful for other people, future class? You have done a bunch of research on DNA, why don’t you want to put all in there?

S1: Because no one's going to read it.

S2: Exactly, no one's going to read it because it gets boring some time...

[Other students echo on the same point]

S3: It's not like we have collected all this information in half an hour. We've done it over weeks. So it is really interesting to us. But if someone had a really big paragraph about it, they're probably going to be either confused or not really interested. And also, it's going to become really big. [Mr. B: Yeah.]

S4: It's like an overview of what you've been working on.

S5: You don’t want people to do the research you have done by reading your note.



Mr. B: Yes, instead of trying to make everyone do the research that you've done through your notes, you want to provide a summary of what the big things you've learned are. That's what will help other people.

The conversation about the purpose and features of super notes helped students to approach their super note writing as individuals and groups. For example, the student group working on a super note about DNA engaged in shared reflection to identify the “big ideas” related to DNA. As a member commented, “We tried to determine the biggest thing in DNA, and how that worked.” They composed their super note focusing on the essence of DNA: “Now we understand that DNA is the building blocks of life. It acts as a code for the cells. This code instructs the cells to produce different body proteins.” As students worked in groups to draft super notes, the teacher walked between groups to understand the questions and ideas synthesized in their inquiry area. Mr. M provided more support to the group process. His support included helping students clarify ideas, refine their questions, and showing the use of scaffolds. Students in the two classrooms generated 16 super notes in total, which presented deep, explanation-seeking questions and elaborated scientific explanations (Zhang et al., 2020).

### Facilitating reflective reading and discussion of super notes from other classrooms

Besides having students engage in spontaneous individual reading of the super notes, each teacher integrated whole class reading and discussion of super notes in the different phases of the cross-classroom interaction, serving several different purposes. Early on during the introduction of the Super View, each teacher facilitated student reading of the super notes from the previous classrooms, as examples of super note writing. Student reading of the super notes triggered their reflection on their own inquiry questions and progress, including considering what they can additionally contribute to the shared knowledge space.

Mr. B: (opens a super note and reads it on the projector screen) So “we used to think that the eye moves without any nerves or veins connected to them. Now we understand that the eye has so many different nerves and veins and muscles in their eye...The optic nerve connects the eye to the brain. When it connects it helps you see and move you your eye.”... So um actually in our class, some people have been working on the eye and many people have read things about the eye. What have we worked on, like who could add something to that about the eye in particular?...

S1: So like, you look, you see it um, it goes then into your pupil and then...what is it, what's the thing around your pupil again? What's it called?

Mr. B: Cornea?

S1: Yeah your cornea, and then so then it goes to your pupil then it goes to... the back of your eye, and then it goes down toward your nerve, and then it goes up to your brain.

Mr. B: That's where you understand that, actually happening in your brain. S2, you have done something about rods and cones that apply to the eye. Can you tell us a little bit about that?

S2: Um well, basically in your eye there are these literally rods and cones. So basically the rods are the things that help you see shapes and black and white. And then the cones consist of the primary colors so and then they all blend together to make the other colors. So for example, if I'm looking at S3, um I can see like the black on her pants and I don't need the cones for that it's because I can see the shades but... I need the cones to see what color S3's shirt is.

Mr. B: Cool! Okay. So there're more details to that question. That's great.

(opens and discusses a super note created by a different school about blood marrow)

Mr. B: I think it would be fair to expect that every person here will be listed at some point as an author on one of these notes. Show that you've made a major contribution to our work together as a class. So I think that we can probably expect that you will see in our main unit that there's a lot to work on and a lot to read and a lot to research.

In this discussion, Mr. B modeled reading the super note about vision from a prior classroom, highlighting the scaffolds “We used to think...Now we understand...” He further modeled building connection with the inquiry of the current class about eyes, encouraging students to add their knowledge. He facilitated student discussion on how eyes work, revealing students' deep understandings of color vision and eye-brain connection. Finally, he highlighted that students take collective responsibility for contributing to “our work together as a class,” expecting that every student would contribute to the super note writing.

Toward the later phase of the human body unit, students in each classroom had created a set of super notes to share their knowledge advances. Each classroom had a whole class discussion about the information gained from the super notes of their partner classroom, followed by further small group discussions focusing on the super notes most relevant to their own inquiry. Through qualitative analysis of the video records of the whole

classroom discussions, we identified specific patterns characterizing how students worked with the super notes to enrich their knowledge building efforts. These included encountering a broadened scope of ideas and interests for mutual enrichment, reflecting on different ideas and perspectives for deeper inquiry, and rising above knowledge from the different areas and communities to formulate a higher-plane view of the human body systems. For example, Mr. B's students conducted specialized inquiry in allergies and created a super note in this area, which was not covered by the students of Mr. M. During the classroom discussion, Mr. M asked his students: "What topic either strikes you as new information or something that you'd like to pick up as a thread and go deeper into? Were there new topics that you've come to?"

- S1: I was really interested in [student]'s note about allergies... Allergies are really interesting to know about. And it could be like basic allergies just like, the symptoms of allergies like sneezing and runny noses. But sometimes, allergies are actually more severe... And then, I started to research about allergies, and how nut allergies [work]... So I am trying to kind of [research] why it is so hard for people to grow out of allergies, like nut allergies, or like peanut allergies... Why it is so much harder and what does it have to do with your immune system?
- S2: Yeah, but, there is so much more to it, like, what causes allergies, also like in your body... sometimes, like nuts, could really change it. And also allergies are really bad for your body, but like we never think that could be that severe.
- S3: Me and S4 are doing immune system, and we saw these notes [from Mr. B's class] about white blood cells, and that was really cool because white blood cells were part of your immune system. We don't really know about them individually...
- Mr. M: Interesting, so kind of related to what you are doing.
- S3: Yeah, it was related but also seeing how it's a tiny little thing, and it has like, there is like a system within a cell.
- S4: And there are like steps that are used...to help like to kill off the bad invaders, so there are five of them, we are trying to understand that a little bit more, and put it into our note, like our words, of course.

While each super note offered a synthesis of a line of inquiry in a classroom, the extended conversation taking off from the super note sharing helped students integrate what they learned, connect the different lines of inquiry, and formulate high-level understanding of how the human body systems work together.

## Discussion

The findings help the field understand how teachers may extend their CSCL practices to incorporate a meta-space of discourse for cross-classroom collaboration. The teachers' pedagogical framing underlined the multiple affordances of the meta-space, including leveraging the social drive and responsibility for building deep knowledge, synthesizing and benchmarking knowledge advances achieved by students in each classroom, and supporting cross-boundary knowledge exchanges beneficial to both the contributors and the readers/learners. The teachers' pedagogical framing informed specific classroom practices to support student interaction in the meta-space. They further shared their understandings of the meta-space with students to guide their participation. Specifically, the two teachers' introduction of the Super View highlighted the social and epistemic expectation for the meta-space: as a higher-level space for sharing significant knowledge advances in "juicy" inquiry areas with the broader classroom communities. Accordingly, the teachers facilitated meta-talk among their students to reflect on their diverse lines (areas) of inquiry and identify knowledge advances. Meta-talk, sometimes referred to as "metadiscourse," is an important feature of knowledge building discourse in which students enact epistemic agency for charting the course of collaborative inquiry (Lei & Chan, 2018; Zhu et al., 2022), yet it rarely occurs in classroom discourse. In this study, the incorporation of the meta-space helped to call out the need of meta-talk for students. Students engaged in meta-talk to review their inquiry themes and advances and generate metacognitive artifacts in the form of super notes, which summarized deep questions and elaborated explanations (see details in Zhang et al., 2020). The super notes posted in the meta-space served as boundary objects (Star & Griesemer, 1989) to bridge the knowledge of the different communities. The scaffolds of super note writing provided a common structure for students from different classrooms to easily understand one another's knowledge advances shared in the online meta-space. The teachers further facilitated face-to-face conversations for students to reflect on what they had learned from the super notes of prior and the parallel classroom.

In conclusion, the current study revealed teachers' productive framing and classroom moves that are prominent in cross-community collaboration in a shared meta-space. These include teacher support of students' meta-talk on knowledge progress and gaps, generation of knowledge artefacts that can serve as boundary objects for cross-community exchanges, and integration of diverse contributions from different students and communities

for idea development. These findings contribute to enriching the literature on teacher practices in CSCL settings, which has focused on small groups in individual classrooms (Murphy et al., 2018; van Leeuwen & Janssen, 2019; Webb, 2009). The extended social interaction through the meta-space provides a way for sustaining CSCL at larger social scales (Chen, Håklev, and Rosé, 2021; Stahl, 2013; Wise & Schwarz, 2017). The social extension of discourse serves to enhance students' metacognition and meta-talk about their ongoing inquiry, leading to the creation and sharing of knowledge artifacts shared between communities. Building on the findings, we continue to examine teacher co-design of cross-community knowledge building in broader settings and create technology systems to support classroom implementation.

## References

- Chen, B., Håklev, S., & Rosé, C. (2021). Collaborative Learning at Scale. In U. Cress, C. Rosé, A. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning*. Springer.
- Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences*, 19(1), 3–53.
- Laferriere, T., Law, N., & Montané, M. (2012). An International Knowledge Building Network for Sustainable Curriculum and Pedagogical Innovation. *International Education Studies*, 5, 148–160.
- Lei, C., and Chan, C. K. (2018). Developing metadiscourse through reflective assessment in knowledge building environments. *Computers and Education*, 126, 153–169.
- Mitchell, I., Keast, S., Panizzon, D., & Mitchell, J. (2017). Using 'big ideas' to enhance teaching and student learning. *Teachers and teaching*, 23(5), 596-610.
- Murphy, P. K., Greene, J. A., Firetto, C. M., Hendrick, B. D., Li, M., Montalbano, C., & Wei, L. (2018). Quality Talk: Developing Students' Discourse to Promote High-level Comprehension. *American Educational Research Journal*, 55(5), 1113-1160.
- Park, H., & Zhang, J. (2023). Teacher Reflective Noticing and Scaffolding for Student-Driven Knowledge-Building Inquiry. In P. Blikstein, J. Van Aalst, R. Kizito, & K. Brennan (Eds.), *Proceedings of the 17th International Conference of the Learning Sciences* (pp. 1054-1057). Montreal, QC: International Society of the Learning Sciences.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97-118). New York: Cambridge University Press.
- Stahl, G. (2013). Learning across levels. *International Journal of Computer-Supported Collaborative Learning*, 8(1), 1–12.
- Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19, 387–420.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks: Sage.
- van Leeuwen, A., & Janssen, J. (2019). A systematic review of teacher guidance during collaborative learning in primary and secondary education. *Educational Research Review*, 27, 71-89.
- Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, 79, 1–28.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467.
- Yuan, G., Zhang, J., & Chen, M.-C. (2022). Cross-community knowledge building with Idea Thread Mapper. *International Journal of Computer-Supported Collaborative Learning*, 17(2), 293–326.
- Zhang, J., Song, G., Underwood, T., Guo, D., Espinosa, K., Hower, L., Muirhead, J., Kerr, J. (2024). Cultivate Research Literacy: High School Students Work with Community Partners to Investigate Real-World Challenges. In Proceedings of the 18th International Conference of the Learning Sciences. Buffalo, NY: International Society of the Learning Sciences.
- Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with Idea Thread Mapper. *Journal of the Learning Sciences*, 27(3), 390-430.
- Zhang, J., Yuan, G., & Bogouslavsky, M. (2020). Give student ideas a larger stage: Support cross-community interaction for knowledge building. *International Journal of Computer-Supported Collaborative Learning*, 15(4), 389–410.
- Zhu G, Scardamalia M, Moreno M, Martins M, Nazeem R., & Lai Z (2022) Discourse moves and emotion in knowledge building discourse and metadiscourse. *Frontiers in Education*, 7:900440.

# Shall We Play a Game? Distributed Games with a Generative AI Player

Jenna Matthews, Ha Nguyen, Hillary Swanson  
jenna.matthews@usu.edu, ha.nguyen@usu.edu, hillary.swanson@usu.edu  
Utah State University

**Abstract:** The prevalence of generative AI tools – especially large language models that enable human-like conversation – provides an opportunity to study how human learners interact with AI as collaborative ideators in the learning process. We expand the framework of distributed games (which initially focused on epistemic games of discovery as distributed processes) to investigate learner interaction with generative AI in design tasks. Through a comparative case study of two undergraduates using a generative AI tool in design tasks, we illustrate different ways in which learners orient themselves to the elements of the game (task, form, other players, etc.) and the corresponding complexity of the game played. We discuss implications for the design of learning environments that facilitate interaction with AI tools in distributed creative tasks.

## Introduction

Advances in generative artificial intelligence, particularly large language models that can generate human-like, conversational text, have promise for supporting teaching and learning (Kasneci et al., 2023). To promote learning, it is critical to examine how learners approach tasks with AI tools. In this study, we explore the way different participants interact with a tool in a distributed design game – a deliberately collective effort to complete a design task. Participants were given a design task within a User Experience Design context and engaged in a brainstorming session with OpenAI’s ChatGPT to develop and refine design ideas. We present two case studies of undergraduate participants. We identify the participants’ different framings of the task and the role of the AI tool and present narratives to show connections between the framings, the design games, and the produced design sketches. The following question guides our research: *How do designers-in-training orient themselves to a design task with a generative AI tool?* Insights from this research have practical implications for guiding students to interact with AI systems in educational contexts.

## Theoretical framework

The model of distributed games focuses on collective activity. We expand an existing framework (*epistemic games*) to describe another type of distributed game (*design games*), in which the goal is to generate a design.

## Distributed games

The conceptualization of knowledge construction as an epistemic game guided by an epistemic form has been in use for several decades (Collins & Ferguson, 1993). This framework visualizes knowledge construction as the result of a scholar making *moves* in a *game*, which are made to fill out a particular template – for example, by filling in the rows of a list. Recent work has expanded this conceptualization with the added lens of distributed cognition (Hutchins, 2008) into a theory of *distributed epistemic games* (Matthews & Swanson, 2023; Matthews, Nguyen, & Swanson, 2023). Distributed cognition adds the importance of other elements in the thinking environment - including tools, artifacts, and other individuals - to the thinking that takes place. In a distributed game, the use of forms with moves made by players persists. The difference is that multiple players are playing the game simultaneously, often filling out multiple forms and creating *artifacts* as repositories of the knowledge generated by the game (*ibid*). Initial work in the distributed game framework has focused on a game which was played exclusively by human players (*ibid*).

In this paper, we expand previous work in two key ways. The first is that we look at games which involve the creation of a *design artifact*. The second is that we look at distributed games including an AI player.

## AI players

Advances in AI, including development in large language models that enable conversational interfaces in the form of generative LLMs, have made it possible to position AI tools as collaborators in the design process (Nguyen & Hayward, 2024; Wang et al., 2020). These tools can not only generate novel ideas, but also engage in idea co-creation with human designers to continuously improve upon the design (Davis et al., 2015; Simeone, Mantelli,

and Adamo, 2022). While promising, these tools can also be challenging to navigate, as designers need to articulate their goals, understand how to approach the tools, and continually assess the AI's outputs (Gmeiner et al., 2023). Thus, scholars have turned to educational research in human-human collaboration to inform the design of human-AI collaboration (Holstein et al., 2020; Schelble et al., 2022). For example, Bansal et al. (2019) investigated how to facilitate shared mental models, or understanding of the task and the AI's capacity, to adapt how human actors collaborate with the tools. Our study contributes to this emerging research through the lens of distributed games. Here, the AI tools can be positioned as another player contributing to the creation of design artifacts.

### Orientation

Orientation plays an important role in the framework. The way players orient themselves to the game and other players impacts the moves that they make. We differentiate this notion of orientation from concepts like task orientation, which focuses on goal-driven learning activities (Pintrich, 2004), and note the similarities between orientation and the different ways in which individuals engage with AI during design tasks (Simeone, Mantelli, and Adamo, 2022). In our definition, a player's orientation within a distributed game reflects their prior experience, expertise, and perspective.

A player's orientation impacts the objects of a player's *attention* - what they attend to such as the environment, tools, and players - as well as their *interpretation* - how they make sense of their objects of attention. Orientation also includes a player's *familiarity* with the objects of attention.

## Methods

### Study setting and participants

This study included 17 designers with different design expertise in Summer 2023 (Institutional IRB #13497). Participants included seven professionals in UX/UI and instructional design, four graduate students, and six undergraduate students in a design program in the Intermountain West region of the United States. Participants were recruited via a flier disseminated through the program's listserv for current students and alumni, to invite individuals interested in using generative AI in design contexts. Participants received a \$10 gift card for completing the interview.

The interviews were conducted via video conference call and lasted 45 minutes on average. They were video and audio-recorded and transcribed automatically by the video conferencing software. During the interviews, participants received a task to redesign the navigation of a learning management system to enhance the user experience of instructors and students. Participants first engaged in individual brainstorming (10 minutes) to gain familiarity with the design space (e.g., identifying key users, researching user needs, brainstorming design solutions). They next used a free generative LLM tool to refine their design ideas (10-15 minutes). The chat interactions with the AI were screenshared, and interviewers prompted participants to think aloud with questions such as "Why did you ask [the AI] that question?", or "What do you think about the AI's response?" Participants had 5-7 minutes to sketch design solutions on paper, based on their individual brainstorming and interactions with the LLM. They verbally presented the sketches and then took photos of the sketches and emailed them to the interviewers. The interviews ended with a debriefing session, where participants outlined what they liked and wanted to improve about the interaction.

### Data sources

Our analysis drew from multiple data sources. To understand the *design moves* that participants engaged in within the distributed design game, we focused on the video recordings and interview transcripts. We specifically examined the human participant's utterances *within* human-AI exchanges (both think-aloud and prompts). Each utterance, prompt, and response was coded as a move in the game.

Additionally, we evaluated the *quality of the design sketches* that participants developed, following their brainstorming with the generative LLM. The rubric (shown in Figure 1) comprised four categories on a scale of 0-5: usefulness, novelty, elaboration, and incorporation of human and AI ideas. The first two measures were linked to how design researchers have defined creativity (Runco & Jaeger, 2012). We added a measure for elaboration, to explore how participants specified details for their design (Dally et al., 2016). Finally, the human-AI idea incorporation category indicated the extent to which the design sketches incorporated both players' ideas, as a marker of design product co-construction. We summed up the scores per participant for the categories:  $M = 10.58$ ;  $SD = 2.49$  (possible maximum score of 20; range of 6-15).

**Figure 1**

*Design Sketch Rubric*

**Idea Description:** A more simplified, easier to navigate version of Canvas dashboard ... divided by subject ... swapping the to-do list up and having it in calendar views. I also included video help and FAQs. And I added a "message the teacher" button. [...] Originally, I was thinking to condense the sidebar menu, but **talking to ChatGPT made me realize that the problems came from navigating Canvas**. So I added the video questions and tutorials. When I started talking about special education educators and **we did the user profiles [for ChatGPT], I thought of accessibility** because it's annoying for my special education students to scroll so much even for one week, so then I thought of the whole week layout.



**Novelty** (0 = solution already exists; 5 = solution is entirely different from what exists)  
Score of 2 (the main solution of swapping or omitting sessions already exists).

**Usefulness** (0 = does not address prompt; 5 = solves the design problem effectively)  
Score of 5 (solves the problem of navigation; considers several use cases for higher education as well as special education students in K-12).

**Elaboration** (0 = does not include details; 5 = includes enough detail that a designer could implement the solution)  
Score of 4 (includes sufficient details in sketches and verbal descriptions for the sidebar menu; needs more details about the FAQ feature and implementation).

**Idea Incorporation** (0 = does not include AI conversations; 5 = incorporates several points from conversations into solution)  
Score of 3 (brings in icons and use cases from conversations with AI; largely one's own thinking about the sidebar menu).

## Analysis

### Case selection

In addressing our research question, we focus this paper on a comparison between two undergraduate students or *designers-in-training*: John and Barbara (pseudonyms). The two participants had similar schooling experiences, as both were entering their senior years in their Design undergraduate degree. They were selected through the extreme sampling method based on the participant scores on the design sketch rubric (John: design score = 13/20; Barbara: design score = 9/20). The participants represented the highest and lowest scores among the undergraduate participants. When asked about their prior experiences with the AI (generative LLM), John brought up personal use such as creative writing & brainstorming, while Barbara had not used or heard of this generative LLM before the interview. Both participants spent roughly the same amount of time with the AI (~12 minutes).

### Analytic procedures

We answered our research question in three steps. First, we developed a codebook for participants' design moves through a grounded theory approach (Charmaz, 2006). The analysis unit was at the sentence level and we focused on participants' utterances during the section of the interview in which they were creating *prompts* for the generative AI. At the beginning of the process, the first author identified various design moves (see Table 1) within the transcript data. These codes were then refined by all authors in three discussion rounds. Next, the first and second authors collaboratively coded both transcripts, using a process of social moderation to discuss and resolve any disagreement (Frederiksen et al., 1998). Next, we wrote up the case narratives for the two selected participants which included a description of the sequences of the design moves during the brainstorming sessions. We followed this with a cross-case comparison to illustrate how participants approached the task and the collaboration with the AI tool.

**Table 1**

*Codebook of Design Moves*

Move	Definition	Example from transcript
Orient to task	Clarify and get information about the task and state perceptions of the AI tool	What do you mean? So, we're attempting to redesign ...
Find approach	Find approaches to working with the AI	We can give it the prompt for what we want it to do [...] so now it will act as a UX/UI designer.
Prompt	Prompt or refine prompts to the AI	Can we ask, is the system useful to students?
Ideate	Brainstorm ideas independently or with AI	I feel like a progress tracking tool would be really good.
Review	Read and review AI's responses for their accuracy and helpfulness	I feel like it's actually everything I said. That's pretty cool.

Reflect	Reflect on the design space, drawing from personal experience or design feasibility.	It's nice to see as a student, a good tracking point.
---------	--	---

## Findings

### Session overview

We first present an overview of the brainstorming sessions with the generative LLM AI based on coding participants' think-aloud utterances. We found iterations of *task orientation*, *finding approach*, and *prompting* in John's session. Meanwhile, Barbara's session was characterized by long periods of *finding approaches*. While participants spent the same amount of time with the AI, John sent seven prompts to the AI, while Barbara sent three prompts.

We delve into these patterns in more detail when analyzing the prompts that participants posed to the AI and how they evaluated the responses. Specifically, we found two macro-level games that participants were playing with the AI tool: *guided roleplay* (John) and *defining the design problem* (Barbara).

Within a distributed game framework, we use *turns* to indicate cycles where *players* make *moves* in sequence. A *player* may make multiple *moves* in a single *turn*. Each *turn* can be a level of analysis, with one or many *turns* comprising a micro-level *game* within the macro-level *game*. In this section, we review the turns of each player's game, with each turn marking a prompt that participants submitted to the AI. A turn can involve multiple moves, such as finding approaches before submitting a prompt, or reviewing response following prompt submission. Each turn is numbered and given a descriptive (and bolded) name, and the moves (as presented in Table 1 above) are also bolded. As in the rest of the paper, we continue to use italics for emphasis.

**Figure 2.**  
*Overview of the Brainstorming Sessions*



### John's distributed design game: Guided roleplay

John's game is marked by a sequence of *finding approaches*, *prompting*, and *reviewing* (see Figure 2). In this section, we present a narrative of the specific moves made during his game play centered on the prompts that he sent. We characterized this sequence of moves, at a macro-level, as fitting patterns of a design game of *guided roleplay*. Here, John specified both role and goal for the AI player, and then moved through a sequence of micro-games, including *setting up roles* (turns 1, 2), *decomposition of design steps* (turns 2-5), *making lists of design challenges* (turn 6), and *compare/contrast* solutions in a problem-centered analysis (turn 7).

**Turn 1. Defining AI role & checking for understanding.** Prior to sending the first prompt, John spent several utterances articulating his *perceptions of the tool* and **finding [his] approach**. He stated: "So there are a couple of ways we can approach it, and it's all in that prompt engineering, right? ... the things I do a lot with Chat GPT, GPT is essentially like that role play idea, like I tell it you're going to act as this kind of person with this kind of knowledge." With this framing, John's first **prompt** to the AI aimed at *defining roles and checking for understanding*. The prompt stated: "For the duration of this conversation, *you'll act as a UX/UI designer with 10+ years of experience*. I'll approach you with a redesign project, and you'll give me a design outline that highlights our user base, what features need to be reworked, and a list of potential solutions. Do you understand?" The AI responded: "Yes, I understand. As a UX/UI designer with 10+ years of experience, I can assist you with your redesign project. Please provide me with the details of the project, including your user base, the features that need to be reworked, and any specific goals or challenges you'd like to address."

**Turn 2. Defining the goal for the conversation.** John then moved on to his second **prompt** to *define the goal of the conversation* and generate a list of the design steps. He specified: "We are *attempting to redesign the Canvas learning management system*, so that it'll be more user-friendly ... We need to figure out what features are less user-friendly and identify solutions that will help Canvas be more competitive in the coming years." The AI player responded with an outline of the design approach, specifying each step such as User Research, Defining User Personas, and Feature Analysis.

**Turn 3. Requesting a mockup of a design step.** John's next **prompt**, after **reviewing** the previous move, was to get more details after the "pretty generic response," and *request a mockup of the user research section*. The AI responded to John's next prompt: "As a text-based AI, I'm unable to provide visual mockups directly." The AI then listed content that the first User Research step might include (e.g., header, introduction, methodology). John **reviewed** the answer and remarked, "Not quite what I was looking for, I think the word mockup threw it off." He stopped the AI mid-response and moved to the next prompt.

**Turn 4. Adjusting the prompt.** John then edited **prompt #3** to request an example of the design step. The AI's response to this included an outline with several elements. For instance, the AI suggested asking: "What are the major pain points or frustrations experienced while using the Canvas system." John **reviewed** the response. "This is a lot better. It's a bit more focused and has given us some pretty decent research questions" and **reflected** on its utility to the design process. "I couldn't just run with this, I'd have to go and create step one myself, but this gives me a good guideline." He then **reviewed** the AI's prior response to prompt #1 (listing the design steps) and continued with his next prompt.

**Turn 5. Reminding the AI of the role; asking for elaboration on the next design step.** John moved to the next step of the design process, Defining User Personas, based on the AI's initial outline. He **prompted**: "Remembering that you are a UX/UI designer with 10+ years of experience, complete step 2 of the outline that you provided, using hypothetical users that represent the Canvas system." In response, the AI created two personas: Emily, a college student, and Professor Mark, an instructor, along with the background, goals, behaviors, and pain points (i.e., challenges/frustrations with the interface) for each user. It identified a challenge for Emily as "difficulty locating specific course materials or assignments," while it suggested that Mark might find the "process for uploading and organizing course materials cumbersome." In **reviewing** the response, John made connections with the ideas he had developed initially in the individual brainstorming session, "the pain points this is highlighting are pretty similar to the outline that I had gone through." At this point John moved on to generating design solutions.

**Turn 6. Requesting a list of solutions to the pain points of each persona.** John continued building on previous exchanges with the AI, and **prompted**, "As a UX/UI designer of 10+ years of experience, take those user personas and generate 4 to 5 solutions per user for their pain points." In turn, the AI generated several solutions to address Emily's and Mark's (the hypothetical personas) challenges with LMS navigation. It proposed developing "enhanced course material organization," including "a tagging or search functionality," to "make it easier for Emily to locate specific resources." In **reviewing** the AI's answer, John exclaimed "Wow, like it's good. It's not detailed, but I'm still kind of impressed." He then strategized with the prompt to gain more details into the solution.

**Turn 7. Requesting detail for design solutions.** John posed a **prompt** to the AI to "give me some example implementation for solution 1." The AI responded: "Certainly, here are some example implementations for solution 1, which focuses on enhancing course material organization," and listed solutions such as "visual folder structure," "tagging system," and "personalized favorites or bookmarks." John **reviewed** the answer and noted what stood out to him, e.g., "A tagging feature, under number 2. I kind of like that. I think that could potentially simplify things. I want to look into that." John then dived into **ideating** and **reflecting** on the design feasibility for several utterances. He concluded the brainstorming session with a reflection note, "This is a really good way to get a good general chunk of ideas that you can kind of start working with."

John's process stood out to us, as he was engaging with the AI in a game of guided roleplay, where the AI was given not only tasks, but a character (role) to *play* for the duration of the *game*. This role was defined at the beginning of the conversation, and John reminded the AI of the role in several prompts. Additionally, John moved through multiple requests to play different mini-games within the larger macro design game of guided roleplay, such as *generating lists*, *requesting a mockup*, and then *creating design solutions*. Following the session, John's design sketches integrated several ideas from his conversation with the AI, including a tag system incorporated into both student-facing and instructor-facing interfaces.

## Barbara's distributed design game: Defining the design problem space

Barbara's macro-level game is a search for a good approach to engage with the task and AI player. Early in the AI-brainstorming session, she worked with the facilitator to review a sample prompt to better understand how the interface works. While John was able to engage with the AI tool independently, Barbara turned to the interviewer in the beginning to request reminders about the task and guidance about how to create a prompt. Barbara's prompts followed a consistent pattern of *list-making* throughout the macro-game of *defining the problem space*.

**Turn 1. Requesting a list of common features of an online course.** Following several utterances to **orient to the task** and **find approaches** to working with the LLM (e.g., "I don't know how to put that in words"), Barbara started with a basic **prompt** of "What are the common stuff in an online course?" She carefully **reviewed**



the AI's response in several utterances to call out the ideas that she liked, and then **ideated** based on the AI-generated features. As an example, she noted that "The progress tracking tool would be really good for a student to see how far along." Not sure what to ask next, she needed reminders from the interviewer about the task (the interviewer prompted: "how might you use those components for your redesign?"). Barbara went through a few utterances to **find approaches** for her next prompt and decided to focus on students as potential users.

**Turn 2. Requesting a list of features most helpful for students.** Barbara **prompted**, "What helps the students, I mean, what elements from above are most important to a student?" In response, the AI highlighted several features such as course materials, assignments, and instructor support. Barbara **reviewed** the response and remarked that "it is everything I actually said. That's pretty cool." She again took a moment to **find approaches** and asked the interviewer if she could get "more insight, kind of more in depth information."

**Turn 3. Requesting information about a specific design feature.** The interviewer encouraged Barbara to articulate the **prompt**, which she stated as "Tell me more in-depth information about progress tracking." Here, we observe that the participant narrowed down her design focus to a specific feature and used the tool to request information. In response, the AI provided several features for progress tracking, including completion tracking, grade tracking, and self-assessment. Upon **reviewing** the response, Barbara noted that "that's really cool. I wish we kinda had that on Canvas." She was particularly interested in student-facing features (e.g., time management tools, assignment tracking), and engaged in several utterances of **reviewing** and **reflecting** on them. She noted, "students have hard times with those [time management]. It's nice to see as a student, a good little tracking point." She continued pursuing these ideas in **ideation**, with features such as personal learning plans, to-do list, and milestones, and **reflected** on how these designs might be useful from a student's perspective.

Barbara spent her game seeking information in various forms. She started very broadly—*orienting to online courses*—and then focused on a specific group of users. For this group, she *investigated a list of features* and then *drilled down for more information about a specific feature*. As the game progressed, she became noticeably more comfortable with both the task and the AI player. We characterized her strategies as finding approaches to define the design space and orient to the AI tool. While she was able to incorporate the progress tracking idea in her final design sketch, her solution did not have as many details as John's, and instead just repeated the AI's feature descriptions.

## Discussion

This work adds two key features to the larger framework of a distributed game. These are the importance of *player orientation* and a way of measuring the *complexity of the game*.

### Orientation

Orientation to the AI showed up for both players (John & Barbara) in different ways over the course of the game. The participants oriented very differently to the AI player, with Barbara spending time building familiarity, while John started with prior experience. These different orientations changed the roles that the participants and AI played. For example, early in the game, John carefully *defined (the) role of the AI* and *checked for understanding*. He started with a very clear role and task and used these to guide the AI through the game. During the game, John maintained his role as the guide in a macro-game of role-play. In comparison, Barbara was not as familiar with the AI, and so needed to spend the time orienting herself to both the AI player and the task. She oriented to the AI as a *source of facts*, whereas John explicitly assigned a role and oriented to it as a *thinking partner*. We note *role definition* as a step which is potentially unique to a *distributed game*.

### Complexity of the game

John iterated quickly through turns within the macro-game, with three of his seven turns consisting of a quick **prompt-review** process. Additionally, John moved through six micro-games, (1) *role definition*, (2) *task definition*, (3) *requesting a mockup* (4) *listing (design) challenges*, (5) *compare/contrast*, and (6) *creating design solutions*. Both the quick iterations and the use of multiple different micro-games signified a more complex macro-game.

Meanwhile, Barbara started the game with much less familiarity & comfort, so she relied on the facilitator during the early stages of the game. As the game progressed, she became more familiar with the AI and focused her interactions there. Even with her increased familiarity, Barbara primarily engaged in only *list-making* micro-games, adding a *create design solutions* micro-game at the end as well. She also took more of the game time to **review, orient, and find approaches** to working with the AI player. Looking at her macro-game as a whole, it is clear that it took the additional time and effort for Barbara to become more familiar with the AI player. Because of the slower iterative process and the focus on a specific type of micro-game, Barbara played a less complex macro-game.

## Implications

Our findings about the different ways the participants oriented themselves to the design task and the AI, as well as the resulting complexity of the games that they played, have implications for educational practice. Specifically, we note the importance of additional support in task structure and facilitation to help students orient themselves to an AI-integrated, distributed game. We observe that guidance from the facilitator helped Barbara get to the point where she could play more directly with the AI. As educators find ways to use and evaluate AI in instructional settings, it is helpful to see how guidance from a facilitator or task design may strengthen students' orientation.

## Limitations and future work

There are some limitations with the current work. First, this study reported on short durations of interactions with the AI player, so we only observe the first stage of the design game (i.e., exploring the design space and ideating). Second, while we have conjectures about the role of orientation to the task from the cross-case analysis, we are not making definite claims. We encourage future work to investigate these conjectures further.

Additionally, future work with theoretical frameworks of distributed games, both *epistemic* and *design*, should include the AI (where present) as a player in the game. This inclusion should involve ways to describe AI *moves* within the game, as well as ways to describe the turns between human and AI players. This will allow the creation of descriptions for how AI players participate in distributed processes of both knowledge generation and design. Finally, researchers might evaluate whether the way players perceive/approach the AI can be guided by a teacher or trainer. It may be that this orientation is malleable and can be directed towards specific goals.

## Conclusion

In this paper, we expand on the framework of distributed games as a collective *knowledge construction* process (Matthews & Swanson, 2023; Matthews, Nguyen, & Swanson, 2023). Players within a distributed *epistemic* game interact to collectively *create* knowledge. We expand the framework by adding a distributed *design* game as collective *creation* of design. We compare two player's games side-by-side, to examine different orientations to the game and game complexity. These differences give us a sample of the range in characteristics that is possible within a distributed game. Additionally, the analysis included an *AI player* as part of design creation. Our work illuminates the role of AI in augmenting participants' moves. Acknowledging the role that AI has in design and the way other (human) players interact with it is a key step in building a framework of distributed, AI-integrated processes of creativity.

## References

- Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019, October). Beyond accuracy: The role of mental models in human-AI team performance. In Proceedings of the AAAI conference on human computation and crowdsourcing (Vol. 7, No. 1, pp. 2-11).
- Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis. Sage.
- Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational psychologist*, 28(1), 25-42.
- Daly, S. R., Adams, R. S., & Bodner, G. M. (2012). What does it mean to design? A qualitative investigation of design professionals' experiences. *Journal of Engineering Education*, 101(2), 187-219.
- Davis, N., Hsiao, C. P., Popova, Y., & Magerko, B. (2015). An enactive model of creativity for computational collaboration and co-creation. *Creativity in the digital age*, 109-133.
- Frederiksen, J. R., Sipusic, M., Sherin, M., & Wolfe, E. W. (1998). Video portfolio assessment: Creating a framework for viewing the functions of teaching. *Educational Assessment*, 5(4), 225-297.
- Gmeiner, F., Yang, H., Yao, L., Holstein, K., & Martelaro, N. (2023, April). Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-20).
- Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human-AI hybrid adaptivity in education. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I 21* (pp. 240-254). Springer International Publishing.
- Hutchins, E. (2008). The role of cultural practices in the emergence of modern human intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499), 2011–2019. <https://doi.org/10.1098/rstb.2008.0003>

- Matthews, J., Nguyen, H., & Swanson, H. (2023, March). Uncovering Features of Discourse that Increase Interactions. In Companion Proceedings 13th International Conference on Learning Analytics & Knowledge (LAK23). Learning Analytics & Knowledge (LAK23).
- Matthews, J. & Swanson, H. (2023, June). The Pivot: Identifying Emergent Tactics in Distributed Epistemic Games. Computer-Supported Collaborative Learning, Montreal.
- Nguyen, H., & Hayward, J. (2024). Reflective practices in designing with conversational artificial intelligence. AERA Annual Meeting 2024. Philadelphia, PA.
- Pintrich, Paul R. "A conceptual framework for assessing motivation and self-regulated learning in college students." *Educational psychology review* 16.4 (2004): 385-407.
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity research journal*, 24(1), 92-96.
- Schelble, B. G., Flathmann, C., McNeese, N. J., Freeman, G., & Mallick, R. (2022). Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1-29.
- Simeone, L., Mantelli, R., and Adamo, A. (2022) Pushing divergence and promoting convergence in a speculative design process: Considerations on the role of AI as a co-creation partner, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), *DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain*. <https://doi.org/10.21606/drs.2022.197>
- Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020, April). From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1-6).

## Developing Participation Equity Criteria: A Fresh Perspective via Turn-Taking Simulation

Liru Hu, University of Texas at El Paso, University of Hong Kong, [liruhu.hku@gmail.com](mailto:liruhu.hku@gmail.com)  
Jiajun Wu, University of Hong Kong, [wujiajun@connect.hku.hk](mailto:wujiajun@connect.hku.hk)

**Abstract:** Participation equity plays a crucial role in shaping students' cognitive and social benefits derived from collaborative efforts. Nevertheless, a lack of quantitative benchmarks for assessing the fairness of group participation persists. This study addresses this gap by employing a simulation-based approach to establish measurable criteria, leveraging empirical data to estimate model parameters. Our research introduces evaluation metrics for gauging participation and opportunity inequalities across varying numbers of turns, derived from the simulation outcomes. The practical significance of our findings lies in providing specific guidelines for evaluating participation equity in collaborative learning and problem-solving contexts. Additionally, from a theoretical standpoint, our study highlights the sensitivity of evaluating participation equity to the number of turns and offers insights into the dynamics of human dialogue. We also discuss limitations and avenues for future research in this area.

### Introduction

Scholars and educators widely acknowledge the effectiveness of learning through specific forms of dialogue in enhancing students' learning, comprehension, and skill development (e.g., Chen et al., 2020; Michaels et al., 2008; Resnick et al., 2015; Webb et al., 2014; Wegerif, 2020). However, the adoption of a dialogic educational approach encounters challenges related to addressing issues of social dominance and isolation within learning interactions (Clarke et al., 2016; Jin, 2017; O'Connor et al., 2017; Shah & Lewis, 2019). Unequal participation among students during collaborative activities can result in information loss, limitations in a team's problem-solving capabilities, and even the emergence of detrimental discourse that undermines group cohesion and hinders collective problem-solving and learning (Borge et al., 2018; Woolley et al., 2010). In virtual group settings, the absence of participation equity can also be a source of frustration (Solomon, 2016).

Moreover, Heinz and Rice (2009) contend that participation equity encapsulates qualities of openness and trust within a group, which are essential affective-motivational factors contributing to positive team outcomes. Furthermore, students' active articulation of their own ideas has been positively associated with their achievement outcomes (O'Connor et al., 2017). Consequently, participation inequality may result in disparities in individual achievement gains within collaborative learning experiences.

Therefore, efforts to cultivate balanced peer interaction have been undertaken through various means. Firstly, researchers have proposed several metrics to assess observed participation equality, often manifested as the degree of variability in the distribution of participation units, such as turns, comments, lines of text, or word counts (e.g., Haines et al., 2014; Reinig & Mejias, 2014). This variability is quantified using measures such as the Gini coefficient (Haines et al., 2014; Reinig & Mejias, 2014), standard deviation (Jahng et al., 2010; Kapur et al., 2008),  $I$  (Reinig & Mejias, 2014), an unbiased estimate of the Gini coefficient, and the coefficient of variation.

Secondly, researchers have incorporated social group awareness tools (Bodemer et al., 2018) and collaboration analytics tools (Martinez-Maldonado et al., 2021) to promote participation equality. These tools employ participation equality as a key metric to regulate participation behavior. For instance, Reflect, an interactive table, enables real-time shared visualization of member participation to encourage balanced engagement among group members (Bachour et al., 2010). Conversely, EQUIP serves as an observation tool designed to monitor the verbal participation of diverse demographic student groups in classroom settings (Reinholz & Shah, 2018). Similar tools include Second Messenger (DiMicco & Bender, 2007) and Conversation Clock (Karahalios & Bergstrom, 2009). Most of these initiatives focus on promoting equal participation among individuals or demographic groups (e.g., Janssen et al., 2007; Ollesch et al., 2021; Strauß & Rummel, 2023; Sukumar et al., 2020). They primarily provide graphical representations of interaction patterns rather than offering guidance for desired states or remedial actions (Hu & Chen, 2021). Social group awareness tools primarily present social information (Bodemer et al., 2018), but not all groups effectively leverage this information (Dehler et al., 2009), and the benefits of such tools rely on students actively engaging with the feedback provided (Lipnevich & Panadero, 2021). However, it can be difficult, particularly for younger students, to accurately assess their level of participation equality based on individual participation rates or statistical measures of participation equality.

One possible reason for the limited provision of advanced behavior scaffolds beyond graphical representation is the challenge of determining when to prompt intervention. While Strauß and Rummel (2021)

introduce adaptive prompts alongside group awareness tools to promote regulation of equal participation, it remains unclear how they decide when to prompt students to discuss responsibility allocation based on their metric of participation inequality. To the best of our knowledge, there are still no established quantitative reference criteria for evaluating whether a group's participation is considered equitable. For example, what level of standard deviation in individual participation rates should be considered indicative of imbalance? Defining which threshold serves as an acceptable criterion to guide student participation or inform the design of social awareness tools remains an open question.

Another limitation in the current literature is the predominant focus on promoting equal participation among individuals or demographic groups, often overlooking other essential dimensions of equitable peer interaction, such as participation opportunities. Shah and Lewis (2019) define participation equity as “a condition where opportunities to participate—and participation itself—are fairly distributed among all students involved in a learning interaction” (p. 428). This definition encompasses two dimensions of participation equity: participation itself and opportunities to participate. In line with this framework, this study employs two quantitative measures to assess a group's level of participation equity. The first measure is participation inequality, quantified by the standard deviation of individual participation rates (Jahng et al., 2010; Kapur et al., 2008). The second measure is opportunity inequality, assessed by the standard deviation of individual opportunity rates. In this context, a participation opportunity is defined as a turn directed towards someone. For instance, in a triadic group consisting of A, B, and C, if A addresses B, it is assumed that A has granted the conversational floor to B, thereby giving B one participation opportunity. It is also common for A to address the entire group rather than specifically targeting B or C, in which case both B and C would each receive one participation opportunity when responding.

In summary, the primary objective of this study is to establish robust criteria for evaluating both participation and opportunity inequality by simulating the turn-taking dynamics in collaborative environments.

## Factors influencing participation equity

Group size is an important factor on participation equity. For larger groups such as online communities or face-to-face group discussions of considerable size, it is quite common for people to participate very differently. Nielsen (2006) observes a recurring phenomenon in online participation, often following a 90-9-1 pattern, where 90% of users are passive observers (lurkers), 9% contribute intermittently, and a mere 1% account for most contributions. This pattern, akin to a long tail or power-law distribution, has been observed across diverse community contexts (Haklay, 2016; Z. Wang et al., 2018; Yuan & Recker, 2014; Konieczny, 2016), although it may deviate from the 90-9-1 ratio in certain communities (e.g., Bonafini, 2018). In face-to-face settings, an earlier study also revealed that individuals' participation rates in large-sized group discussions exhibit an exponential decline, indicative of a power-law distribution that underscores substantial participation inequality. In contrast, smaller groups offer a more conducive environment for achieving participation equality. For instance, in dyads (groups of two), equal participation is readily attainable through turn-taking. However, in relatively larger groups, such as those comprising three or four members, the likelihood of dominance or isolation by one or a few individuals may increase. This study focuses specifically on small-sized groups, where there is a greater emphasis on the effectiveness of collaboration and the issue of participation equity, in contrast to larger-sized groups.

Students' engagement in collaborative activities is closely associated with their individual characteristics. Previous research has highlighted the influence of students' status within a group on their participation dynamics. Cohen and Lotan (2014) have identified several indicators of status, including academic status (referring to one's academic performance ranking within the group), peer status (indicative of attractiveness or popularity among peers resulting from interactions within and outside the school context), and social status (which encompasses distinctions related to social class, gender, race, and ethnic background). Information about students' status can be collected through demographic surveys that capture parental educational backgrounds, gender, ethnicity, or by soliciting students' self-reported levels of friendship with fellow group members. Alternatively, academic status can be inferred from recent academic scores. Furthermore, students' levels of interest, comprehension, and self-confidence play a significant role in shaping their participation behaviors (Hu & Chen, 2022; Blue et al., 1998). Students are more likely to engage actively when they possess the ability to articulate their thoughts, possess subject knowledge, and do not fear appearing unintelligent in front of peers or instructors (Jin, 2012). Gathering this information can be achieved through the utilization of various psychological scales, which assess students' self-concept, enjoyment, personality traits, and levels of social anxiety.

Various types of tasks may also result in different levels of participation equity among the same group of students. Research has shown that ill-structured problems tend to stimulate more problem-centered discussions and higher participation inequality compared to well-structured problems (Kapur & Kinzer, 2007). Some scholars recommend the use of open-ended questions that encompass a wide range of student competencies to prevent the

dominance of high-achieving students (Boaler, 2008; Cohen & Lotan, 1995), or the implementation of jigsaw activities to position students as experts in various aspects (Klein, 2018).

## Methods

### Setup of the model

The simulated turn-taking model classifies all personal factors influencing participation equity into two categories: Type I and Type II factors. Type I factors pertain to elements that impact an individual's possibility to be the speaker. For instance, one's self-concept primarily influences their confidence in their academic abilities. Students with lower levels of self-concept may hesitate to express their ideas or comment on others' contributions due to insecurity about their own ideas and a fear of making errors. Social anxiety is another characteristic that falls within the Type I factors. Students with high levels of social anxiety tend to exhibit passivity in collaborative settings due to discomfort associated with public speaking.

Type II factors, on the other hand, influence an individual's likelihood of being the addressee. An example of a Type II factor is academic status, which exerts an influence on the likelihood of others initiating communication with an individual. In collaborative settings, students tend to be more inclined to engage with those they perceive as academically proficient, particularly when confronted with challenging tasks or when prioritizing aspects such as problem-solving speed and desired outcomes. Conversely, students who are perceived as academically disadvantaged often assume follower roles with limited participation opportunities. Additionally, students may have varying degrees of friendship with one another, which can influence their chances of being addressed. For instance, students are more likely to receive participation opportunities if they share a closer relationship with the speaker compared to other group members. This phenomenon may be attributed to the sense of psychological safety associated with conversing with friends and reflects the principle of "reciprocity" in dialogue, wherein people tend to engage more with those who have interacted with them extensively in the past (Leenders et al., 2016).

In addition to individual characteristics, our simulation model incorporates contextual and temporal aspects of turn-taking. We distinguish between three types of reciprocal patterns, as outlined in Table 1. One such pattern is termed "turn-receiving" (Hu & Chen, 2022; Gibson, 2003, 2005). In this pattern, when individual B is addressed by individual A in the previous turn, it is highly likely that B will take the conversational floor and continue speaking, as supported by the organizational principles governing human turn-taking (e.g., Sacks et al., 1974; Schegloff, 2007). Therefore, our model includes a consideration of the rate at which the previous addressee becomes the subsequent speaker, denoted as " $P_{\text{receive}}$ ." This consideration complements the personalized Type I factors that influence an individual's likelihood of assuming the role of the speaker.

**Table 1**  
*Temporal Factors Influencing Turn-Taking Patterns.*

Pattern	Rate	Representation	Interpretation
Turn-receiving	$P_{\text{receive}}$	A-> (B) -> <b>B</b> -> X	A talks to B, B then talks to X (could be A)
Talk-back	$P_{\text{talkback}}$	A -> (B) -> X -> ( <b>A</b> )	A talks to B, then X (could be B) talks to A
Talk-to-group	$P_{\text{talktogroup}}$	A -> (Group) -> X	A talks to the group without nominating the next speaker

The rate at which individuals respond to the last speaker is another significant feature that influences turn-taking patterns. Human interactions inherently possess a reciprocal nature (Blau, 1964; Gergen et al., 1980), with feedback being a fundamental component of conversations. Consequently, when individual A serves as the last speaker, it is highly probable that A will become the addressee in the next turn. To account for this, our model incorporates a consideration of the rate at which the last speaker assumes the role of the new addressee, denoted as " $P_{\text{talkback}}$ ." This addition complements the personalized Type II factors when determining the addressee.

Additionally, there are instances where the current speaker does not specifically nominate the next speaker in a group discussion, essentially relinquishing the conversational floor to the entire group. This approach ensures that all other group members have equal participation opportunities. In our study, we regard this rate, denoted as " $P_{\text{talktogroup}}$ ", as an intrinsic feature of dialogue and will estimate its values based on empirical data.

Furthermore, prior research has proposed the concept of general reciprocity, which relates to the likelihood of speaker A yielding the conversational floor to speaker B based on the total number of times B has yielded the floor to A in the past (Leenders et al., 2016). We consider this feature to be reflective of personal characteristics rather than an inherent aspect of human dialogue. It shares similarities with individuals' degrees of friendship, which can be quantified based on cumulative interactions over time. Therefore, we categorize general reciprocity as a Type II factor rather than an intrinsic feature of dialogue dynamics.

In summary, our turn-taking simulation in this study focuses on three rates related to features intrinsic to human dialogue. “ $P_{\text{receive}}$ ” influences the selection of speakers, “ $P_{\text{talkback}}$ ” and “ $P_{\text{talktogroup}}$ ” influence the choice of addressees. The decision-making process for turn-taking in our model is presented in Table 2.

**Table 2**  
*Turn-Taking Simulation Procedure and Rules.*

Steps	Turn	Speaker	Addressee	Criteria of choice
1	1	Speaker <sub>1</sub>	→	Type I factors
2	1		Addressee <sub>1</sub> →	Type II factors
3	2	Speaker <sub>2</sub>	→	$P_{\text{receive}}$ and Type I factors
4	2		Addressee <sub>2</sub> →	$P_{\text{talktogroup}}$ , $P_{\text{talkback}}$ and Type II factors
...	...			
2n-1	n	Speaker <sub>n</sub>	→	$P_{\text{receive}}$ and Type I factors
2n	n		Addressee <sub>n</sub> →	$P_{\text{talktogroup}}$ , $P_{\text{talkback}}$ and Type II factors

In this study, our focus is on triadic groups, and we aim to simulate ideally equitable group discussions. To achieve this, we exclusively consider dialogue-related features ( $P_{\text{receive}}$ ,  $P_{\text{talktogroup}}$ , and  $P_{\text{talkback}}$ ) in our turn-taking simulations and do not incorporate personalized factors (Type I and Type II). The detailed simulation process for an ideally equitable triadic group discussion unfolds as follows:

Determine if it is the first turn?

If yes, randomly select one member as the first speaker.

Randomly designate either the entire group or another member as the addressee.

If not, the previous addressee has a  $P_{\text{receive}}$  rate to become the current speaker. Note that the previous speaker cannot assume the role of the speaker again in the next turn. The remaining member has a rate of  $(1 - P_{\text{receive}})$  to become the current speaker.

Determine the addressee for the current turn based on rates: addressing the entire group ( $P_{\text{talktogroup}}$ ), addressing the previous speaker ( $P_{\text{talkback}}$ ), addressing the remaining member  $(1 - P_{\text{talktogroup}} - P_{\text{talkback}})$ .

## Parameter estimation from empirical data

This study focuses on three parameters, namely  $P_{\text{receive}}$ ,  $P_{\text{talkback}}$  and  $P_{\text{talktogroup}}$ , which are reflective of inherent characteristics of human interaction. To appropriately set these parameters, we began by examining their distributions within empirical data.

Our dataset comprises 117 fourth-grade primary school students from two classes in the mainland of China. We collected recent academic grades in mathematics and Chinese from their respective teachers. Additionally, students were asked to nominate three classmates they preferred to be grouped with and three peers they frequently consulted for mathematics-related questions. To ensure comparability across groups, we categorized students within each class into three levels based on their overall scores in mathematics, Chinese, and individual reasoning ability. Subsequently, we formed groups of three students, with one student from each level. To encourage participation among low-level students, we prioritized their grouping preferences when selecting middle- and high-level group members. Furthermore, we aimed to ensure that each group included at least one male and one female student. Group assignments were adjusted based on teacher recommendations and significant feedback from students regarding interpersonal relationships within the groups.

Students were tasked with solving three mathematical problems and one reading comprehension problem. Following these tasks, they were regrouped under similar rules and assigned two additional mathematical problems and one more Chinese problem. The mathematical problems were process-open and varied in terms of difficulty and complexity. For instance, one task involved predicting the shape of a knot when its two ends were pulled taut, while another required students to create unique patterns by combining six identical triangular tiles. The reading tasks involved sequencing scrambled sentences to construct paragraphs for two different stories. Teachers emphasized the importance of mutual respect, active and equitable participation, and joint efforts. Students were informed that their performance would be assessed not only based on their solutions but also on their participation and communication during the problem-solving process.

To evaluate the three parameters in triadic group discussions, we calculated them at the individual level. Specifically, individual  $P_{\text{receive}}$  quantifies how frequently someone takes the assigned floor from the previous speaker relative to their frequency of being the addressee. Individual  $P_{\text{talktogroup}}$  measures how frequently someone addresses the entire group compared to their total speaking turns. Individual  $P_{\text{talkback}}$  quantifies how often someone addresses the previous speaker relative to their total speaking turns. These individual-level rates may be adjusted based on individual characteristics. For example, students with lower academic grades or higher social anxiety

may have a lower rate of seizing given participation opportunities, resulting in a lower individual  $P_{\text{receive}}$  compared to their more capable peers.

To estimate the three parameters inherent to the dialogue itself, we excluded individual rates if their group had a member who contributed fewer than 10 turns. This exclusion criterion was applied because individual rates may exhibit significant fluctuations if a student participated very minimally, and the group dynamics might resemble a dyadic interaction if one member spoke extremely infrequently during problem-solving tasks. Subsequently, we calculated the means of individual rates by averaging them across various types of tasks and different group compositions. Finally, we averaged all rates from all individuals to derive estimates of the parameters. This approach aimed to mitigate the effects of task variations and subjective factors (see Table 3).

**Table 3**  
*Descriptive Statistics and Distributions of Three Parameters*

Parameter	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max	95% <i>CI</i> of <i>M</i>	Normality
$P_{\text{receive}}$	115	0.609	0.109	0.36	0.88	[0.589, 0.629]	$p = .799$
$P_{\text{talkback}}$	115	0.451	0.099	0.22	0.68	[0.433, 0.469]	$p = .445$
$P_{\text{talktogroup}}$	117	0.446	0.119	0.15	0.75	[0.424, 0.468]	$p = .225$

The results indicate that the distributions closely approximate normal distributions. Specifically, the  $p$ -values obtained from the Shapiro-Wilk tests suggest that there are no significant deviations from normality in the distributions of all three average individual rates. The 95% confidence interval for the mean values demonstrates fluctuations within a very narrow range. In our simulation model, we employ the means of these three parameters to replicate ideally equitable triadic group discussions.

## The simulation model

In our research, we are primarily concerned with simulating the dynamics of ideally equitable turn-taking within triadic groups, specifically focusing on the inherent characteristics of human dialogue while excluding the influence of personalized factors. To establish criteria for equitable triadic group discussions, we conducted 1000 simulations across various scenarios, considering common numbers of turns typically observed in classroom group activities, including 30, 40, 50, 100, 150, 200, 300, 400, 500, and 600 turns. From these simulations, we derived thresholds for participation inequality and opportunity inequality by selecting the 95<sup>th</sup> percentile data points, which serve as practical guidelines.

In addition to defining these criteria, we sought to evaluate the reliability of our simulation models. To do so, we recalculated the three key parameters using the simulated data to assess their consistency with the original input parameters. Furthermore, we conducted sensitivity analyses to explore how variations in these parameters within their 95% confidence intervals might impact the simulation results.

## Results

Our findings reveal that as the number of turns increases, both participation inequality and opportunity inequality tend to approach zero (see Table 4). Notably, there is a substantial degree of fluctuation in both of these indicators during the initial stages of the discussion, typically within the first 50 turns. However, as the number of turns progresses, these indicators exhibit slower rates of change (see Figure 1a). Importantly, the distributions of these indicators are not normal; they exhibit left-skewness. For instance, Figure 1b illustrates the distribution of participation inequality for a scenario involving 200 turns.

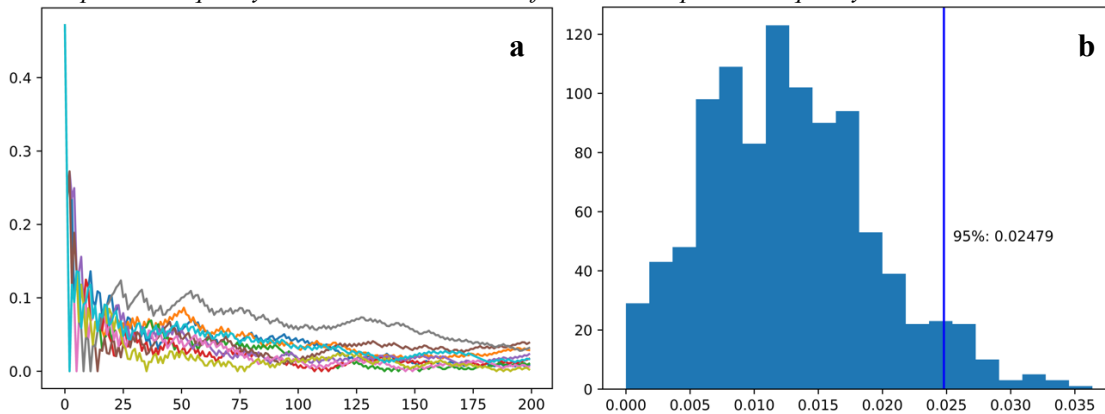
**Table 4**  
*Simulation Results for 1000 Times*

Turn	<i>n</i>	Participation inequality				Opportunity inequality			
		<i>M</i>	<i>SD</i>	95%th	Normality	<i>M</i>	<i>SD</i>	95%th	Normality
30	1000	.047	.026	.094	< .001	.033	.017	.063	< .001
40	1000	.042	.021	.077	< .001	.029	.015	.057	< .001
50	1000	.037	.019	.068	< .001	.026	.014	.049	< .001
100	1000	.026	.013	.050	< .001	.018	.009	.035	< .001
150	1000	.021	.012	.043	< .001	.015	.008	.028	< .001
200	1000	.018	.009	.035	< .001	.013	.006	.025	< .001
300	1000	.015	.008	.029	< .001	.010	.005	.019	< .001
400	1000	.013	.007	.026	< .001	.009	.005	.018	< .001
500	1000	.012	.006	.023	< .001	.008	.004	.016	< .001
600	1000	.011	.006	.022	< .001	.007	.004	.014	< .001



**Figure 1**

Participation Inequality for 200 Turns and 1000 Simulation Times. a. Temporal Evolution of First 10 Participation Inequality Scores. b. Distribution of 1000 Participation Inequality Scores.



Using the quantitative criteria established in Table 4 as a benchmark, we assessed both participation inequality and opportunity inequality within our empirical dataset (Table 5). The results indicate that roughly 80% of the groups exhibited varying degrees of participation inequality issues, while approximately 60% of the groups displayed issues related to opportunity inequality.

**Table 5**

Evaluation of Participation Inequality and Opportunity Inequality Based on the Established Criteria.

Turn	n	Participation inequality			Opportunity inequality		
		M	SD	Number of qualified groups (%)	M	SD	Number of qualified groups (%)
1	44	0.156	0.078	7(16%)	0.091	0.070	18(41%)
2	31	0.119	0.060	8(26%)	0.061	0.044	20(65%)
3	35	0.119	0.072	10(29%)	0.063	0.038	13(37%)
4	105	0.105	0.072	25(24%)	0.061	0.055	37(35%)
5	35	0.086	0.063	8(23%)	0.051	0.046	13(37%)
6	9	0.084	0.038	1(11%)	0.087	0.035	0(0%)
7	4	0.095	0.083	2(50%)	0.072	0.061	1(25%)
Total	263	0.114	0.073	61(23%)	0.066	0.054	104(40%)

We recalculated the three parameters using simulated data to assess the model's accuracy. The results demonstrate that the recalculated parameter means align closely with the input values (Table 6). Interestingly, as the number of turns increased, the distributions of these parameters tended toward normality.

**Table 6**

Descriptive Statistics of Three Parameters Based on Simulated Data.

Turn	n	P <sub>receive</sub>			P <sub>talkback</sub>			P <sub>talktogroup</sub>		
		M	SD	Normality	M	SD	Normality	M	SD	Normality
30	1000	.604	.122	.017	.454	.093	< .001	.439	.092	< .001
40	1000	.613	.107	.234	.447	.081	< .001	.445	.079	< .001
50	1000	.608	.094	.008	.453	.069	< .001	.443	.069	< .001
100	1000	.610	.067	.398	.452	.049	.003	.445	.048	.004
150	1000	.610	.052	.051	.450	.041	.021	.446	.041	.035
200	1000	.612	.046	.006	.453	.036	.102	.444	.036	.011
300	1000	.609	.039	.550	.452	.029	.077	.445	.028	.072
400	1000	.610	.031	.407	.451	.026	.013	.445	.025	.081
500	1000	.610	.030	.760	.451	.022	.302	.446	.022	.327
600	1000	.007	.004	.329	.450	.020	.498	.447	.020	.684

Upon altering the input parameters to their respective limits within the 95% confidence intervals, we observed that the variations in the 95<sup>th</sup> percentiles of participation inequality and opportunity inequality were not

highly sensitive (see Table 7). These values remained relatively stable, with minimal changes in the second digit after the decimal point and only slight adjustments to the third digit.

**Table 7**  
*Sensitivity Analysis for 200 Turns and 1000 Simulations.*

Output	[ $P_{receive}$ , $P_{talkback}$ , $P_{talktogroup}$ ]						
	[.609, .451, .446]	[.589, .451, .446]	[.609, .451, .424]	[.609, .433, .446]	[.609, .469, .446]	[.609, .451, .468]	[.629, .451, .446]
95%th of Participation inequality	<b>.0379</b>	.0368	.0352	.0358	.0368	.0366	.0371
95%th of Opportunity inequality	<b>.0255</b>	.0245	.0256	.0253	.0247	.0244	.0247

## Discussion

This study introduces quantitative criteria for assessing participation inequality and opportunity inequality within triadic group discussions using a simulation-based approach. Our findings indicate that both participation inequality and opportunity inequality tend to approach zero as the number of turns increases. However, achieving perfect equity in limited school group discussion time is challenging. Therefore, when evaluating whether students have similar numbers of speaking turns or receive comparable participation opportunities, it is essential to consider the total number of turns as a contextual factor influencing equity in group discussion. Disparities in speaking turns and received opportunities can be more pronounced in discussions with a limited number of turns. Consequently, we should exercise caution when assessing participation equity in group discussions with a limited number of turns, as the standard for ideally equitable group discussions should be less strict for groups with fewer turns compared to those with a greater number of turns. This insight offers valuable guidance for designing advanced scaffolding in social group awareness tools, which should employ less stringent standards when providing early-stage prompts related to participation equity in group discussions due to the substantial fluctuations observed in both participation inequality and opportunity inequality in the initial phases of discussion.

Our findings contribute to a deeper understanding of the early lock-in phenomenon observed in participation equality within both online small group collaboration (Kapur et al., 2008) and face-to-face small group discussion (Lämsä et al., 2018). These prior studies visually depict cumulative individual participation rates, revealing that participation inequality, as measured by the standard deviation of individual participation rates, tends to stabilize in the early stages of discussion. In our simulated ideal group discussions, we observed similar patterns, with both indicators exhibiting significant fluctuations in the initial phase with a limited number of turns. Moreover, our quantitative analysis reveals that, following initial fluctuations during the first 50 or so turns, there is a gradually diminishing trend towards achieving perfect equity in group discussions.

The quantitative criteria established in this study are expected to serve as valuable tools for supporting the design of advanced scaffolds in social group awareness tools (Bodemer et al., 2018) and visual learning analytical tools (Hu & Chen, 2021). These criteria can assist designers in determining when to implement appropriate scaffolds for groups. Furthermore, given the decreasing rate of change in participation inequality and opportunity inequality, we recommend that technology designs implement real-time monitoring of participation equity using a sliding window strategy of approximately 30 to 50 turns. This approach can enhance the detection and timely intervention for potential participation inequity issues within group discussions.

The proposed criteria also offer valuable insights for both students and teachers, fostering a better understanding of equal participation and equal participation opportunities. Students, especially young students, tend to perceive themselves as not participating equally if their speaking turns are not identical. Our criteria indicate that demonstrating equality in group discussions with limited time is a challenging task. In practice, ideal participation equality and opportunity equality do not necessitate absolute uniformity but should fall within reasonable ranges. With reference to these criteria, students can gain a better understanding of their performance regarding participation equity in group discussions. Moreover, the quantitative evaluation criteria offer teachers a valuable tool to assess the degree of inequity in group discussions. This resource can aid teachers in delivering timely and effective scaffolding interventions to support groups more efficiently.

Furthermore, beyond the primary findings concerning the quantitative criteria, it is noteworthy that the distributions of the three rates ( $P_{receive}$ ,  $P_{talkback}$ ,  $P_{talktogroup}$ ) in triadic groups exhibited no substantial deviations from normal distributions in empirical data. Our subsequent recalculations of these three parameters based on simulated data, utilizing constant parameters, revealed distributions that closely approximated normal distributions as the

number of turns increased. This observation provides additional evidence supporting our hypothesis that these three parameters are indicative of the fundamental characteristics of human dialogue.

## Limitations and future research

This study acknowledges several limitations that warrant attention in future research. Firstly, the parameters of the simulation model are derived from empirical data involving fourth-grade primary school students solving a limited range of tasks. It is imperative to validate these parameters across various contexts and with a broader spectrum of task types to enhance the generalizability of the findings. Secondly, the study exclusively focuses on triadic groups, while group size is a pivotal factor influencing participation equity. Future research endeavors should aim to establish quantitative evaluation criteria for larger group sizes, such as four, five, and six-person groups. However, it is important to note that expanding the analysis to larger groups may introduce greater complexity to the simulation model and necessitate additional empirical data for parameter estimation. Lastly, this study exclusively simulates conditions for ideally equitable group discussions. Future research will consider the incorporation of type I and type II personal factors to establish more nuanced quantitative criteria capable of detecting various degrees of participation inequality and opportunity inequality. This includes categories like slight, moderate, and strong inequality, thereby offering more specific guidance for practical applications and the design of adaptive feedback within visual learning analytical tools. Moreover, these quantitative evaluation criteria, derived from simulations, should undergo further validation through qualitative research methods.

## References

- Boaler, J. (2008). Promoting “relational equity” and high mathematics achievement through an innovative mixed-ability approach. *British Educational Research Journal*, 34(2), 167–194.
- Bonafini, F. C. (2018). Characterizing super-posters in a MOOC for teachers’ professional development. *Online Learning Journal*, 22(4), 89–108. <https://doi.org/10.24059/olj.v22i4.1503>
- Hu, L. & Chen, G. (2021). A systematic review of visual representations for analyzing collaborative discourse. *Educational Research Review*, 34, 100403. <https://doi.org/10.1016/j.edurev.2021.100403>
- Hu, L. & Chen, G. (2022). Exploring turn-taking patterns during dialogic collaborative problem solving. *Instructional Science*, 50(1), 63–88. <https://doi.org/10.1007/s11251-021-09565-2>
- Jahng, N., Nielsen, W. S., & Chan, E. K. H. (2010). Collaborative learning in an online course: A comparison of communication patterns in small and whole group activities. *Journal of Distance Education*, 24(2), 39–58.
- Kapur, M., Voiklis, J., & Kinzer, C. K. (2008). Sensitivities to early exchange in synchronous computer-supported collaborative learning (CSCL) groups. *Computers & Education*, 51(1), 54–66. <https://doi.org/10.1016/j.compedu.2007.04.007>
- Lämsä, J., Hämäläinen, R., Koskinen, P., & Viiri, J. (2018). Visualising the temporal aspects of collaborative inquiry-based learning processes in technology-enhanced physics learning. *International Journal of science education*, 40(14), 1697-1717.
- Leenders, R. T. A., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review*, 6(1), 92-115.
- Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education*, 6, 720195. <https://doi.org/10.3389/feduc.2021.720195>
- Ollesch, L., Heimbuch, S., & Bodemer, D. (2021). Improving learning and writing outcomes: Influence of cognitive and behavioral group awareness tools in wikis. *International Journal of Computer-Supported Collaborative Learning*, 16(2), 225–259. <https://doi.org/10.1007/s11412-021-09346-6>
- Shah, N., & Lewis, C. M. (2019). Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework. *Cognition and Instruction*, 37(4), 423–452. <https://doi.org/10.1080/07370008.2019.1631825>
- Strauß, S., & Rummel, N. (2021). Promoting regulation of equal participation in online collaboration by combining a group awareness tool and adaptive prompts. But does it even matter? *International Journal of Computer-Supported Collaborative Learning*, 16(1), 67–104.
- Sukumar, P. T., Reinholz, D., & Striegel, A. (2020). *Visualizing participatory inequities in classroom data. In proceedings of the annual IEEE visualization conference (VIS 2020)*. Salt Lake City, UT, USA.
- Wegerif, R. (2020). Towards a dialogic theory of education for the internet age. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 14–26). Routledge.

# A Knowledge Building-Modeling Approach to Scientific Inquiry

Dina Soliman, University of Toronto, dinaa.soliman@utoronto.ca  
Xueqi Feng, Southern University of Science and Technology, SUSTech Affiliated Experimental School,  
fengxueqi@hotmail.com

**Abstract:** This study explores how a Knowledge Building - Modeling approach (KBM) was used to enhance grade 5 students' understanding of dynamic mechanisms related to natural hazards. Forty-three students collectively engaged in an iterative KBM process involving idea-driven discourse on Knowledge Forum. Additionally, students created models in response to emerging theories and questions. Iterations were driven by students' discourse and supported by learning analytics to advance idea and model building. Results show that students engaged in deeper discourse practices over time. Their models increasingly reflected complex causal reasoning and their knowledge of natural hazards improved. Analysis revealed that knowledge building discourse was a good predictor of both scientific understanding and modeling practices. We discuss the work that informed our intervention and highlight the KBM approach. Implications of designing knowledge building environments enriched with modeling to promote complex reasoning and modeling practices are discussed.

## Introduction

The world is composed of complex, interconnected parts whose behaviors reinforce or balance each other in dynamic and often non-intuitive patterns. Understanding a complex phenomenon requires viewing it as a system consisting of multiple interdependent elements whose relationships are usually not obvious (Hmelo-Silver & Pfeffer, 2004). For example, earthquakes constitute a complex, interconnected system involving various dynamic mechanisms and consequences. Tectonic plate movements trigger seismic events as stress accumulates along fault lines, releasing energy as seismic waves, which can then trigger cascading events like mudslides and tsunamis (Stern, 2007). The environmental effects can involve altered landscapes, changes in ecosystems, and release of harmful gases. Hence, to making sense of complex system a person needs to mentally form interconnections between different aspects of a phenomena (Penner, 2000). Wilensky (2003) argues that bridging the gap between the existing, static curricula and the complexities that students experience in the real world necessitates a new form of literacy that engages students in exploring dynamics of complex systems. One notable challenge for students to reason with complex systems is their non-linearity. Students tend to think of complex problems in a reductive way, assuming a linear, sequential cause-and-effect (Assaraf & Orion, 2005; Hmelo-Silver and Pfeffer, 2004; Jacobson & Wilensky, 2006). In contrast, complex systems are characterized by non-linearity, involving feedback loops and changes over time (Jacobson et al., 2016; Meadows, 2008). Pedagogies that can facilitate the conceptual shift towards non-linearity are therefore required if students are to reason with complex systems. Hmelo-Silver et al. (2000) suggest that students should engage in modeling activities as well as discussions that help them understand casual behavior.

## Complex system understanding via knowledge building and modeling (KBM)

### Knowledge building

Knowledge Building is an educational approach which engages students in collaborative problem solving aimed at improving ideas and advancing community knowledge (Scardamalia, 2002). In Knowledge building communities, students assume epistemic agency, setting their own goals and identifying next steps. Knowledge Building prepares students to tackle complex problems by emphasizing working with ideas in design mode – theorizing, synthesizing, and identifying new areas to be explored (Bereiter & Scardamalia, 2014). It fosters idea diversity by encouraging students to engage in non-linear thinking, exploring multiple idea trajectories, connections, and possibilities concurrently. Students take collective responsibility for synthesizing ideas, thereby developing their capacity to navigate complexities inherent in real-world problems.

Knowledge Forum (KF), the technology designed to facilitate Knowledge Building (Scardamalia, 2004), includes a suite of analytics to help community members assess their discourse patterns and knowledge growth. In this study, we utilize a social-semantic platform called Knowledge Building Discourse Explorer (KBDeX) (Feng et al., 2021; Oshima, 2012) to generate a word-network visualization that highlights relationships among ideas in the community. This visualization offer students an accessible means to conceptualize discourse by

mapping collective ideas and identifying idea boundaries, thereby supporting them in navigating complex systems and serving as a grounding for classroom discussion around ideas and their implications.

## Modeling and scientific inquiry

Researchers view modeling as the practice of purposefully developing a representation of one's conception of a reality - the properties, processes, relationships, behaviors, and conditions that surround a system (Harrison & Treagust, 2000). These models are used to describe, communicate, or investigate complex phenomena. Prior research shows that modeling enables students to innovate and engage in sustained inquiry and evaluation, rather than focus on correct answers to problems posed by the teacher (Lehrer and Schauble, 2005). Specifically, engaging students in creating models for explaining mechanisms could support their epistemic agency by empowering them to construct their own knowledge as they examine complex phenomena. In our research, we adopt Knuuttila's (2011) idea of models as knowledge-creating entities that can be used as objects of discourse that can be improved as students engage in collaborative discourse.

This study explores how Knowledge Building supported by learning analytics and enriched with modeling practices can facilitate the understanding of dynamic mechanisms in a grade 5 science unit. We adopt an iterative Knowledge Building – Modeling (KBM) approach, which starts with students generating questions and theories about phenomena. Students then develop models that represent ideas and theories and use their models to generate more questions and ideas, which become part of the community's discourse that students can advance and explore in subsequent model iterations. We anticipate that visualizations produced by learning analytics could help students identify emerging ideas and relationships as they engage with more in-depth discourse about the natural hazards, which they could then integrate in their models. The expectation is for students to understand dynamic mechanisms associated with plate tectonics, along with an understanding of the environmental, geological, and humanitarian impacts.

In this study, we investigate how students engage in deeper causal reasoning (Perkins & Grotzer, 2005) – moving beyond simple cause-and effect towards more dynamic mechanisms – as they engage in iterative KBM practices supported by learning analytics. We aim to answer the following questions: 1) Following the KBM approach, did students' models, KF discourse, and domain understanding develop over the course of the study? 2) What were the predictors of modeling ability and domain understanding? 3) How did students' models develop through knowledge building practices?

## Methods

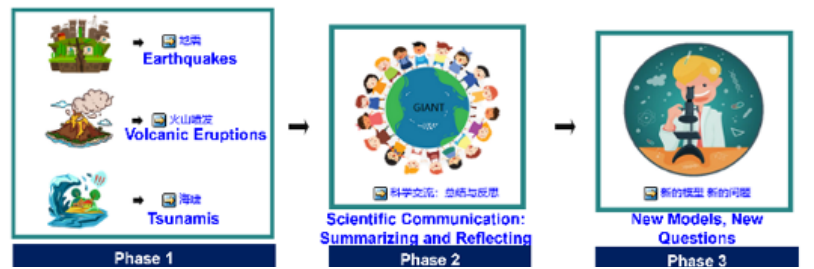
### Participants and context

Participants include grade 5 students ( $n=43$ ) in Shenzhen, China, exploring an earth science unit focused on natural hazards. For 7, 45-minute sessions, students generated questions, ideas and models related to earthquakes, volcanoes, and tsunamis. They actively participated in knowledge building discourse both in class and on KF. Individually, students drew paper-and-pencil models to illustrate their understanding of elements, relationships, and mechanisms stemming from their discourse. Models were also used to stimulate further ideas and questions, which served as objects for ongoing discourse.

### The KBM environment design

Students engaged in KBM practices over three phases (see boxes in figure 1). The first model was part of their domain pre-test, which was used to elicit students' prior knowledge. Subsequent model iterations were driven by students' collective discourse, where they generated new questions and theories concerning the three phenomena. Students created models on prompt sheets with scaffolds such as "I want to modify my original model like this because", and "By constructing this scientific model, I have updated ideas/questions" designed to facilitate model-based reasoning (Lehrer & Schauble, 2005) and epistemic modeling practices (Schwarz et al., 2009). New questions were later moved to KF to facilitate collective inquiry.

**Figure 1**  
*Home Page of the Knowledge Forum community Showing Three Phases (partial view)*



**Phase 1: Initial ideas and idea synthesis (Sessions 1-3)**

The unit started by stimulating students' thinking by showing them images and videos of earthquakes, tsunamis, and volcanoes. Students then engaged in collective inquiry on KF by posing their questions and preliminary ideas in three different KF views, one for each phenomenon. Students also created notes which synthesized their collective ideas, after which they constructed their second model. To help facilitate model construction, students were introduced to scientific modeling by showing them pictures of models they are familiar with (for e.g., cell wall model) and engaging them in discussions around the roles of models in explaining theories and discoveries.

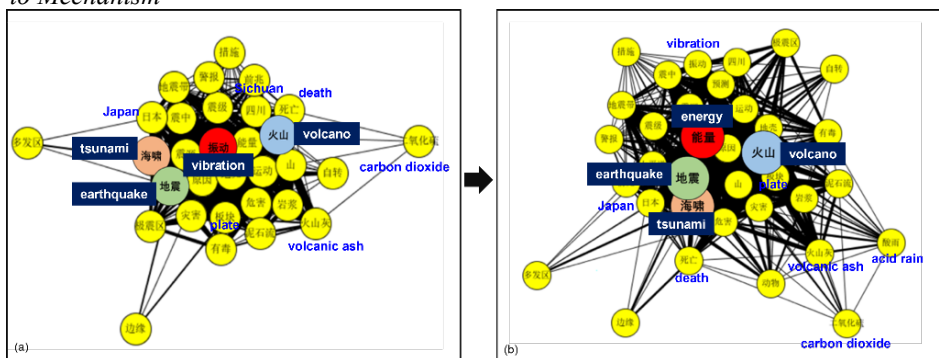
**Phase 2: Modeling and Inquiry through learning analytics and meta-discourse (Session 4-5)**

In phase 2, students deepened their inquiry through identifying categories of questions and ideas around the three phenomena, posting their notes in a view called “Scientific Exchange: Summary and Reflection”. Students also collectively agreed on some criteria of good models, such as that models should explain/predict issues and be clear and easy to understand. The researchers generated a KBDeX visualization to highlight connections between keywords used by students in their discourse (figure 2a). The network was shared with students to facilitate classroom discussion and to support and prompt their third model.

**Phase 3: New learning analytics, new questions (Session 6-7)**

In phase 3, students generated more questions and ideas in a KF view called “New Model, New Problem”, driven by their work on previous models. The researchers generated an updated network using KBDeX to support and prompt their fourth model (figure 2b). At the end of this phase students also took a domain post-test to assess the knowledge gained from working in the KBM environment.

**Figure 2**  
*Word Networks (a) Phase 2 Showing the “Vibration” as the Central Concept of Students’ Discourse (b) Phase 3, Showing Energy at the Center, Indicating Advancing from the Phenomena to Mechanism*



**Data sources**

We included the following data sources: (1) Pre and post-domain tests designed to examine students' knowledge of natural hazards. The test consists of five open ended questions (for e.g. Why do earthquakes, volcanoes, and tsunamis happen?). The test also included one question asking students to represent their understanding through a drawing – which was considered their first model iteration; (2) Three sets of prompt sheets for each student, containing the model drawings for remaining iterations; models were coded to determine levels of causal reasoning; (3) KF notes contributed by students were coded to assess depth of students' discourse at different

phases; (4) teacher's slides shared with students each lesson to help identify key themes in the different phases. All data was translated from Chinese to English by a researcher who is also a certified translator.

## Results and analysis

### Development of students' models, KF discourse, and domain understanding

#### Models

Student models were coded using a framework adapted from the work of Cabello et al. (2021) – see table 1. A total of 200 student models were coded. Two researchers coded 40% of the models independently. Inter-rater reliability was achieved ( $\kappa = 0.75$ ). After disagreements were discussed, the remaining models were coded by one researcher.

**Table 1**

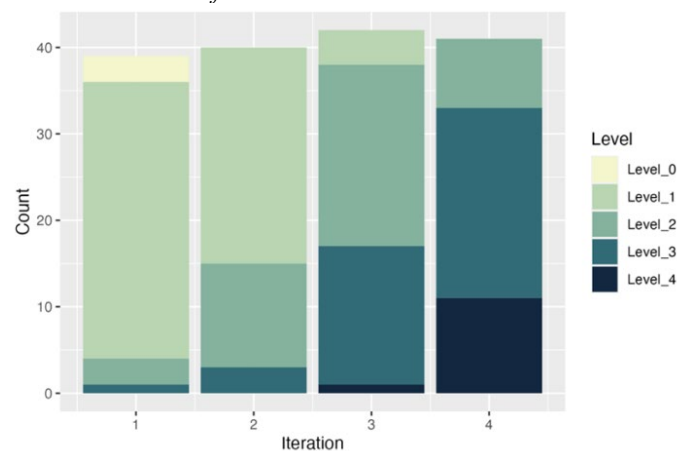
*Coding Rubric for Primary Students' Models Inspired by Cabello et al. (2021).*

Level 0	No Understanding - No recognizable understanding of the topic.
Level 1	Sensory Level of Understanding - Recognizes phenomena based on sensory perceptions with a focus on immediate and visible consequences.
Level 2	Causal Understanding - Understands cause-and-effect relationships without delving into dynamic mechanisms or non-visible theories.
Level 3	Causal Understanding with Dynamic Mechanisms - Recognizes dynamic mechanisms and processes, going beyond simple causality.
Level 4:	Systemic Understanding - Recognizes interconnected systems, secondary consequences, and feedback loops within the topic, demonstrating a systemic understanding.

Our analysis reveals a notable increase in levels as iterations unfold as demonstrated in figure 4. We note that by the third iteration, 40.5% of students who submitted models have achieved a level 3 or 4, followed by 80.4% in iteration 4. Attaining a level 4 involves a more sophisticated engagement with systems thinking, involving feedback loops and changes over time. This is an advanced skill that is difficult for grade 5 students to achieve without prior introduction and appropriate scaffolding. However, by the fourth iteration we observed that 27.8% of students were able to attain this level.

Specifically, we were interested in identifying points at which students' models improved most across different iterations. We conducted Friedman test on the complete set to examine the differences in student scores across the four different points in time. Results suggest a statistically significant difference in levels between at least two iterations ( $\chi^2 = 73.1, p < 0.05$ ). We then conducted a paired Wilcoxon signed-rank test to assess changes between successive iterations. Results show a significant change in levels between iterations 1 and 2 ( $r = 0.432, p < 0.05$ ). However, a larger change was observed between iterations 2 and 3 ( $r = 0.77, p < 0.001$ ) and iteration 3 to 4 ( $r = 0.643, p < 0.001$ ). This suggests that the use of analytics to visualize collective ideas helped students create more refined models which demonstrate deeper causal understanding of the topics.

**Figure 4**  
*Distribution of Model Levels Across Iterations*



#### KF discourse

The depth of knowledge building discourse was assessed by analyzing and coding all KF notes. Notes were coded by two researchers independently and achieved inter-rater reliability ( $\kappa = 0.77$ ). The researcher then met and resolved disagreements. We adopted ten sub-codes categorized into *Questioning*, *Theorizing*, and *Community* discourse moves. As indicated in table 3 students engaged in low-level *theorizing* (37.63%) during phase 1. In phases 2 and 3, students significantly engaged in high-level community synthesis (55.56% and 40.48%, respectively). In phase 3, students additionally engaged in community lending support (23.81%).

**Table 2**

*Numbers and Percentages of KF Discourse Moves. (\* represents high-level discourse moves)*

Category	Discourse Move	Phase 1	Phase 2	Phase 3
Questioning	Q1 - fact-oriented	10 (6.4%)	0 (0%)	0
	Q2 - explanation seeking	18 (11.5%)	0 (0%)	3 (5.7%)
	Q3 - sustained inquiry*	15 (9.6%)	0 (0%)	<b>29 (55.8%)</b>
Theorizing	T1 - simple claim	3 (1.9%)	0 (0%)	0 (0%)
	T2 - proposing an explanation	<b>70 (44.6%)</b>	9 (20.9%)	8 (15.4%)
	T3 - supporting an explanation*	26 (16.6%)	3 (7%)	5 (9.6%)
	T4 - improving an explanation*	12 (7.6%)	2 (4.7%)	<b>7 (13.5%)</b>
Community	C1 - bridging knowledge*	15 (9.6%)	2 (4.7%)	5 (9.6)
	C2 - Synthesis*	14 (8.9%)	<b>23 (53.5%)</b>	0 (0%)
	C3 - lending support*	5 (3.2%)	5 (11.6%)	0 (0%)

One-way MANOVA results showed a significant difference in students' use of knowledge building discourse moves across phases,  $F(20, 150) = 9.170, p < .001$ , Wilks'  $\Lambda = 0.202$ , partial  $\eta^2 = 0.55$ . Follow-up univariate ANOVAs revealed that students used significantly more proposing an explanation moves in phase 1 ( $p < 0.05$ ) than in phase 1 and 3. In phase 2, students used significantly more community-synthesis moves ( $p < .001$ ). The use of sustained inquiry moves initially decreased from phase 1 to phase 2, then rebounded in phase 3, with post-hoc analysis showing statistically significant increase between phase 1 and 3 ( $p < 0.001$ ). These findings suggest a shift from idea improvement using questioning and theorizing in phase 1, to deeper collective knowledge advancement activities involving synthesis in phase 2, followed by more sustained inquiry in phase 3.

### Domain understanding

Results from a paired-samples *t*-test show significant differences between students' pre-test ( $M = 40.39, SD = 16.93$ ) and post-test ( $M = 67.25, SD = 12.86$ ),  $t(33) = 9.79, p < .001$ , Cohen's  $d = 1.80$ . Results indicate students obtained higher scores in the post-tests with their experience in the designed knowledge-building environment.

## Predictors of modeling and domain understanding

### Predictors of modeling ability

Hierarchical regression show that students' prior modeling predicted their post modeling ( $F(1, 35) = 7.25, p = .011$ ) and explained 17.2% of the variance; when students' high-level discourse moves were entered, an additional 20.4% of the total variance was explained ( $F(7, 29) = 2.50, p = .039$ ). These findings indicate that students' in-depth knowledge building discourse was a good predictor of students' modeling over and above their prior modeling ability.

### Predictors of scientific understanding

Hierarchical regression show that students' prior scientific understanding predicted their scientific understanding ( $F(1, 32) = 8.12, p = .008$ ) and explained 20.2% of the variance; when students' post-modeling values were entered, an additional 11.1% of the total variance was explained ( $F(2, 31) = 7.05, p = .003$ ); when students' high-level discourse moves were entered, another 15.5% of the total variance was explained ( $F(8, 25) = 2.75, p = .025$ ). These findings indicate that students' in-depth knowledge building discourse was a good predictor of scientific understanding over and above their prior scientific understanding and their modeling.

## Models and knowledge building practices

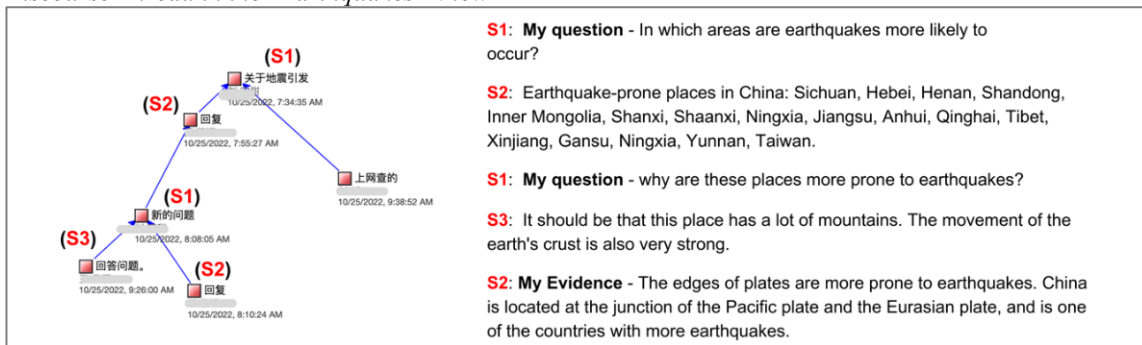
Three themes emerged from the qualitative analysis of students' models, KF contributions, and teacher slides.

### Theme 1: Knowledge building discourse for modeling.



In phase 1, students proposed questions like “How do earthquakes form?” and proposed theories like “earthquakes are mostly accompanied by tsunamis and volcanic eruptions”. Figure 5 shows a discourse thread which started by S1 asking about areas which are prone to earthquakes, followed by a response from S2 listing earth-prone locations. This triggered S1 to further question why these places are prone to earthquakes, to which S2 responded by explaining that it is because they are located at plate junctions, and that plate edges are more prone to earthquakes. This demonstrates students moving from questioning to explaining mechanisms, attributing earthquake susceptibility to the specific geographical locations where plates meet or interact.

**Figure 5**  
Discourse Thread in the “Earthquakes” View

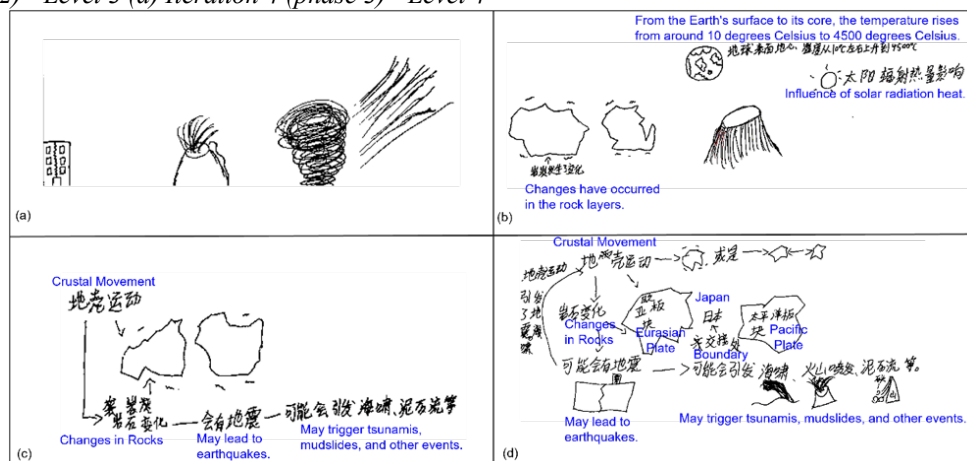


Later, another student synthesized peers’ ideas related to this question.

After discussion, I found that most of the places where earthquakes often occur are "located at the junction of the Eurasian plate and the Pacific plate, and the earth's crust is active" ("Earthquakes often occur in Japan" by [Student]) - "The main cause of earthquakes is that the earth's plates squeeze and collide with each other, causing dislocation and rupture at the edge of the plates and within the plates." ("The Causes of Earthquakes" by [Student])

In this phase, students drew their second iteration after improving ideas and synthesizing notes. Figure 6a and b shows the first two iterations by student S2. The first iteration created in the pre-test depicts the phenomena at a sensory level. In the second iteration created in phase 1, the student selected the two questions to demonstrate through his model “Why are there frequent earthquakes, volcanic eruptions and tsunamis in Japan? Why is magma hot?”. The drawing depicts two adjacent moving plates, which shows the student attempting to incorporate ideas emerging from discourse.

**Figure 6**  
S2’s Models (a) Iteration 1 (pre-test) - Level 1 (b) Iteration 2 (phase 1) - Level 2 (c) Iteration 3 (phase 2) - Level 3 (d) Iteration 4 (phase 3) - Level 4



Theme 2: Learning analytics for synthesis and model improvement.

In phase 2, students synthesized ideas from the three phenomena views into the view “Scientific Exchange: Summary and Reflection” and identified 10 categories of questions related to the three phenomena. Student S2 whose models are in figure 6 constructed the following note

A better idea - why Japan often has earthquakes, volcanic eruptions and tsunamis: because the island of Japan is right at the junction of the Eurasian plate and the Pacific plate, so the Eurasian plate collides with the Pacific plate when it moves.

The word network analytic shared with students in this phase (figure 3a) shows the word “vibration” at the center of the three phenomena. S2 then constructed her third model (figure 6c), representing vibration as a mechanism which could trigger earthquakes in areas located at the junction of plates. The student also highlighted that earthquakes could trigger other hazards like tsunamis and mudflows. This suggests movement towards a more advanced causal understanding with dynamic mechanisms (connecting vibration to earthquake triggers).

### Theme 3: KBM for sustained collective knowledge advancement and deeper modeling practices.

In phase 3 students deepened their inquiry by highlighting ideas and questions which emerged as they constructed models. Students posed more questions related to mechanisms, such as “What is the relationship between energy and plate motion?”. An updated visualization of KF discourse was shared with students (figure 3b), showing “energy” at the center of the network. Figure 7 shows an example of a students’ final model showing energy released by plate vibrations.

**Figure 7**  
*Model Iteration 4 Showing Energy Released by Plate Vibration.*

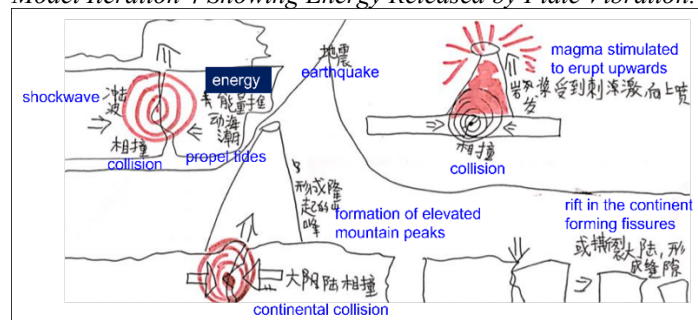
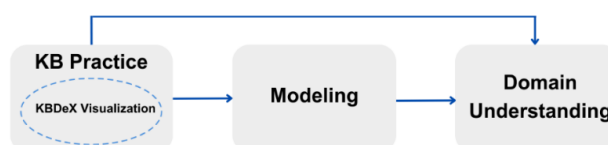


Figure 6d shows the final model produced by student S2, showing a systemic understanding of the phenomena. The model represents students’ understanding of how crustal movements can lead to earthquakes and subsequently influencing more crustal movement resulting in more earthquakes (feedback loop). The student also highlighted more phenomena associated with plate collision like mudflows and volcanoes. These results suggest a more comprehensive grasp of the interconnections and mechanisms underlying natural hazards, supported by the word network visualization.

### Implications and significance

This study contributes to both the modeling and Knowledge Building literature, two key areas in the learning sciences. It investigated how an iterative KBM approach aided by learning analytics visualizations helped students develop scientific understanding of dynamic mechanisms related to natural hazards – which represents a complex system. Our findings suggest that visualizations of knowledge building discourse played a significant role in promoting deeper engagement with modeling practices and improving students’ domain understanding. Figure 8 describes our hypothesis, which was supported by our findings.

**Figure 8**  
*Discourse Visualization Enabling Deeper Modeling Practices and Improved Domain Understanding*



We hypothesized that engaging students in discourse and showing them discourse visualizations can help facilitate epistemic knowledge work where students can reflect on the changing status of their collective ideas,

which can help them identify next steps in their inquiry. Emerging themes of our study underscore the impact of learning analytics in providing students with an accessible way to visualize their collective ideas, which could be particularly helpful in large classes where it is difficult for students to read all notes. Visualizations also helped students understand relationships between key concepts. Visualizations generated as discourse proceeded allowed students identify new emerging concepts and relationships, which students can then integrate into their models.

This line of research explores the synergies in the key areas of Knowledge Building, learning analytics, and modeling to facilitate understanding of complex, dynamic systems. Further research is being conducted in this space, with specific focus on computation modeling, where students could use simulations to test theories and predictions that arise during knowledge building discourse, where knowledge emerges from social interactions.

## References

- Assaraf, O. B.-Z., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, 42(5), 518–560.
- Bereiter, C., & Scardamalia, M. (2014). Knowledge Building and Knowledge Creation: One Concept, Two Hills to Climb. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge Creation in Education* (pp. 35–52). Springer Singapore.
- Cabello, V. M., Moreira, P. M., & Griño Morales, P. (2021). Elementary Students' Reasoning in Drawn Explanations Based on a Scientific Theory. *Education Sciences*, 11(10), Article 10.
- Feng, X., van Aalst, J., Chan, C. K. K., & Yang, Y. (2021). Developing progressive knowledge building through idea-friend maps and opportunistic collaboration. In d. Vries, H. E., Y., & A. J. (Eds.), *15th International Conference of the Learning Sciences (ICLS) 2021* (pp. 370-377). International Society of the Learning Sciences.
- Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22(9), 1011–1026.
- Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to Learn About Complex Systems. *Journal of the Learning Sciences*, 9(3), 247–298.
- Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, 28(1), 127–138.
- Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing Debates in Learning and Educational Research: Toward a Complex Systems Conceptual Framework of Learning. *Educational Psychologist*, 51(2), 210–218.
- Jacobson, M. J., & Wilensky, U. (2006). Complex Systems in Education: Scientific and Educational Importance and Implications for the Learning Sciences. *Journal of the Learning Sciences*, 15(1), 11–34.
- Knuutila, T. (2011). Modelling and Representing: An Artefactual Approach to Model-Based Representation. *Studies in History and Philosophy of Science Part A*, 42(2), 262–271.
- Lehrer, R., & Schauble, L. (2005). Cultivating Model-Based Reasoning in Science Education. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (1st ed., pp. 371–388). Cambridge University Press.
- Meadows, D. H. (2008). Thinking in systems: A primer (D. Wright, Ed.). Chelsea Green Pub.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60(5), 903-921.
- Perkins, D. N., & Grotzer, T. A. (2005). Dimensions of Causal Understanding: The Role of Complex Causal Models in Students' Understanding of Science. *Studies in Science Education*, 41(1), 117–165.
- Penner, D. E. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching*, 37(8), 784–806.
- Scardamalia, M. (2002). Collective Cognitive Responsibility for the Advancement of Knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago: Open Court.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Stern, R. J. (2007). When and how did plate tectonics begin? Theoretical and empirical considerations. *Chinese Science Bulletin*, 52(5), 578–591.
- Wilensky. (2003). Statistical Mechanics for Secondary School: The GasLab Multi-agent Modeling Toolkit. *International Journal of Computers for Mathematical Learning*, 8(1), 1–41.

## Exploring Teacher Orchestration Load in Scripted CSCL: The Role of Heart Rate Variability

Lubna Hakami, Universitat Pompeu Fabra, lubna.hakami01@estudiant.upf.edu

Batuhan Sayis, Universitat Pompeu Fabra, batuhan.sayis@upf.edu

Ishari Amarasinghe, Radboud University, ishari.amarasinghe@donders.ru.nl

Davinia Hernandez-Leo, Universitat Pompeu Fabra, davinia.hernandez-leo@upf.edu

**Abstract:** Integrating Computer-Supported Collaborative Learning (CSCL) tools in the educational environment is considered to enhance collaborative learning experiences. However, previous research findings have indicated that classroom orchestration adds to the workload of the teachers, therefore potentially introducing their perceived stress levels. This study investigates the extent to which teachers experience stress when orchestrating CSCL activities using multimodal data. Physiological data such as heart rate variability (HRV) was used in this study in addition to subjective data such as self-reported questionnaires and observation notes. A combination of multimodal data and single-subject research design (SSRD) allowed us to investigate the influence of the CSCL orchestration tool on teachers' stress levels. Based on the collected multimodal data, the findings show that the studied tool does not increase the stress levels of the teachers.

### Introduction

The field of Computer-Supported Collaborative Learning (CSCL) aims to bring computer support to enhance collaborative learning situations (Stahl et al., 2006). Within the context of CSCL, group learning can be pre-emptively structured using collaborative learning scripts (by defining activity sequences, distribution of roles, groups, and resources) therefore creating opportunities for productive interactions to occur among learners (Dillenbourg et al., 2002; Fischer et al., 2007). For instance, Pyramid script provides a structure for collaboration that guides students to achieve a gradual consensus on a complex problem within different social levels (from individual to small groups and later in larger groups) (Hernández-Leo et al., 2006). Teachers may use CSCL tools to design and deploy scripted activities in everyday teaching and learning situations.

The notion of orchestration load aims to capture the attentive effort teachers encounter when regulating or orchestrating multiple learning activities and processes (Prieto et al., 2015). Literature provides multiple definitions for orchestration load (Cuendet et al. 2013; Prieto et al. 2017) and broadly it can be perceived as a concept that contributes to both physical and cognitive effort teachers experience when regulating learning activities in real-time (Amarasinghe et al., 2021). In scripted CSCL situations, the notion of orchestration load becomes a relevant aspect as teachers are required to manage and divide their attention across multiple groups simultaneously to take actions, while also focusing and adapting to the other contextual constraints. For example, the time available to finish the learning activity, and the support provided by technological tools (mirroring vs guiding tools) play a role in influencing the orchestration load experienced by teachers (Amarasinghe et al., 2021).

From a theoretical perspective, as outlined in Activity-Centered Analysis and Design (ACAD) framework (Goodyear et al, 2021), teachers may focus on different facets of a learning situation during its run-time, i.e., *epistemic*, *social*, and *set* aspects. The *epistemic* aspect refers to the learning task at hand, the *social* aspect refers to the structure of collaboration and the *set* aspect includes the technological tooling which supports the deployment of CSCL activity itself. In the present study using multimodal data collected from five teachers, we aim to study how the *set* aspect (which includes the technological tooling available to deploy and monitor CSCL activities) influences the stress levels of the teachers in real-time (while epistemic and social aspects remain constant, i.e., similar open-ended tasks were given to students during CSCL sessions (*epistemic* aspect) and all activities were scripted according to the pyramid script hence following similar group structures (*social* aspect)). Therefore, this study aims to explore whether CSCL orchestration tools influence the stress levels of the teachers, which would help to inform design guidelines for similar tools in the future.

Moreover, previous research has attempted to estimate orchestration load using different data sources, such as log data, post-activity questionnaire responses as well as observation notes in the form of audio and video data (Prieto et al., 2015). Despite the rich detailed quantitative (e.g., log data) and qualitative information (e.g., self-reports and observations) that could be collected using such means, those may not always produce reliable and valid estimates on orchestration load due to many reasons. For instance, teachers may answer the self-reporting questionnaires at the end of the CSCL activity to reflect on overall experience as it is difficult to employ

questionnaires multiple times throughout the collaboration process to report specific moments that contributed to their orchestration load. Transcribing observation notes and qualitative coding can also introduce errors and biases. On the other hand, physiological measures may provide an alternative yet unobtrusive approach to collecting data from individuals continuously and non-intrusively without disturbing teachers' orchestration actions in CSCL activities. Despite the potential for data collection, research regarding the use of wearable sensors to estimate teachers' orchestration load and perceived stress in authentic learning situations remains largely underexplored. However, some studies have used sensors for teachers and students in learning environments as the sensors are increasingly being utilized to enhance insights into student learning experiences and teacher practices. On the students' side, previous studies used sensors to track students' physiological data (Giannakos et al., 2020) and quantify the classroom emotional climate (Gashi, Lascio & Santinii., 2018). On the teachers' side, there is a study that used wearable sensors to increase teachers' interest in living healthy lifestyles (Ertzberger et al., 2016). In addition, a study by Liang et al., (2019) used sensors to increase teacher-student interaction within the classroom by developing a smart interactive educational system.

To this end, this study aims to use physiological sensing technologies, following the research that has increasingly validated the use of sensors to measure stress (Kyriakou et al., 2019; Di Campli San Vito et al., 2023). Specifically, we use the Cosinuss<sup>o</sup> sensor (1) to capture the physiological data (i.e., heart rate variability (HRV)) when monitoring CSCL activities. Studies have demonstrated the efficacy of HRV metrics as indicators of stress levels (Hjortskov et al., 2004; Weber et al., 2010). In addition to the physiological data, we collect multimodal data during the CSCL activities. The term "multimodal data" is used to describe data that is collected using several techniques. According to Prieto et al. (2018), audio, video, electrodermal activity data, eye-tracking, user logs, and click-stream data are examples of multimodal data. In the education field, several studies have multimodal data. For example, in our previous study (Hakami et al., 2022) we used data collected from different modalities (i.e., video and audio recordings, log data, and observation notes) to study teachers' orchestrations in different learning settings. Another study by Chejara et al. (2020) used multimodal data (i.e., audio and log data) to show that it is possible to assess collaboration quality and its sub-dimension quantitatively. In this study, we use multimodal data from the teachers during the activities (i.e., self-reported questionnaires, log data, and observation notes) to investigate whether CSCL orchestration tools affect the stress levels of the teachers. We use multimodal data analysis and single-subject research design (SSRD) to answer the following research question:

- To what extent are teachers experiencing reasonable levels of stress when orchestrating CSCL activities?

The rest of the paper is organized as follows: First, we describe the methods we used to answer the research questions. This section explains the study design, data collection, and materials used. Then, we present and discuss the study's findings. Lastly, we conclude the study and present the future work.

## Methods

This study employs a combination of multimodal data analysis and SSRD to explore the orchestration load experienced by teachers and whether the CSCL tools affect the stress levels of the teachers, taking into account both objective (i.e., physiological data from sensors and log data) and subjective (i.e., self-reported data and observation notes) measures.

### Data collection and procedures

In this study, we incorporated *objective data* collected from the participating teachers (i.e., physiological data). Participants are monitored by sensors placed in a non-invasive manner, allowing us to collect data without interrupting them and allowing them to interact naturally. Specifically, the Cosinuss<sup>o</sup> sensor (see Figure 1) was used to capture physiological data (i.e., heart rate variability HRV). HRV refers to the variation in the time intervals between consecutive heartbeats (Shaffer & Ginsberg, 2017). The sensor was placed in the teacher's ear. Before each session, the sensor is carefully cleaned with a damp cloth or a cotton tip moistened with a disinfectant product (i.e., hydroalcoholic gel) before placing it. Moreover, log data was collected to track various aspects of the teachers' actions during CSCL activities. The log data included the duration of each task, the number of script phases, the number of students who participated in the activity, the login mode from students, and the specific CSCL task given for each session. Second, *subjective data* were collected during the sessions. First, self-report questionnaires were given to the teachers after the activity to capture the teachers' perception of the activity and their experiences of stress. For example, the teachers have to respond to a question in the questionnaire: "Could you describe it in detail focusing on the specific emotion (frustration, confusion, etc.)? What was the trigger?". The qualitative data were analyzed using content analysis (Hsieh & Shannon, 2005), which is an approach for the interpretation of textual data. Another type of data collected is observation notes taken by the researchers during the CSCL activities. We used the observation notes to capture relevant details about the teachers' behavior and reactions to the activities. These behaviors were classified as teacher motion (e.g., the teacher walking around the

classroom), teacher interaction with students (e.g., the teacher provides directions to the class on how to use the tool), and announcements to class (e.g., time remaining for the activity). Thus, the Institutional Committee for Ethical Review of Projects (CIREF), with the assigned approval number 280, approved this study to collect data ethically.

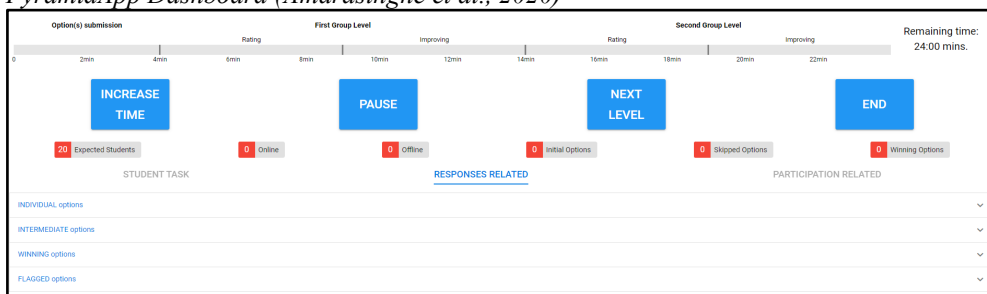
**Figure 1**  
*Cosinuss<sup>o</sup> Sensor*



PyramidApp tool was used to deploy Pyramid script-based CSCL activities in learning situations (Manathunga & Hernández-Leo, 2018). The tool consists of an authoring space that facilitates activity authoring; an activity enactment space for students; and a teacher-facing dashboard that provides orchestration support, such as information about students' activity participation, in addition to functionalities to adapt the script's flow in real-time. The Pyramid flow begins with students individually working on a task. In subsequent levels, students join in small groups to engage in discussions and collectively enhance their solutions. Later larger groups are formed by merging smaller groups to reach a consensus answer on the given task. During the activities, the teachers will monitor in real-time the activities done by students through the PyramidApp dashboard (see Figure 2). It displays information regarding the activity duration, the size of the class, and how many answers were received from the students (i.e., initial, skipped, and winning answers) in each activity. Moreover, a group of controls e.g., increase time, pause, move to the next level, and end control allows the teacher to adapt the learning flow in response to the classroom dynamics.

Five teachers from a Spanish public university, who had prior experience with the PyramidApp tool, participated in the sessions (2 females and 3 males). Each teacher conducted 5-7 classroom CSCL sessions. A total of 40 sessions were initially conducted. However, due to technical issues with the Cosinuss<sup>o</sup> sensor and missing data, 8 sessions had to be removed to maintain the integrity and accuracy of the dataset. A total of 32 sessions from five teachers were retained for further analysis. All the sessions were with a bachelor course except 4 sessions for teacher A were with a master course. Each CSCL session lasted around 10-35 minutes.

**Figure 2**  
*PyramidApp Dashboard (Amarasinghe et al., 2020)*



## Data preprocessing and analysis

The analysis of the physiological data was done using Kubios software, which supports the analysis of the data from heart rate monitors (Tarvainen et al., 2014). Cosinuss<sup>o</sup> sensors provide various data (e.g., HRV, accelerometer, temperature data, etc.). Before starting the analysis, we used the Cosinuss<sup>o</sup>One app to record and extract the data. Then, due to the low signal quality of the data collected, we used a filter from Kubios software for artifact correction. The Kubios offers a range of threshold-based artifact correction filters for HRV data analysis (Alcantara et al., 2020). These filters include the following options: (1) None (no correction applied), (2) Very Low, (3) Low, (4) Medium, (5) Strong, and (6) Very Strong filter. In our analysis, we utilized the "Very Strong" filter for the artifact correction. Subsequently, 46 features were extracted from the data including common

HRV metrics such as Standard Deviation of NN Intervals (SDNN), Root Mean Square of Successive Differences (RMSSD), Very Low-Frequency Power (VLF), Low-Frequency Power (LF), High-Frequency Power (HF), standard deviation of the IBIs for all sinus beats (SDRR), and the LF/HF ratio. Therefore, RMSSD was used in this study which is one of the metrics in Time-Domain Measurements (Shaffer & Ginsberg, 2017). It is worth mentioning that higher stress is associated with lower RMSSD values. In contrast, lower stress is associated with higher RMSSD values (Shaffer & Ginsberg, 2017).

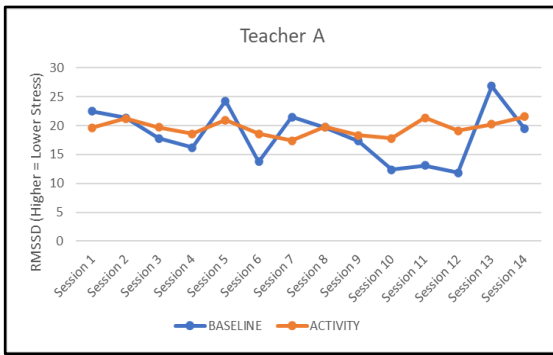
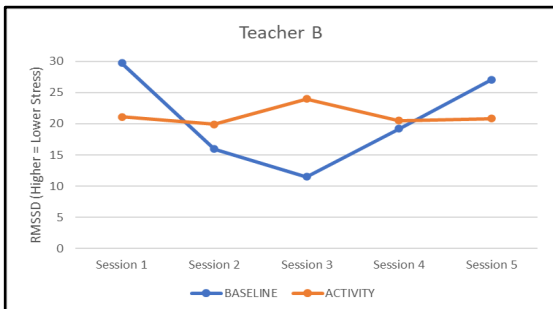
We used a visual inspection method, which is common in SSRD (Price et al., 2015). Visual inspection involves plotting individual participants' data and examining the data. According to Price et al (2015), three factors should be considered when using the visual inspection which are changes in the level of the dependent variable, the gradual increases or decreases in the dependent variable (i.e., trend), and the time it takes for the dependent variable to begin changing after a change in conditions (latency). In this study, we used trend analysis to investigate whether the PyramidApp tool affects teachers' stress levels by comparing baseline and CSCL activities. A baseline was established by using the first minute of each activity as a baseline. This allowed us to take into account individual differences and generate a baseline. At the beginning of each activity, the first minute was spent preparing the activities and providing the students with instructions on how to log in to the activity. Therefore, by plotting the RMSSD values across sessions for each teacher, we will be able to visually inspect the changes in the stress levels for each teacher.

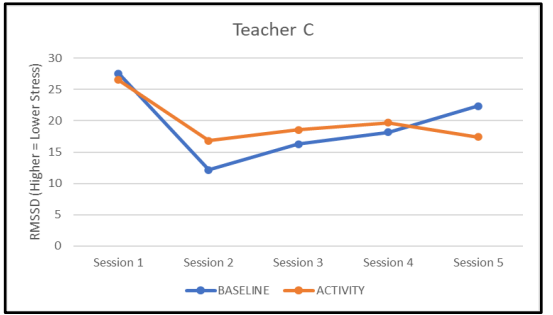
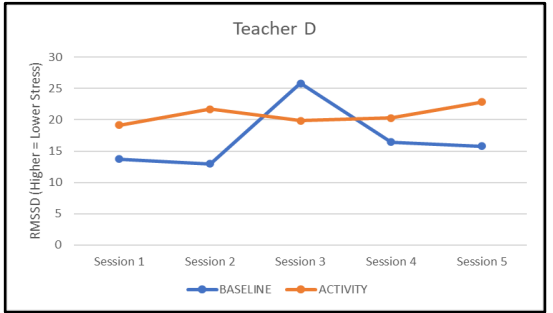
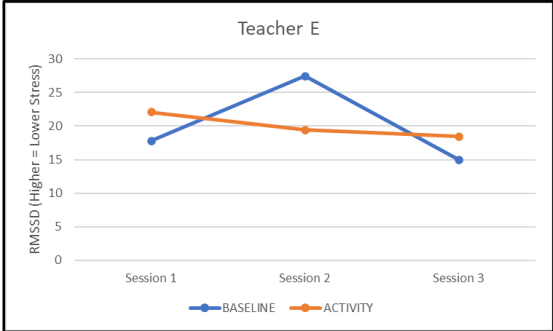
## Results

### Trend analysis: baseline and activity comparison

We present the trend analysis of RMSSD values for each teacher across the sessions to compare the baseline (i.e., the first minute of the activity) and the rest of the activity for each teacher. Table 1 provides the results obtained from the trend analysis to compare the baseline and the activity across sessions for each teacher.

**Table 1**  
*Baseline vs. Activity RMSSD Comparison*

Explanation	Figures																																													
<p><b>Teacher A:</b> The activities generally show increased RMSSD values compared to the baseline, which suggests that the activities are typically associated with a decrease in stress for Teacher A. However, there are exceptions; sessions 1, 5, 7, and 13 exhibit lower RMSSD values during the activity than during the baseline, indicating that the activity may have increased stress levels in those instances. Overall, the variation of measurements during the activity across sessions is small as well as differences with the baseline, which suggest that stress levels triggered by the use of the PyramidApp tool are moderate.</p>	 <table border="1"> <caption>Teacher A RMSSD Data</caption> <thead> <tr> <th>Session</th> <th>Baseline</th> <th>Activity</th> </tr> </thead> <tbody> <tr><td>Session 1</td><td>22</td><td>19</td></tr> <tr><td>Session 2</td><td>21</td><td>21</td></tr> <tr><td>Session 3</td><td>18</td><td>19</td></tr> <tr><td>Session 4</td><td>16</td><td>20</td></tr> <tr><td>Session 5</td><td>24</td><td>21</td></tr> <tr><td>Session 6</td><td>14</td><td>18</td></tr> <tr><td>Session 7</td><td>21</td><td>17</td></tr> <tr><td>Session 8</td><td>19</td><td>18</td></tr> <tr><td>Session 9</td><td>17</td><td>17</td></tr> <tr><td>Session 10</td><td>12</td><td>17</td></tr> <tr><td>Session 11</td><td>13</td><td>21</td></tr> <tr><td>Session 12</td><td>11</td><td>18</td></tr> <tr><td>Session 13</td><td>27</td><td>20</td></tr> <tr><td>Session 14</td><td>19</td><td>21</td></tr> </tbody> </table>	Session	Baseline	Activity	Session 1	22	19	Session 2	21	21	Session 3	18	19	Session 4	16	20	Session 5	24	21	Session 6	14	18	Session 7	21	17	Session 8	19	18	Session 9	17	17	Session 10	12	17	Session 11	13	21	Session 12	11	18	Session 13	27	20	Session 14	19	21
Session	Baseline	Activity																																												
Session 1	22	19																																												
Session 2	21	21																																												
Session 3	18	19																																												
Session 4	16	20																																												
Session 5	24	21																																												
Session 6	14	18																																												
Session 7	21	17																																												
Session 8	19	18																																												
Session 9	17	17																																												
Session 10	12	17																																												
Session 11	13	21																																												
Session 12	11	18																																												
Session 13	27	20																																												
Session 14	19	21																																												
<p><b>Teacher B:</b> Overall, in 3 out of the 5 sessions (sessions 2, 3, and 4), RMSSD values during the activity were higher than during the baseline, indicating that the activity could be associated with moments of relaxation or decreased stress. In the case of sessions 1 and 5, the difference with the baseline is relatively small and the measurements are within the range of measurements for that teacher.</p>	 <table border="1"> <caption>Teacher B RMSSD Data</caption> <thead> <tr> <th>Session</th> <th>Baseline</th> <th>Activity</th> </tr> </thead> <tbody> <tr><td>Session 1</td><td>30</td><td>21</td></tr> <tr><td>Session 2</td><td>16</td><td>20</td></tr> <tr><td>Session 3</td><td>11</td><td>24</td></tr> <tr><td>Session 4</td><td>19</td><td>20</td></tr> <tr><td>Session 5</td><td>27</td><td>21</td></tr> </tbody> </table>	Session	Baseline	Activity	Session 1	30	21	Session 2	16	20	Session 3	11	24	Session 4	19	20	Session 5	27	21																											
Session	Baseline	Activity																																												
Session 1	30	21																																												
Session 2	16	20																																												
Session 3	11	24																																												
Session 4	19	20																																												
Session 5	27	21																																												

<p><b>Teacher C:</b> Like teacher B, teacher C shows higher RMSSD values during the activity compared to the baseline in 3 out of the 5 sessions (sessions 2, 3, and 4), which could imply that the activity is associated with decreased stress levels or increased relaxation during these sessions. In the case of sessions 1 and 5, the difference with the baseline is very small, suggesting that the increase in stress is moderated.</p>	
<p><b>Teacher D:</b> Across all sessions except for session 3, the RMSSD values during the activity are higher than during the baseline. This consistently higher value during the activity implies that teacher D typically experiences lower stress levels during the activity than at baseline. When compared across sessions, overall, the variation of measurements during the activity is small.</p>	
<p><b>Teacher E:</b> In some instances, the activities seem to induce less stress than the baseline (2 out of 3 sessions). Overall, the variation of measurements during the activity across sessions is small.</p>	

In summary, the RMSSD values when analyzed across multiple sessions for teachers A through E suggest that the activities performed generally resulted in higher RMSSD values compared to the baseline measurements (i.e., 21 out of 32 CSCL sessions). Since higher RMSSD values are typically associated with lower stress levels, this pattern indicates that the activities were more often associated with a state of relaxation when compared to the state of the teacher before the activity. The constant tendency across several teachers where the activity phases resulted in increased RMSSD values in the majority of the sessions. Teacher A displayed increased RMSSD values during activities in most sessions, with exceptions that warrant closer examination for stress factors. Teachers B and C showed higher RMSSD values during activities in three out of five sessions, reinforcing the trend of activities being less stressful. Teacher D presented a consistent pattern where almost all sessions during the activity phases had higher RMSSD values than the baseline, except for a slight decrease in session 3. Similarly, teacher E also had higher RMSSD values during activities in two out of three sessions, indicating lower stress levels during those activities.

In the cases where measurements indicate increased levels of stress (i.e., decrease in RMSSD values) when compared to the baseline, differences are relatively small for all teachers, being especially small for specific teachers (A, C, E) but quite small also for B and C. Moreover, the variation across sessions of measurements during the activity is small. The general stability and higher RMSSD values during most activities, when compared to the baseline, suggest that the activities, facilitated by a teacher CSCL tool such as PyramidApp tool, do not increase the stress of the teachers or raise only moderate levels of stress. This result shows that the CSCL orchestration tool (i.e., PyramidApp) contributes to a comfortable and controlled (in terms of pressure) environment for teachers during its operation.

## High-stress sessions analysis



Thus, in exploring the differences between the baseline and the activity, we observed particular sessions for each teacher where the RMSSD values were lower during the activity than in the baseline, suggesting an increase in stress level. For instance, for teacher A, this was seen in sessions 1, 5, 7, and 13; for teacher B, sessions 1 and 5; for teacher C, only in session 5; for teacher D, only in session 3; and for teacher E, only in session 2. Therefore, we corroborated the self-reported data and observation notes from these sessions to provide additional context that may help us to find the potential causes for the increased stress levels. According to self-reported data, teacher A indicated several sessions where stressful moments were experienced, aligning with the sessions where lower RMSSD values were observed, except for one session. In contrast, teachers B, C, D, and E did not report any stressful moments during the sessions. However, we used the observation notes to find what caused increased stress levels. Following, we analyze qualitative data for each teacher across those stressful sessions.

For teacher A, first, in session 1, the teacher reported two stressful moments. The first moment was experienced due to the students having problems accessing Google Forms provided by the teacher, which occurred after the activity. This particular instance of stress was not a direct result of the activity's structure, or the demands of the PyramidApp tool used. For the second moment, the teacher noted the cognitive effort involved in deciding which options to flag (this feature is used in the PyramidApp tool to mark the answers provided by students for further discussion) during the Pyramid activity. The stressful moments were not related to the functionality of the PyramidApp tool. However, this feature is optional in the PyramidApp tool. Similarity two stressful moments were reported in session 5. The first moment was because of managing the inclusion of online students, which required the teacher to stay within the proximity of a laptop, which added to the teacher's demands, potentially splitting their attention between in-person and online students. This was the only session that included online students besides the in-person students. Another moment happened when the teacher realized that the laptop was running out of power induced logistical stress, as the teacher had to locate a power source while controlling the flow of the session. We can categorize these moments as technological difficulties. Therefore, these moments were not related to the activity or the specific functionalities of the PyramidApp tool.

When it comes to session 13, the session presented a different situation where the teacher reported that he/she experienced a small moment of stress due to a personal issue which was cleaning glasses and a problem with the bag clipper. This instance, while indeed leading to a momentary increase in stress, was unrelated to the activity or the functionalities of the PyramidApp tool. Such personal inconveniences, although minor and transient, underscore the variety of stressors teachers may encounter that are not directly linked to the orchestration. This momentary stress, as reflected in the RMSSD data, demonstrates the sensitivity of physiological measures in detecting even the subtlest fluctuations in stress levels.

Lastly, although the teacher did not report any stressful moments during session 7, the RMSSD data suggests otherwise, showing decreased values that are typically associated with higher stress levels. This discrepancy between the subjective report and the objective physiological data could suggest that the teacher may not have been consciously aware of the stress experience. Thus, we used the observation notes taken by researchers in the classroom to find out if there were things that were happening that caused the stress. Based on the observation notes, the teacher was moving around the class with the tablet to explain and control the activity through the orchestration PyramidApp dashboard until the activity finished. The physical activity of moving around the class may be a factor that can elevate physiological stress indicators like RMSSD. This highlights a critical aspect of teacher stress that can be present even without conscious recognition, and that objective measures like RMSSD can offer an understanding of teachers' stress levels that might otherwise go unnoticed.

Regarding teacher D, session 3 as indicated by RMSSD has high stress, but there were not any stressful moments reported by this teacher. However, the teacher answered open-ended questions in the questionnaire "Is there anything else worth mentioning about this session?", and the teacher reported observing better student behavior during activities conducted in the student's non-anonymous login mode. The session was conducted in an anonymous mode (i.e., the students did not have to provide their names in the activity), which, as suggested by the RMSSD data, may have contributed to an elevated stress level for the teacher.

Teacher E experienced high stress in session 2. However, the teacher did not report any stressful moments in the questionnaire. When we looked at the observation notes, we found that this high stress could happen because of the physical activity of the teacher. The teacher was walking around the classroom until the activity finished. This is similar to teacher A behavior in session 7.

Teachers B and C, despite the absence of stressful moments reported or noted stress indicators in the observation notes, the RMSSD data indicated heightened stress levels. This discrepancy suggests that teachers may not always be conscious of their physiological stress responses, as RMSSD is known to capture subtle autonomic changes that can occur.

To conclude, it became evident from the trend analysis of RMSSD values that instances of elevated stress among teachers were largely unconnected to the functionalities of the PyramidApp tool or the activities. Despite

lower RMSSD values suggesting increased stress levels during certain sessions, the self-reported data, particularly from teacher A, linked these moments to external technological issues (e.g., the issue with the laptop battery) rather than the PyramidApp tool or the activities. However, in one situation the teacher reported one stressful moment about using the flagging feature in the PyramidApp tool (which is an optional feature in the tool). Furthermore, teachers B, C, D, and E did not report any moments that contributed to increasing the stress level, indicating that the PyramidApp did not inherently contribute to stress. This finding is supported by the observation notes, which often associated increased RMSSD readings with factors external to the teaching activity, such as physical movement or the students' login mode, rather than the PyramidApp. The consistency across multiple sessions in demonstrating that activities facilitated by the tool do not raise significant stress levels suggests that the tool is capable of supporting teachers in a CSCL environment without adding workload.

## Conclusion

The study explores the orchestration load experienced by teachers and whether the CSCL tools affect the teachers' stress levels using a multimodal data analysis approach. This approach proved to be particularly effective in understanding the teacher's stress in the CSCL environment considering both objective (i.e., physiological data from sensors and log data from PyramidApp database) and subjective (i.e., self-reported data and observation notes) measures. The findings indicate the stability of higher RMSSD values during the majority of activities (i.e., 21 out of 32 CSCL sessions), relative to baseline measurements, indicates that the use of the PyramidApp tool during CSCL sessions does not significantly elevate teacher stress levels and, in some instances, is associated with only a modest increase in stress. In addition, the higher levels of stress as indicated by lower RMSSD, were generally not associated with the functionalities of the PyramidApp tool or the structure of the activities itself. Instead, stress levels were influenced by external factors such as technological difficulties, actions by students, physical actions, or personal issues. Overall, we observed that the PyramidApp tool used by teachers in CSCL activities did not increase teacher stress.

The qualitative data for this study was limited to self-reported questionnaires and observation notes. In-depth interviews with teachers can be used in future work to offer a richer understanding of teacher's experiences and support the subjective measures. Moreover, future work should consider incorporating physiological measures, such as electrodermal activity (EDA), also known as skin conductance, which could enhance the understanding of teacher stress in CSCL environments. The number of EDA peaks can provide an indicator of how emotionally evocative certain stimuli are for subjects (Salim & Partnerships, 2020). Therefore, EDA peaks will be analyzed to see how dashboard design (e.g., mirroring and guiding) and learning settings (e.g., face-to-face, online, and hybrid) can affect the teachers' orchestration load during CSCL activities. In addition, combining EDA with a multimodal approach that assesses emotional valence (e.g., self-report) can eventually provide deeper insights into both emotional intensity and valence (Salim & Partnerships, 2020).

## Endnotes

(1) <https://www.cosinuss.com/en/>

## References

- Amarasinghe, I., Hernández-Leo, D., Michos, K., & Vujovic, M. (2020). An actionable orchestration dashboard to enhance collaboration in the classroom. *IEEE Transactions on Learning Technologies*, 13(4), 662-675.
- Amarasinghe, I., Hernández-Leo, D., & Ulrich Hoppe, H. (2021). Deconstructing orchestration load: comparing teacher support through mirroring and guiding. *Intern. J. Comput.-Support. Collab. Learn* 16, 307-338
- Alcantara, J. M., Plaza-Florido, A., Amaro-Gahete, F. J., Acosta, F. M., Migueles, J. H., Molina-Garcia, P., ... & Martinez-Tellez, B. (2020). Impact of using different levels of threshold-based artefact correction on the quantification of heart rate variability in three independent human cohorts. *Journal of clinical medicine*, 9(2), 325.
- Chejara, P., Prieto, L. P., Ruiz-Calleja, A., Rodríguez-Triana, M. J., Shankar, S. K., & Kasepalu, R. (2020). Quantifying collaboration quality in face-to-face classroom settings using mmla. In *International Conference on Collaboration Technologies and Social Computing* (pp. 159-166). Springer, Cham
- Cuendet, S., Bonnard, Q., Do-Lenh, S., & Dillenbourg, P. (2013). Designing augmented reality for the classroom. *Computers and Education*, 68, 557-569.
- Di Campli San Vito, P., Shakeri, G., Ross, J., Yang, X., & Brewster, S. (2023). Development of a Real-Time Stress Detection System for Older Adults with Heart Rate Data. In *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 226-236).

- Ertzberger, J., & Martin, F. (2016). Use of fitness bands by teachers in the classroom. *TechTrends*, 60(4), 392-397.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three worlds of CSCL. Can we support CSCL*, 61-91.
- Fischer, F., Kollar, I., Mandl, H., & Haake, J. M.(Eds.).(2007). *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (Vol.6).
- Gashi, S., Di Lascio, E., & Santini, S. (2018). Using students' physiological synchrony to quantify the classroom emotional climate. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*.
- Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., & Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies*, 136,102384.
- Hakami, L., Amarasinghe, I., Hakami, E., & Hernandez-Leo, D. (2022). Exploring Teacher's Orchestration Actions in Online and In-Class Computer-Supported Collaborative Learning. In *European Conference on Technology Enhanced Learning* (pp. 521-527). Cham: Springer International Publishing.
- Hernández-Leo, D., Villasclaras-Fernández, E. D., Asensio-Pérez, J. I., Dimitriadis, Y., Jorriñ-Abellán, I. M., Ruiz-Requies, I., & Rubia-Avi, B. (2006). COLLAGE: A collaborative Learning Design editor based on patterns. *Journal of Educational Technology & Society*, 9(1), 58-71.
- Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, 92, 84-89.
- Hsieh, H. F., Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research* 15(9), 1277-1288 .
- Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., ... & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors*, 19(17), 3805.
- Liang, J. M., Su, W. C., Chen, Y. L., Wu, S. L., & Chen, J. J. (2019). Smart interactive education system based on wearable devices. *Sensors*, 19(15), 3260. <https://doi.org/10.3390/s19153260>
- Manathunga, K., & Hernández-Leo, D. (2018). Authoring and enactment of mobile pyramid-based collaborative learning activities. *British Journal of Educational Technology*, 49(2), 262-275.
- Price, P. C., Jhangiani, R., & Chiang, I. C. A. (2015). *Research methods in psychology*. BCCampus.
- Prieto, L. P., Sharma, K., & Dillenbourg, P. (2015). Studying teacher orchestration load in technology-enhanced classrooms: a mixed-method approach and case study. In *Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15-18, 2015, Proceedings 10*(pp.268-281). Springer International Publishing.
- Prieto, L. P., Sharma, K., Kidzinski, Ł., & Dillenbourg, P. (2017). Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. *IEEE Transactions on Learning Technologies*.
- Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning*, 34(2), 193-203.
- Salim, R., & Partnerships, D. O. (2020). *What is EDA Peak Detection and How Does it Work*.
- Shaffer, F., & Ginsberg, J. P. (2017). *An overview of heart rate variability metrics and norms*.
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge: Cambridge University Press.
- Tarvainen, M. P., Niskanen, J. P., Lipponen, J.A., Ranta-Aho, P.O.,&Karjalainen, P. A.(2014).Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine*, 113(1),210-220.
- Weber, C. S., Thayer, J. F., Rudat, M., Wirtz, P. H., Zimmermann-Viehoff, F., Thomas, A., ... & Deter, H. C. (2010). Low vagal tone is associated with impaired post stress recovery of cardiovascular, endocrine, and immune markers. *European journal of applied physiology*, 109, 201-211.

## Acknowledgment

This work has been partially funded by the Planetary Wellbeing program of UPF (PLAWB00322). It has been also partially supported by MICIU/AEI/10.13039/501100011033 (PID2020-112584RB-C33, CEX2021-001195-M) and the Department of Research and Universities of the Government of Catalonia (SGR 00930). DHL also acknowledges the support by ICREA under the ICREA Academia programme.

# Short Papers



# Enhancing Argumentative Essay Structure and Quality through a Collaborative Argumentation Environment in University EFL Students

Li-Jen Wang, Language Center, National Central University, Taiwan, [tommywang@ncu.edu.tw](mailto:tommywang@ncu.edu.tw)

Ying-Tien Wu, Teng-Yao Cheng  
[ytwu@cl.ncu.edu.tw](mailto:ytwu@cl.ncu.edu.tw), [ttcheng@cc.ncu.edu.tw](mailto:ttcheng@cc.ncu.edu.tw)

Graduate Institute of Network Learning Technology, National Central University, Taiwan

**Abstract:** Previous research has highlighted significant challenges faced by university-level English as a Foreign Language (EFL) students in their argumentative essay writing. In this study, a collaborative argumentation environment was adopted where students could collectively generate and refine writing ideas before composing their argumentative essays. A quasi-experimental study was conducted to examine the impact of this environment on the EFL university students' argumentative writing performance in terms of structure and quality. Two groups of students participated: A control group (n=30) exposed to a face-to-face discussion environment, and an experimental group (n=25) using the collaborative argumentation environment. Both groups underwent pre- and post-study evaluations of their argumentative essays. The results showed that there were no statistically significant differences between the two learning environments when it came to improving argumentative essay structure and claims for both groups. In terms of grounds and rebuttals, the experimental group outperformed the control group ( $P < 0.05$ ).

## Introduction

Effective argumentation is crucial in today's knowledge-driven economy, especially in higher education, which is vital for success (Matos, 2021; Mitchell, 2000). Scholars stress its importance, leading to extensive research (Liu & Stapleton, 2020). Written argumentation, especially in essays, is a focus, particularly among EFL students, reflecting on their writing proficiency (Gabinete, 2021). Standardized tests like TOEFL and IELTS include argumentative writing tasks, emphasizing opinion expression and support. Despite its significance, research on EFL argumentative writing is limited, with students facing challenges in structure and quality (Lam et al., 2018). Effective teaching of argumentative writing is also challenging, with limited exploration of social interaction or cooperative learning (Azimah & Marlizayati, 2023).

Collaborative argumentation, involving dialogue and diverse viewpoints, is gaining traction (Wu & Tsai, 2011), aiming to broaden perspectives. Knowledge-building pedagogy, emphasizing collaboration and dialogue, shows promise (Scardamalia, 2002). The "Knowledge Forum" platform embodies this, providing scaffolds for collaborative argumentation (Scardamalia & Bereiter, 2003). It offers a collaborative environment for learners to exchange arguments and supporting ideas before writing argumentative essays. To this end, this study explored the impact of a collaborative argumentation environment on university EFL students' essays regarding structure and quality. Two research questions are addressed:

1. How does the experimental group's argumentative essay structure differ from the control group's after using the collaborative argumentation environment?
2. How does the experimental group's argumentative essay quality differ from the control group's after using the collaborative argumentation environment?

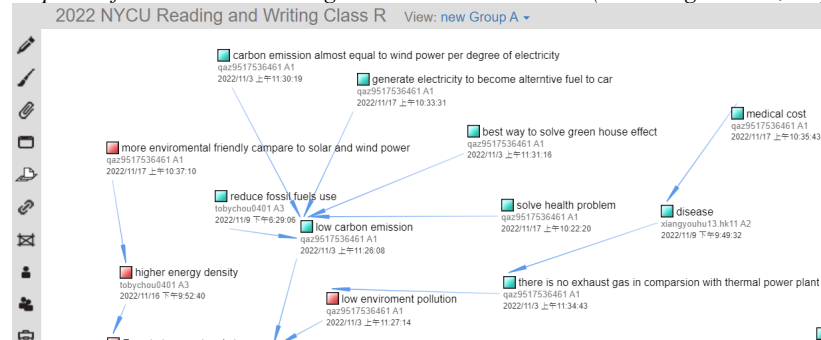
## Methods

### Research design and participants

This study conducted a quasi-experimental study with 55 university students from two classes in northern Taiwan during the spring semester of 2023. Participants, aged 20 or older and native Mandarin speakers, received two hours of weekly instruction over eight weeks to enhance argumentation and argumentative essay writing skills. They were randomly assigned to experimental (n=25) and control (n=30) groups. Pretests and post-tests were administered in weeks one and eight, respectively, to assess argumentative essay performance. In weeks 2 and 3, the experimental group learned collaborative argumentation using Knowledge Forum (KF), while the control group received traditional instruction. Both groups discussed the Social Scientific Issue (SSI) topic of Nuclear

Power and conducted research on sources. From weeks 4 to 7, the experimental group utilized KF for classroom discussions, while the control group engaged in face-to-face pair discussions. The experimental group, consisting of 3 to 4 students per group, actively searched for information online, posted their notes synchronously and asynchronously on KF, and responded to each other's contributions using embedded scaffolds like "My idea", "A better idea", "New information", "This idea cannot explain", "I need to understand", and "Put out knowledge together". These scaffolds facilitated the argumentation process (see Figure 1). Conversely, the control group discussed writing ideas and sources only in pairs during class sessions. Finally, in week 8, both groups wrote argumentative essays on the SSI topic.

**Figure 1**  
*Snapshot of the Collaborative Argumentation Environment (Knowledge Forum, KF)*



## Instruments

The impact of various environments on university students' English argumentative writing was analyzed through pretest and post-test essays, using Kathpalia and See's (2016) writing model, adapted from Toulmin's (1958). Essays were classified into five levels (1 to 5) based on the clarity of claims, supporting evidence, and rebuttals. Higher levels indicate better organization. Quality assessment focused on claims, grounds, and rebuttals. Weak claims received 1 point, while strong claims received 2 points. Grounds were categorized into five sub-categories: no evidence (0), faulty evidence (1), personal opinion only (2), attribute only (3), and both attribute and personal opinion (4), with higher scores for essays providing both objective and subjective justifications. Rebuttals were evaluated based on the inclusion of counter-arguments and rebuttals, with points assigned accordingly (0, 1, or 2).

## Data collection and analysis

This study applied Kathpalia and See's (2016) coding scheme to assess argumentative essays in both the pretest and post-test phases. The evaluation focused on structural elements and overall essay quality. The authors and a fellow EFL lecturer independently coded and assessed essays. Inter-coder reliability exceeded 0.8, indicating consistent interpretation. Discrepancies were resolved through rigorous deliberation. The chi-square test analyzed data for significant relationships between categorical variables.

## Major findings

### Students' argumentative essay structure

This study categorized students' argumentative essay structure into lower, intermediate, and higher levels. Chi-square tests were conducted in pretest and post-test to compare students' proportions in the two environments. Results (see Table 1) showed no significant difference in the pretest ( $\chi^2 = 0.42$ ,  $p > 0.05$ ), indicating similar argumentation structure levels before the interventions. In the collaborative argumentation environment, 80% achieved a higher level, while in the traditional face-to-face discussion environment, 77% achieved a higher level, with 17% at an intermediate level. No significant difference was found in the post-test ( $\chi^2 = 0.20$ ,  $p > 0.05$ ), suggesting both environments benefited students' essay structure.

**Table 1**  
*Comparisons on the Students' Argumentative Essay Structure*

	Structure level			$\chi^2$ value	
	Group	Lower (n, %)	Intermediate (n, %)		Higher (n, %)
Pretest	Control group	10 (33%)	12 (40%)	8 (27%)	

	Experimental group	10 (40%)	8 (32%)	7 (28%)	0.42 (n.s.)
Post-test	Control group	2 (6%)	5 (17%)	23 (77%)	0.20
	Experimental group	1 (4%)	4 (16%)	20 (80%)	(n.s.)

Note. n.s.: non-significant

### Students' argumentative essay quality

To assess the impact of collaborative argumentation on students' essay writing, we evaluated and compared the quality of argumentative essays from two groups. Essays were analyzed based on claim and ground quality and the strength of rebuttals. Chi-square tests were conducted to compare proportions between the pretest and post-test groups. Results (see Table 2) indicate no significant differences in claim quality between groups in either the pretest ( $\chi^2 = 0.18$ ,  $p > 0.05$ ) or post-test ( $\chi^2 = 2.34$ ,  $p > 0.05$ ). Notably, most students in both groups—77% in control and 92% in experimental—generated strong claims.

**Table 2**

*Group Comparisons on the Quality of Claims in the Three Argumentative Essays*

	Structure level		Weak Claim	Strong Claim	$\chi^2$
	Group				
Pretest	Control group		10(33%)	20(67%)	0.18 (n.s.)
	Experimental group		7(28%)	18(72%)	
Post-test	Control group		7(23%)	23(77%)	2.34 (n.s.)
	Experimental group		2(8%)	23(92%)	

n.s.: non-significant

Table 3 shows no significant differences between the two groups in pre-test ground quality ( $\chi^2 = 1.38$ ,  $p > 0.05$ ), indicating similar proficiency levels. However, in the post-test, significant differences emerged ( $\chi^2 = 16.5$ ,  $p < 0.05$ ). 76% of students in the collaborative argumentation group utilized attributes and personal grounds, enhancing persuasiveness. In contrast, only 23% of students in the traditional group did the same. Thus, collaborative argumentation is more effective in improving ground quality in university essays than traditional methods.

**Table 3**

*Group Comparisons on the Quality of Grounds in the Three Argumentative Essays*

	Grounds				$\chi^2$
	Group				
	Faulty evidence	Personal only	Attribute only	Attribute and personal	
Pretest	Control group				1.38 (n.s.)
	Experimental group				
Post-test	Control group				16.5*
	Experimental group				

n.s.: non-significant; \* $p < 0.05$

Table 4 shows no significant differences initially between the two groups rebuttal quality ( $\chi^2 = 0.04$ ,  $p < 0.05$ ). However, a notable contrast emerged later ( $\chi^2 = 4.28$ ,  $p < 0.05$ ), akin to differences in argument quality. About 80% of students in collaborative argumentation demonstrated improved counterargument usage, enhancing their persuasive abilities. Conversely, only 53% of those in traditional instruction effectively employed rebuttals. This suggests that incorporating collaborative argumentation benefits EFL university students in enhancing essay rebuttal quality.

**Table 4**

*Group Comparisons on the Quality of Rebuttals in the Three Argumentative Essays*



	Group	Grounds	Weak Rebuttals	Strong Rebuttals	$\chi^2$
Pretest	Control group		28(93%)	2(7%)	0.04 (n.s.)
	Experimental group		23(92%)	2(8%)	
Post-test	Control group		14(47%)	16(53%)	4.28*
	Experimental group		5(20%)	20(80%)	

n.s.: non-significant; \* $p < 0.05$

## Discussion

This study examined the impact of collaborative argumentative learning on university students' English writing by analyzing pre- and post-test essays. Both experimental and control groups improved their essay structure post-test, supporting Osborne et al. (2014) on the importance of explicit argumentation instruction. More experimental group students (80%) reached advanced levels compared to control (77%), likely due to autonomy and exposure to diverse ideas (Kathpalia & See, 2016). Quality, based on claims, grounds, and rebuttals, improved post-test. Experimental group essays had higher standard evidence (60%) versus control (30%) and excelled in rebuttals (91% vs. 76%). These findings underscore the benefits of collaborative discourse and a knowledge-building environment (Scardamalia, 2002).

## Conclusion and limitations

This study integrates knowledge-building (KB) pedagogy into a collaborative argumentation environment for university-level EFL students, improving essay quality, especially in grounds and rebuttals. The findings provide insights for researchers and educators to enhance collaborative argumentation environments. However, the study's context-specific nature limits its broader applicability. Further research is needed to explore the relationship between argumentative discussions and essay writing and to assess the long-term impacts of this collaborative approach.

## References

- Azimah, A., & Marlizayati, M. (2023). Exploring the effectiveness of collaborative argumentative writing in EFL classrooms. *Journal of Language Teaching and Research*, 14(1), 23-37.
- Gabinete, A. (2021). EFL students' argumentative writing: Challenges and pedagogical implications. *Asian EFL Journal*, 28(3), 67-89.
- Kathpalia, S. S., & See, E. K. (2016). Developing argumentation skills in an EFL setting: The case of Toulmin's model. *System*, 61, 13-26.
- Liu, Y., & Stapleton, P. (2020). Connecting past and present: A cross-sectional study of argumentative writing in Chinese EFL students. *Journal of English for Academic Purposes*, 46, 100872.
- Matos, L. (2021). Argumentative writing proficiency among EFL learners: An overview. *Journal of Applied Linguistics*, 6(1), 89-108.
- Mitchell, S. (2000). *Writing to learn across the curriculum*. Pearson.
- Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2014). Learning to argue and arguing to learn: Argumentative discourse as a vehicle for higher-level thinking in science classrooms. *Journal of Research in Science Teaching*, 51(8), 992-1023.
- Scardamalia, M. (2002). *Collective cognitive responsibility for the advancement of knowledge*. In B. Smith (Ed.), *Liberal Education in a Knowledge Society* (pp. 67-98). Open Court.
- Scardamalia, M., & Bereiter, C. (2003). *Knowledge building*. In J. W. Guthrie (Ed.), *Encyclopedia of Education* (2nd ed., pp. 1370-1373). Macmillan.
- Wu, Y. C., & Tsai, C. C. (2011). The effect of individual and group argumentation on collaborative discourse. *Educational Technology & Society*, 14(2), 89-99.

## Acknowledgements

Funding of this research work was supported by the National Science and Technology Council, Taiwan, under grant numbers NSTC 111-2410-H-A49-024 and NSTC 113-2914-I-008-006-A1.

# Using Pose Data to Explore Changes in Students' Self-Efficacy and Community Belonging in Makerspaces

Bertrand Schneider, Harvard University, [bertrand\\_schneider@gse.harvard.edu](mailto:bertrand_schneider@gse.harvard.edu)  
Edwin Chng, Harvard University, [chng\\_weimingedwin@g.harvard.edu](mailto:chng_weimingedwin@g.harvard.edu)

**Abstract:** Constructionist learning (Papert, 1980) can be a powerful way of fostering students' self-efficacy and their connection to a community. Makerspaces and digital fabrication labs epitomize this philosophy, by encouraging community-based, project-oriented learning through the creation of both physical and digital products. Because makerspaces are open-ended, messy learning environments, understanding why some students have transformative experiences, while others do not, remains a challenge. In this paper, we explore how multimodal learning analytics can provide a complementary lens to quantify and visualize students' learning trajectories. We collected students' 3D poses from video data collected during a semester-long course and used unsupervised machine learning approaches to identify behavioral states. We present preliminary correlational results with students' reports of self-efficacy and community belonging. We discuss these findings and how they could be used to support constructionist learning and teaching.

## Introduction and literature review

One of the most compelling aspects of open-ended learning environments such as makerspaces and digital fabrication labs is that they integrate formal STEM concepts with social emotional and 21st century skills (Dede, 2010). In these environments, students learn complex technical skills in teams, create unique artifacts, and develop self-directed, creative solutions to problems in their communities. However, capturing evidence of 21st century skills development remains a challenge in makerspaces. In a review, Timotheou & Ioannou (2019) reported that 88% of researchers relied on either qualitative or mixed methods, even though "we now have enough evidence of the value of computation making, allowing for scaling-up the impact and measurement via quantitative studies" (2019, p. 227). The contribution of this paper is to explore whether quantitative measures from pose data can be used to study self-efficacy and community belonging in makerspaces. For example, the time spent in the makerspace using various machines can provide indicators of mastery, and social interactions with peers and facilitators can provide a proxy for social persuasion, vicarious experiences, or a feeling of belonging to the community. In the sections below, we describe our research questions, the setting for data collection, and our data collection and processing pipeline. We conduct a preliminary data analysis to test whether pose-derived metrics can be related to self-efficacy and community belonging in a semester-long course using correlational methods.

## Methods

### Participants

Participants were 18 students from a course on digital fabrication at a graduate school of education of a private U.S. university in the spring of 2022. Students were between 25 and 35 years of age (1 male, 17 female). Two students had some prior experience with digital fabrication, 16 of them had never been in a makerspace before. Due to an oversubscription to the course during enrolment period, students were chosen for acceptance into the course based on a central lottery organized by the school.

### Procedure

Data collection took place during a semester-long course on digital fabrication. Students learned about laser cutting, 3D printing, vinyl cutting, sewing, electronics, micro-controllers, and various sensing technologies. 18 students were enrolled in the course, which is a typical size for this kind of hands-on course on digital fabrication. During the first half of the semester, each week involved a mini project that taught students a specific tool. The second half of the semester focused on a final project, where they were tasked to design a learning toolkit or environment of their choice using the tools of the makerspace. Students were required to write a blog post about the work they did in the makerspace every week and were encouraged to comment on their classmates' posts.

## Dependent variables

The data collected for this study includes pre- and post-surveys and periodic surveys. Pre-, mid-semester, and post-surveys used validated instruments to measure student self-efficacy (General Self-Efficacy Scale; Schwarzer & Jerusalem, 1995) and community belonging (Brief Sense of Community Scale; Peterson et al., 2008). Additionally, every week students self-reported their level of enjoyment and stress on a 1-5 Likert scale when submitting their project description. For the scope of this paper, we look at the gain from pre to post-test for self-efficacy and community belonging, and the average enjoyment / stress for the entire semester.

## Data collection and processing

The makerspace is instrumented with 8 high resolution cameras that capture synchronized video recordings during the operating hours of the makerspace (Fig. 1). We collected video data over 14 weeks, which resulted in approximately 10TB of video. The video recordings were first used to extract 2D pose data (xy location of 17 body joints) using AlphaPose (Fang & al., 2022).

**Figure 1**

*The Makerspace is Equipped with 8 High Resolution Cameras that Can Be Processed to Extract 3D Poses*



We captured the data 15 times per second using 8 cameras over 93 days, which resulted in 810'120'120 observations (i.e., rows of data in csv files). The 2D pose data was then processed to reconstruct 3D coordinates of body joints using poseconnect (Quinn, 2023). After this step, the dataset shrunk to ~100'000'000 observations. These observations were grouped into 96,188 “tracks”, which represents a person who was continuously tracked in the space and assigned a unique identifier. We manually labeled 36,081 (37.51%) pictures of makerspace users to train a person re-identification algorithm. The accuracy of the algorithm on seen data was 94.41%, and the accuracy on unseen data was 92.26% (tested on a sample of 2,493 tracks). The data was then downsampled for each second. Our final dataset comprised 6'751'001 observations of the x,y,z coordinates of 17 body joints (nose, left\_eye, right\_eye, left\_ear, right\_ear, left\_shoulder, right\_shoulder, left\_elbow, right\_elbow, left\_wrist, right\_wrist, left\_hip, right\_hip, left\_knee, right\_knee, left\_ankle, right\_ankle) with a timestamp and the predicted identity of the person. We included data from the 18 students enrolled in the course, 1 lab manager, 2 Teaching Fellows, and 2 instructors. We augmented our dataset with joint angles (neck, right / left shoulders, elbows, hips, knees) and a “movement” measure, which captures the distance between two consecutive observations from the same participant. Our final dataset contained 17 xyz coordinates, 9 joint angles, and 1 movement measure.

## Results

To analyze our dataset, we conducted a correlational analysis with simple features from our 3D pose data. We then used unsupervised machine learning algorithms on these features to find prototypical states (inspired by Schneider & Blikstein, 2015). We looked at how much time each student spent in these states and correlated these measures with our dependent variables.

### Correlations with simple features


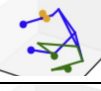


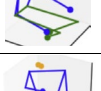
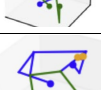

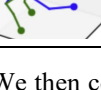
There was no significant correlation between our four dependent variables (self-efficacy gain, community belonging gain, stress, enjoyment;  $p > 0.1$ ), which suggest that these measures are capturing different constructs. We then computed simple features from our dataset: time spent in the space, time spent in different areas (e.g., collaboration space, by the laser cutter, 3D printer, etc.), time spent interacting with others (i.e., close proximity with another makerspace user), how much each person moved in the space, their average joint angles, and so on.

We found no significant correlations with self-efficacy and only one significant correlation with community belonging (the amount of time spent by the 3D printers, which is close to the entrance of the makerspace and was often used as a “social” space:  $r(16) = 0.53, p < 0.05$ ). This is most likely due to the inadequacy of *overall* correlations to 1) uncover finer-grained relationships that exist for different groups of student features, and 2) unpack the temporal aspects of these student features. For these reasons, we further analyzed our dataset with more sophisticated methods, such as unsupervised machine learning algorithms (to discern between different groups of student features).

### Correlations with cluster centroids

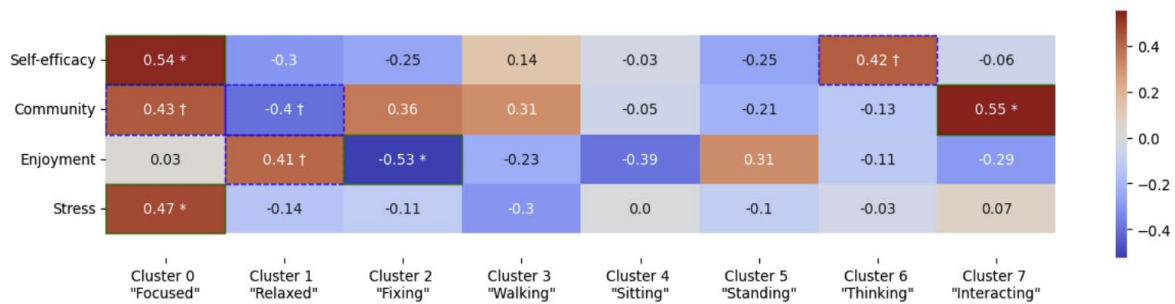
Before clustering, we used a z-score normalization to standardize features and constructs since they are on different scales. Because of the considerable size of the dataset, we ran a mini-batch k-means algorithm to find cluster centroids based on the nose xyz coordinates, the 9 joint angles, and the movement measure. A silhouette analysis suggests that the best number of clusters was 8. The resulting 8 cluster centroids are shown in Table 1. While we added labels to the centroids (based on their 3D posture appearance), they should *not* be taken as “truth”; these labels are merely suggestive of the students’ state and are provided to facilitate interpretation of the results. We provide a description of the clusters in Table 1 below.

**Table 1**  
*Description of the Main Clusters Identified from the Pose Data*

Cluster Centroid	Data	Correlates with (* $p < 0.05$ , † $p < 0.01$ )	Interpretation	Proposed label
	Low nose_y, low movement, low joint angles	* Self-efficacy (+) * Stress (+) † Community (+)	Posture shows a sitting position, still, with joints slightly bent. Correlations with stress and self-efficacy suggest a “focused” state.	“Focused”
	High knee / hip angles (bent), low movement	† Community (-) † Enjoyment (+)	Sitting position with legs bent, suggesting a “relaxed” position	“Relaxed”
	High nose_y, low movement	* Enjoyment (-)	Standing position, with arms forward, and low movement; negative correlation with enjoyment suggests “fixing” something	“Fixing”
	High nose_y, high movement	None	Posture shows a standing position with high movement, suggesting a person “walking” around the space	“Walking”
	Average values overall	None	Posture has average values of a person sitting at a table	“Sitting”
	Average values overall	None	Posture has average values of a person standing in the space	“Standing”
	Low nose_y coordinate (sit), low movement	† Self-efficacy (+)	Posture shows a sitting position with low movement, possibly with hand touching the face; marginal correlation with self-efficacy could suggest a “thinking” state	“Thinking”
	Low movement, high joint angles	* Community (+)	Posture shows a person with straight arms and legs, correlation with community belonging suggests that g” the student is “interacting” with others	“Interacting”

We then computed the ratio of time spent in each cluster for each student. We computed Pearson’s correlations between these measures and our dependent variables (Fig. 2). We found that spending more time in the “Focused” cluster was significantly correlated with gains in self-efficacy, but also with stress; additionally, spending more time in the “Interacting” cluster was significantly correlated with community belonging. Being in the “Fixing” cluster was negatively correlated with enjoyment. Finally, there are other marginally significant relationships (indicated as †  $p < 0.01$  on Fig. 2) that could potentially be significant with a larger sample size, but we leave it to the reader to hedge their interpretation of these results.

**Figure 2**  
*Correlations between Cluster Centroids and Dependent Variables (\* indicates  $p < 0.05$ , †  $p < 0.01$ ).*



## Discussion

In this paper we explored a large dataset of pose data collected during a semester-long course on digital fabrication and we used multimodal learning analytics methods as an additional tool to measure and explore the learning paths of students. We found that simple pose features were not very informative (possibly because of their inability to reflect finer-grained relationships that exist for different student features and temporal details) but using unsupervised machine learning methods revealed some prototypical states that could indicate self-efficacy growth and becoming part of a community. In particular, we found that spending more time in a “Focused” state was correlated with gains in self-efficacy, but also with stress; spending more time in the “Interacting” cluster was correlated with community belonging. Additionally, staying in a “Focused” state, moving from a “Relaxed” to “Thinking” state, and from “Walking” to “Focused” were all correlated with self-efficacy gains, which could indicate pathways to increased self-efficacy. On the other hand, moving from “Focused” to “Fixing” or “Sitting”, from “Interacting” or “Sitting” to “Fixing”, or from “Thinking” to “Walking” were all negatively associated with self-efficacy gains. These results suggest that being focused on or thinking about a project might put students on a path toward feeling a sense of mastery in the makerspace (which has been found to contribute to self-efficacy in prior work; Bandura, 2001; Chng, Seyam, Yao & Schneider, 2022). While these findings are just a first step and need to be replicated, they show promise in identifying meaningful student states in makerspace environments and potential for the use of quantitative measures to understand student learning trajectories in open-ended learning environments.

## References

- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.
- Chng, E., Seyam, M.R., Yao, W. & Schneider, B. (2022). Toward capturing divergent collaboration in makerspaces using motion sensors. *Information and Learning Sciences*, Vol. 123, No. 5/6, pp. 276-297
- Dede, C. (2010). Comparing frameworks for 21st century skills. *21st century skills: Rethinking how students learn*, 20(2010), 51-76.
- Fang, H. S., Li, J., Tang, H., Xu, C., Zhu, H., Xiu, Y., & Lu, C. (2022). Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic books, 2020.
- Peterson, N. A., Speer, P. W., & McMillan, D. W. (2008). Validation of a brief sense of community scale: Confirmation of the principal theory of sense of community. *Journal of community psychology*, 36(1), 61-73.
- Quinn, T. (2023). Poseconnect. <https://github.com/WildflowerSchools/poseconnect>.
- Schneider, B., & Blikstein, P. (2015). Unraveling Students’ Interaction Around a Tangible Interface using Multimodal Learning Analytics. *International Journal of Educational Data Mining*, 7(3), 89-116.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user’s portfolio. *Causal and control beliefs*, 35, 37.
- Timotheou, S., & Ioannou, A. (2019). On making, tinkering, coding and play for learning: A review of current research. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part II 17* (pp. 217-232). Springer Publishing.

## Toward an Assessment of Students' (Social) Computational Literacy

Michael Foster, Rochester Institute of Technology, [mrfcos@rit.edu](mailto:mrfcos@rit.edu)  
Matthew Dunham, CSU Monterey Bay, [mdunham@csUMB.edu](mailto:mdunham@csUMB.edu)  
Christian Cammarota, Rochester Institute of Technology, [cmccos@rit.edu](mailto:cmccos@rit.edu)  
Mike Verostek, University of Rochester, [mveroste@ur.rochester.edu](mailto:mveroste@ur.rochester.edu)  
Ben Zwickl, Rochester Institute of Technology, [ben.zwickl@rit.edu](mailto:ben.zwickl@rit.edu)  
Tony E. Wong, Rochester Institute of Technology, [aewsma@rit.edu](mailto:aewsma@rit.edu)

**Abstract:** Computational tools, particularly computer programming and statistical software, have enabled new forms of computational course assignments using tools such as Jupyter Notebooks and Minitab. With computation becoming ubiquitous within STEM, and society writ large, computational literacy – or the skills, knowledge, and beliefs students need in order to use computation – is an important construct to assess. Indeed, there is a need for an assessment rubric that can be used across multiple computational tools to better understand students' performances with these tools. In this study, we focused on the development of an assessment rubric for students' social computational literacy. Building off Odden et al.'s (2019) framework, we analyzed three students' artifacts and interviews for social practices, knowledge, and beliefs. We present the beginning phase of an assessment rubric identifying social computational literacy through students' social knowledge and beliefs.

### Introduction

Computation, which we take to include the use of computers or programming tools to solve problems and draw conclusions, plays a significant role in research and industrial professions. As such, there is a need to better understand learning from, and through, computation. Indeed, several projects have sought to explore the use of computational thinking in STEM education (Weintrop et al., 2016). Further research in this field is vital to ensure that students obtain the necessary computational skills to pursue a wide range of occupations within STEM that use computation.

Assessing students' progression towards mastering computational practices through their use of computational tools (e.g., object-oriented programming and menu-driven statistical software) is of acute interest in education. Indeed, in their review of the literature on assessment of computational thinking, Tang et al. (2020) concluded that there is a need for assessment items that can be used across computational tools (e.g., Python or Minitab) and devices to better compare students' performances across different contexts. As such, the goal of this project is to operationalize a framework for the assessment of students' *computational literacy* (defined below). In particular, we are interested in the identification and assessment of how students communicate their use of computational tools. As we integrate computational tools in collaborative environments, assessment of students' communication through and with computational tools warrants further inquiry. Thus, the goal of this study is to begin operationalizing a rubric for the assessment of the *social* elements of students' computational literacy.

### Computational literacy

One way in which computation has been conceptualized within education is through the lens of *computational literacy*. Introduced by diSessa (2001), computational literacy (CL) draws parallels between written/verbal literacy and computation. DiSessa (2018) defined literacy as the “adoption by a broad cultural group ... of a particular infrastructural representational form for supporting intellectual activities” (p. 4). Our civilization is in the process of adopting computational tools as both an “infrastructural” and a “representational form for supporting intellectual activities” (p. 4) ranging from research in STEM to business.

This trend has continued, and grown, since diSessa' coined CL, as evidenced by the increased emphasis within curriculum development on *computational thinking* (Weintrop et al., 2016). Computational thinking (CT) is a construct that captures what it means to think “like a computer scientist” (Wing, 2006, p. 34). Since Wing's initial formulation, CT has served as an important lens into students' development of algorithmic thinking and been broadened to include a myriad of other skills associated with computation (Weintrop et al., 2016).

While CT is a productive framework for thinking about students' uses of computational tools, we follow diSessa's use of CL, in part, because of the explicit inclusion of a social lens in a literacy (diSessa, 2001). In defining what constitutes a literacy, diSessa identified three pillars relevant to all forms of literacy: social,

material, and cognitive. Each of these pillars offers a powerful lens into students' use of computational tools. In particular, social CL (SCL), which refers to the communication of, with, and about one's use of computation, guides our identification and assessment of students' communication through and with computational tools.

Odden et al. (2019) expanded on diSessa's CL pillars in the context of undergraduate physics and identified three CL sub-pillars: practices, knowledge, and beliefs. CL practices relate primarily to observable actions while programming. For example, social practices include the presentation of a model. CL knowledge relates to the conceptual understandings required for practices to emerge. For example, social knowledge includes an awareness of effective communication strategies (e.g., how to create a research report presenting computational findings). CL beliefs refer to the attitudes/feelings about the use of a computational tool. For example, social beliefs include one's attitude toward how code should be displayed and communicated.

## Methods

This investigation focuses on three students who were recruited from a university in the northeastern United States. We sought out multiple undergraduate STEM courses in which a computational tool was being used. One pre-freshmen student, Caleb, was recruited from a general science course during the summer before his first semester and two senior mathematics students, Jen and James, were recruited from an upper division mathematics course on stochastic processes. From each participant, one homework item (henceforth, referred to as an artifact) was collected and a semi-structured interview was conducted. The artifact from these courses was a Python assignment in a Jupyter Notebook, which is a dynamic electronic notebook that integrates written explanations and problem statements with computer code, output, and mathematical expressions). During the interview, students were asked about their background (e.g., "What prior experiences do you have with computation tools?"), and they were asked guided reflection questions about their artifact (e.g., "How did you think about this task?").

Analysis was a two-phase process. During phase 1, we expanded upon Odden et al.'s (2019) framework of physics CL to code the participants' artifacts and interviews for SCL. Drawing on CL and CT literature, we identified a number of practices that related to SCL. Important social practices include presentation of results from a computational model and the use of formatting. Presentation of results includes creating visualizations or the production of written/oral reports and the use of formatting includes the use of logical structure within the documentation of one's use of a computational tool (Gambrell & Brewe, 2023; Odden et al., 2019; Weintrop et al., 2016). Importantly, these practices were identified as broadly applicable to varying computational tools and across contexts. We expanded social knowledge beyond Odden and colleagues' definition to capture an explicit awareness of the practices and the purpose of these practices as a means/strategy of communication. Similarly, we identified social beliefs when the participants expressed an attitude or feeling about the use of social practices.

Phase 2 is where our rubric began to take shape. Using the forms of SCL identified in Phase 1, a dichotomous scale was created to assess students' SCL as either proficient or emergent. To identify what constitutes evidence of proficient or emergent SCL, we turned to the literature on verbal and written literacies. Whitehurst and Lonigan (1998) define emergent literacy as "the skills, knowledge, and attitudes that are presumed to be the developmental precursors to conventional forms of reading and writing" (p. 849). Analogously, we take SCL to be evidenced as emergent when precursors to the generally accepted forms of social practices, knowledge, and beliefs are present. Importantly, what is "conventional" and what is a "precursor" is subjective and context dependent. As such, the practices we identified in the literature served as conventional, or proficient, forms of SCL within our rubric. When the participants used, had an awareness of, or expressed an attitude about these forms of communication, then they were assessed as evidencing proficient SCL. Emergent SCL was then evidenced through social practices, knowledge, or beliefs that were unused, used for nonstandard purposes, or exhibited negative attitudes. Future implementation of our assessment rubric may require adaptation depending on what is important to the context of its use.

## Results

The results section is broken into two subsections, covering both the knowledge and beliefs sub-pillars. Given the scope of this proceeding and the field's previous work identifying practices (e.g., Weintrop et al., 2016), we do not explicitly illustrate social practices here. Within each subsection we illustrate how our rubric can be used for assessing instances of proficient or emergent SCL.

### Knowledge

Social knowledge is primarily evidenced through social practices (i.e., what we know is evidenced in what we do). As such, the identification/assessment of social practices plays a role in the assessment of social knowledge. Additionally, knowledge can also be evidenced through students' reflections about communication techniques.

Caleb, the participant from the pre-freshman STEM course, evidenced proficient SCL through his social knowledge during his interview. Specifically, Caleb expressed an awareness that his use of structure and commenting in his code was for communicative purposes. When asked about the organization of his code, Caleb said that he organized his code because “[the code] started to look nice.” Similarly, Caleb said that “[the comments are] so I don’t get confused. If I look back on it, I know what’s going on. It’s just keep[ing] track.” In both responses, Caleb identified that his social practices were a part of communicating his code. Communication to his future self to help “not get confused” is an important metacognitive skill, and the emphasis on the readability for himself improves the readability, and thus the communication, for others. Thus, Caleb’s awareness of the purpose of organization and comments to aid in the communication of one’s code evidenced proficient SCL.

Similarly, Jen, a participant from the upper-division class using Jupyter Notebooks, evidenced proficient SCL knowledge while explaining her code for a stochastic simulation. Specifically, Jen evidenced awareness and an understanding of communicating through comments. Figure 1 contains a segment of Jen’s work where she defined what her variables represented and a block of her code where she used comments to document what each line of code represented. During her interview, Jen’s social knowledge became explicit when asked about her use of comments. Jen stated, “I wanted to be more organized ... I tried to use a bunch of comments to make it clear what I was doing.” Later, she reflected that her comments and variables help “keep it very straightforward just to be able to look back on it.” Here, Jen articulated a social knowledge that comments help with organization and for the readability of code. Thus, Jen’s awareness of the purpose of commenting for organization and readability were forms of social knowledge that evidenced proficient SCL.

**Figure 1**

*Two Segments from Jen’s Jupyter Notebook Documenting Her Variable Names*

```
If we let  $D_{i,j}$  be the driving distance from location  $i$  to  $j$  and our locations are  $R$ : intersection of
Jefferson and John St.,  $S$ : S lot,  $F$ : F lot, and  $H$ : H lot.
If we let  $S$  be the time to drive around the lot to find a spot we can approximate the time to
follow the distribution  $S \sim \text{norm}(2, 0.5)$ 
else: #now you have to drive to F Lot
    time+=np.random.normal(5,2) #D_{S,F}
    time+=np.random.normal(2,0.5) #S
```

In contrast, James, another participant from the upper-division class using Jupyter Notebooks, evidenced emergent SCL knowledge. While working on programming a stochastic simulation, James’ code contained a single comment (Fig. 2). In this comment, James indicated that he made a poor choice in his variable assignment. Importantly, James’ comment does not serve to communicate his use of his computational tool, rather it served to communicate his awareness of his own mistake. During the interview, when asked the purpose of commenting, James reflected, “I’ll be honest, I usually don’t comment nearly as much as I should. So, it [commenting] was more for grading purposes.” Through James’ reflection, he evidenced an awareness that he should comment, but his reason for commenting indicated that his comments were not for the communication of his code and were for an auxiliary purpose (i.e., grading). In summary, James evidenced social knowledge of commenting as a communication tool, but his inattentiveness to purpose of commenting to aid in the communication of his code evidenced emergent SCL.

**Figure 2**

*James’ Comment Describing an Issue with His Chosen Variable Name*

```
Df = np.random.choice(Df_choices, p=Df_probs) # I realized
    that my variable assignment cause this one to be Note
    that this is not a dataframe object.
```

## Beliefs

Social beliefs include one’s attitudes and feelings about social practices. For example, Jen and Caleb evidenced proficient SCL through the social beliefs identified within their interviews. Toward the end of her interview, Jen reflected that her practice of using comments, which she used on every single line of her code, was important for organization. When asked by the interviewer about what she learned from this assignment, Jen stated, “keeping everything organized was helpful.” Similarly, Caleb’s social beliefs could be identified through a convergence in his knowledge and practices, which both evidenced proficient SCL. Caleb evidenced proficient SCL through his social practice of using meaningful variable names. Similarly, Caleb expressed a social knowledge about the use of meaningful variables names and a social belief that it is important to not use random or arbitrary variables in



code. Specifically, during the interview Caleb reflected, “I just put [the variable name] as what it is. If you find something random, like a single letter or something, it gets kind of confusing.” In this response, Caleb expressed an attitude that it is important to use meaningful variable names to avoid confusion. Combined with his practice of using meaningful variable names, this belief evidenced proficient SCL.

James, on the other hand, evidenced emergent SCL through his social beliefs. James lone comment (Fig. 2) illustrated his infrequent use of this social practice, but his social knowledge showed he knew he “should” comment more frequently. In fact, James continually evidenced an awareness that his social practices were lacking. When James was asked by the interviewer about his formatting of an array he acknowledged his lack of formatting when he stated, “Honestly, I was sloppy here. Usually, [the professor] would format it so that ... it would be even and stuff like that.” Similarly, when probed further about his comments, James stated, “when it comes to like homework assignments, I’m usually pretty on my game with commenting and stuff. But in terms of personal stuff, I could definitely be a little better about it.” In this way, James evidenced a knowledge of social practices, but his knowledge did not align with his practices. First, he knew that his unformatted array was sloppy, and that commenting is important, but he also acknowledged that he does not enact that knowledge. As such, there is a disconnect between his knowledge and practice. Possibly, if James had more robust social beliefs in the importance of the social practices that he has knowledge of, then he would have put that knowledge into practice. This tension, possibly a result of the short vs long-term benefits of commenting, is particularly evident in his comment (Fig. 2). While he evidenced knowledge about meaningful variable names, he did not put his knowledge into practice and change his variable name. As such, his social beliefs evidence emergent SCL.

## Discussion

This work provided an overview of the first phase of the design of a rubric for students’ CL. Focusing on the pillar of social CL, we illustrated how our participants evidenced proficient and emergent SCL through their knowledge and beliefs. This work offers a contribution by expanding upon Odden et al. (2019) beyond a physics context, through the inclusion of social practices (and the implied knowledge and beliefs) that are generalizable to non-programming-tools (e.g., statistical software) and practices that have been emphasized in CT. Furthermore, our work contributes to Tang et al.’s (2020) call for the creation of an assessment tool that can be applied across multiple forms of computational tools. Future work will begin to illustrate our rubric’s application to both programming and statistical software tools. Importantly, we acknowledge that the practices, knowledge, and beliefs we identified are context dependent. For example, the set of possible social practices available in a menu-driven environment are much smaller than the set of possible social practices available in a Jupyter Notebook. Similarly, the context of why a student is using a social practice also impacts assessment. For example, a student may be instructed to present results in a certain way, or they may be using comments or formatting because they believe their professor wants them too (as opposed to it being important for communication purposes). As such, we forward our rubric not as a tool to be directly applied, but rather adapted to each contextual social objective. Finally, we observe that the artifacts and interviews played complementary roles in identifying features of SCL. The interviews played an important role in eliciting participants’ beliefs and their knowledge. The artifacts, on the other hand, made practices explicit and social knowledge implicitly identifiable. While interviews were useful for assessing SCL, they are not practical in most instructional situations. Future work will consider how to frame reflection questions within homework to better identify social knowledge and beliefs within artifacts.

## References

- DiSessa, A. (2001). *Changing minds: computers, learning and literacy*. MIT Press.
- DiSessa, A. (2018). Computational literacy and “the big picture” concerning computers in mathematics education. *Mathematical Thinking and Learning*, 20(1), 3-31.
- Gambrell, J., & Brewes, E. (2023). Towards a Generalized Assessment of Computational Thinking for Introductory Physics Students. arXiv preprint arXiv:2308.03593.
- Odden, T. O. B., Lockwood, E., & Caballero, M. D. (2019). Physics computational literacy: An exploratory case study using computational essays. *Physical Review Physics Education Research*, 15(2), 020152. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020152>
- Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, 148, 103798.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.
- Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, 25(1), 127–147. <https://doi.org/10.1007/s10956-015-9581-5>

## Examine Student Resource Uses in a Game-based CSCL

Xiaotian Zou, Daeun Hong, Feng Chen, Cindy E. Hmelo-Silver  
xz107@iu.edu, dh37@iu.edu, carrfeng@iu.edu, chmelosi@indiana.edu  
Indiana University

Krista Glazewski, Tianshu Wang, Bradford Mott, James Lester  
kdglazew@ncsu.edu, twang43@ncsu.edu, bwmott@ncsu.edu, lester@ncsu.edu  
North Carolina State University

**Abstract:** This paper explores student resources utilization and its impact on student learning outcomes in a game-based CSCL context from a sociocultural perspective. Analysis of group-level resource utilization revealed three distinct patterns that differentiate high-performing from low-performing groups. Qualitative analysis of group problem-solving responses provided further evidence of the association between group-level resource use and subsequent group performance. The study sheds light on the association between resource utilization and learning success in collaborative learning settings, offering valuable guidance for enhancing the pedagogical design of learning resources and teacher scaffolding to promote effective resource utilization.

### Introduction

Sociocultural learning theories emphasize the significance of tools and artifacts in mediating learning (Danish & Gresalfi, 2018). In game-based CSCLs, these tools and artifacts often take the form of learning resources which may include video clips, visual representations, feedback from in-game characters, and digital notebooks designed to afford student learning opportunities (Jeong & Hmelo-Silver, 2010). Learning in these contexts may rely on how well students utilize available learning resources and tools important to the learning objectives (Jeong & Hmelo-Silver, 2010; Danish et al., 2022). However, CSCL environments tend to emphasize student agency and are characterized by student-centered inquiry, which necessitates that students effectively appropriate and engage with embedded resources relevant to learning and problem-solving goals (Law et al., 2021). Furthermore, the digital interactions and multimodal learning resources can introduce extra layers of complexity and challenges, and students may fail to engage deeply with the resources to achieve a sophisticated understanding of the materials (Jeong & Hmelo-Silver, 2010). Ultimately, these complexities pose challenges when teachers hope to monitor and facilitate effective resource uses within CSCL classrooms (Furberg, 2016). This study aims to explore measures of students' interactions with learning resources and utilization patterns crucial for collective learning achievements in resource-rich CSCL environments. We propose that productive resource use correlates with learning outcomes, reflected in student performances such as final assessments and group problem-solving activities. By investigating the relationship between resource utilization and knowledge acquisition in game-based CSCL settings, this study seeks to inform the design of automated tools to aid teachers and students in fostering more effective resource utilization in CSCLs.

### Tools and resources in collaborative learning

The sociocultural perspective highlights that learning is fundamentally a social process, in which individuals learn through observation, negotiation, social interactions, sense-making, and solving collective problems (Pea, 2004). Within social learning processes, tools and resources help people organize and regulate their goal-directed activities, allowing them to form new representations as a group (Danish & Gresalfi, 2018). In CSCLs, the process of knowledge co-construction involves individuals making sense of digital resources and representations and negotiating their understanding with others to collectively construct shared knowledge (Stahl & Suthers, 2014). However, due to the diversity of personal and cultural experiences learners bring to the learning contexts, their goals and approaches when engaging with various resources and tools may differ from each other and from the intentions of teachers or co-created objectives (Furberg & Ludvigsen, 2013). Therefore, researchers (e.g., Danish et al., 2022) interested in studying the affordances of resources and tools tend to ask: why a particular group of learners tend to use certain tools in certain ways. In this paper, we investigate the nature of learners' interaction with the available resources, and patterns of interactions that are crucial to both group problem-solving processes and final learning success. To these ends, the present study is guided by the following research questions (RQs): RQ1) In what ways do groups with different learning outcomes vary in their utilization of learning resources over

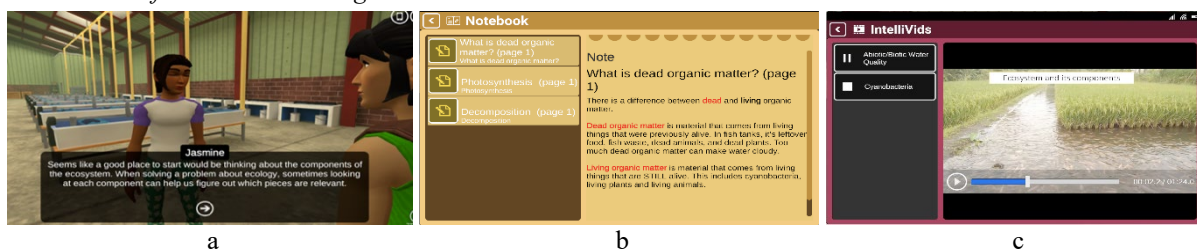
time? RQ2) How do different levels of resource utilization manifest in subsequent collaborative problem-solving activities?

## Methods

### Participants and context

The study took place in a science game-based learning environment: CRYSTAL ISLAND: ECOJOURNEYS (Figure 1), designed to teach students knowledge about aquatic ecosystems. In this research, we collected data from 156 middle school students from six science classrooms in midwestern and southern US. In the game, students worked together in small groups (n=3 or 4) to solve complex problems related to sick tilapia fish at a local farm in the Philippines. The game consists of a tutorial section followed by three quests focusing on interrelated problems. Students begin with solo investigation by gathering data, receiving notes, watching learning videos, and interacting with non-player characters (NPCs). Students also engage in collaborative activities known as *Deduce* and *TIDE* (see also Hong et al., 2023). At the end of both *Deduce* and *TIDE*, students respond to the questions collectively using the notes and data collected. A pre-test and a post-test were administered before and after the game. Group learning gain refers to student's average pre- and post-test percent difference. Three types of learning resources within ECOJOURNEYS are included: informational videos offered in each quest; excerpts of key concepts presented in notebook entries collected during individual investigation; and in-game NPCs that provide key information. Figure 1 displays the screenshots of each of these resources.

**Figure 1**  
Screenshots of In-Game Learning Resources within ECOJOURNEYS:



Note: a. In-game NPC providing information; b. Notes stored in the Tablet; c. Informational video in quest 3.

### Research design, data sources, and analysis

In this paper, we utilized two data types: log data recording human-computer interactions and students' collective responses to questions at the end of *TIDE* and *Deduce*. Using log data, we identified four variables indicative of how small groups interact with learning resources in ECOJOURNEYS: group average time spent with NPC characters, group average time spent reading notes during game, whether skip a video in each quest by group, and average number of notes revisited by a group. After excluding missing data, log files of 113 students (28 groups) was utilized in the study. For RQ1, we employed log data regarding group-level resource utilization of each quest to examine the pattern of change in their interaction with the resources across the game. We centered the time\_NPC and the time\_note variables around the average time and number of resource usage of each quest due to the differences among each quest. To compare cases, we selected 6 high-performing groups and 6 low-performing groups from the entire dataset. The high-performing groups consisted of groups with average pre-/post-gains above the 75th percentile, while the low-performing groups are below the 25th percentile. No statistically significant difference in pretest scores between the high-performing and the low-performing groups was confirmed by an independent t-test ( $p > .01$ ). To address RQ2, we qualitatively analyzed group responses to problem-solving questions in *Deduce* and *TIDE* from both high-performing and low-performing groups and identified and compared patterns in responses from each group.

## Results

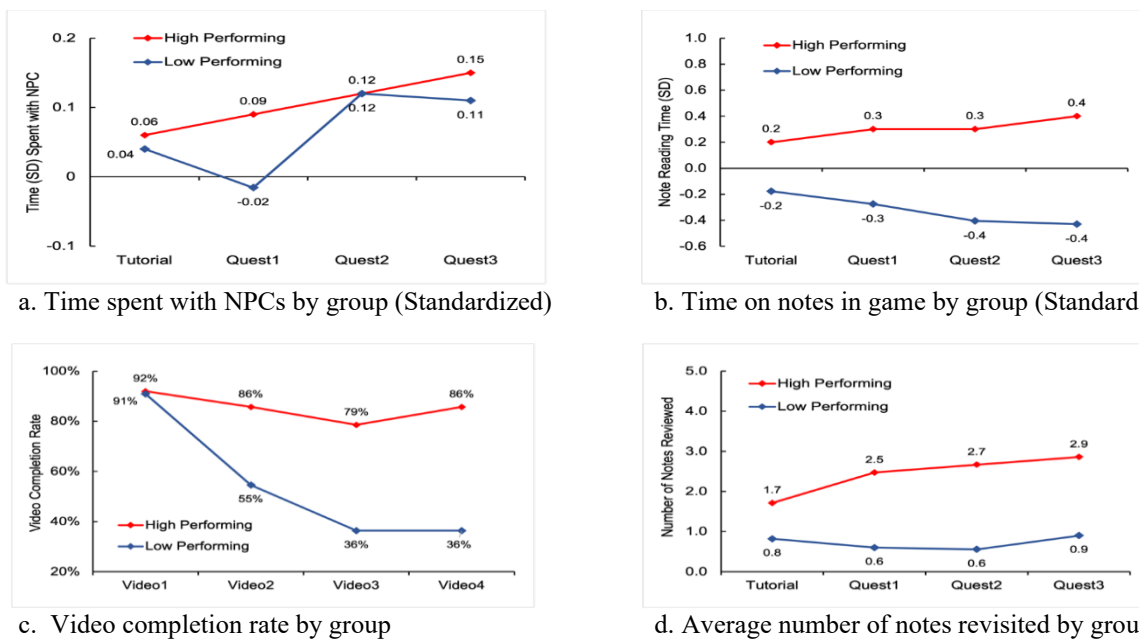
### RQ1: Patterns of group resource utilization over time

We explored how group-level resource utilization evolved over time and what patterns tended to be important to group learning success. Figure 2 presents four trajectories of resource utilization over time by the high-performing and the low-performing student groups. Overall, three patterns of resource utilization were observed that differentiated the high and low performing groups. First, high-performing groups seemed to consistently utilize resources over time. For example, while both groups are seen a similar level of interaction with NPCs in tutorial,

the low-performing group experienced a decline (-0.02 SD) in the time investment within quest 1, and the level of utilization went back up in the last two quests. Another pattern is that the high-performing group appeared to be more likely to complete required resources rather than skipping them or going through them quickly without completing them. For instance, while both groups have seen decline in video completion rates over time, a significant portion of the low-performing group began skipping required videos in quest one, with over half of them skipping videos in the last two quests. In comparison, the high-performing group maintained a consistently high completion rate throughout the game. The third pattern is the increased resources utilization over duration of the game in the high-performing group. For example, the high-performing group exhibited an upward trend in both the frequency of revisiting their notes (see Figure 2d) and the average time spent with the notes (see Figure 2b) as the game progressed. The increase in note revisits may be due to their growing need to reinforce knowledge as collaborative problems became more complex in later quests. They might have also become increasingly adept at navigating and locating resources as the game advanced. However, the low-performing group, despite participating in the activities in the same order, did not show a similar trend of resource utilization.

**Figure 2**

*Patterns of Learning Resource Use Over Time by High-Performing and Low-Performing Groups*



## RQ2: Group resource utilization in problem-solving

Next, we explored how different levels of learning resources utilization can manifest in students' problem-solving abilities. To investigate what might be the evidence of productive uses of resources, we qualitatively analyzed the constructed responses at the end of each *Deduce* and *TIDE* activities from both high-performing (HG) and low-performing groups (LG) and decided to present the sample responses in *Deduce* due to their representativeness (see Table 1). For example, when students were asked to justify the statement regarding the bacteria at the beach, we can observe how the two groups differ not only in the quality of their responses, but also how learning resources (see bold texts) were indexed in their answers. The absence of references to resources in the responses of low-performing groups is consistent with the patterns of their incomplete and inconsistent interactions with learning resources in this group, leading to their inability to use them. In addition, when asked to explain the “cloudiness” phenomenon in the water, the high-performing groups seemed to employ more essential components and related scientific concepts (see italic texts). Whereas the low-performing groups lack mentioning of scientific concepts or oversimplify the underlying mechanisms, even though the concepts and key components were covered in different learning resources (e.g., videos, notes) throughout the quest. This example provides further evidence for how different levels of interactions with learning resources would affect subsequent groups' problem-solving performances.

**Table 1**

*Sample Responses from High- and Low-Performing Groups in Deduce*

<p>HG2: “We know this because we tested a <b>sample in the lab</b>”</p> <p>HG3: “We as a team think that it is bacteria because <b>sulu's notes</b> explained the effects of bacteria in water”</p> <p>HG4: “It is bacteria because it matches the shape and color of the image in the <b>notebook</b>”</p> <p>HG6: “The water is becoming cloudy because of the excess of <i>cyanobacteria</i> and <i>dead organic matter</i>.”</p> <p>HG5: “There is bacteria and <i>dead organic matter</i> in the water making it cloudy.”</p>	<p>LG1: “I think there's bacteria at the beach because the boats traveling back and forth could've caused the green bacteria to appear”</p> <p>LG3: “They keep talking about bacteria”</p> <p>LG5: “It looks like bacteria.”</p> <p>LG2: “if more dead things are in the water it becomes more cloudy which make more deAd feish”</p> <p>LG3: “More <i>dead organic matter</i> = cloudy”</p> <p>LG6: “turbidity”</p>
--	--

## Discussion

In our study, we examined how learners utilize resources within a game-based collaborative learning environment and its impact on group learning outcomes. Drawing from sociocultural learning theories, we emphasize the significance of resource utilization in shaping learning experiences and collective knowledge building. Our analysis of group resource usage over time and problem-solving qualities suggests a link between resource engagement and group learning performance. However, there are limitations to consider, such as the lack of comprehensive understanding behind resource utilization patterns derived solely from log trace data. For instance, additional evidence is needed to elaborate on what motivated high-performing groups to increasingly revisit their notes, and what factors led low-performing groups to skip a greater number of videos.

To conclude, the findings of our study provide insights into how students use (or do not use) resources and how patterns of resource use influence their subsequent group performance in complex collaborative environments like game-based CSCL environments. Our research sheds light on the role of diverse resources in facilitating knowledge acquisition at group levels. To enhance resource utilization effectively, teacher guidance and support are crucial. Our findings can inform the development of tools for teachers to understand and improve student resource utilization. Additionally, student-facing tools can empower learners to monitor and reflect on their resource interactions during collaborative learning. Overall, this study emphasizes the importance of resource-driven factors in collaborative learning and suggests avenues for further research to optimize resource usage for better learning outcomes.

## References

- Jeong, H., & Hmelo-Silver, C. E. (2010). Productive use of learning resources in an online problem-based learning environment. *Computers in Human Behavior*, 26(1), 84-99.
- Danish, J. A., & Gresalfi, M. (2018). Cognitive and sociocultural perspective on learning: Tensions and synergy in the learning sciences. *International handbook of the learning sciences*, 34-43.
- Danish, J. A., Anton, G., Mathayas, N., Jen, T., Vickery, M., Lee, S., ... & Ryan, Z. (2022). Designing for shifting learning activities. *Journal of Applied Instructional Design*.
- Furberg, A. (2016). Teacher support in computer-supported lab work: Bridging the gap between lab experiments and students' conceptual understanding. *International Journal of Computer-Supported Collaborative Learning*, 11, 89-113.
- Furberg, A., Kluge, A., & Ludvigsen, S. (2013). Student sensemaking with science diagrams in a computer-based setting. *International Journal of Computer-Supported Collaborative Learning*, 8, 41-64.
- Hong, D., Zou, X., Wang, T., Hmelo-Silver, C. E., Glazewski, K., Uttamchandani, S., ... & Lester, J. (2023). Towards Understanding Collaborative Scientific Inquiry Practices in CSCL Classrooms With In-Game Data. International Society of the Learning Sciences.
- Law, N., Zhang, J., & Peppler, K. (2021). Sustainability and scalability of CSCL innovations. *International handbook of computer-supported collaborative learning*, 121-141.
- Ludvigsen, S., & Steier, R. (2019). Reflections and looking ahead for CSCL: Digital infrastructures, digital tools, and collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 14, 415-423.
- Stahl, G., Koschmann, T., & Suthers, D. D. (2014). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. pp. 479-500). NY: Cambridge University Press.

## Acknowledgments

This study was supported by US National Science Foundation DRL-1561486, DRL-1561655, and DRL- 2112635.

# Measuring Help-seeking in Online Course Discussion Forums with Privacy-preserving Large Language Models

Nigel Bosch, Destiny Williams-Dobosz, Michelle Perry  
pnb@illinois.edu, destiny7@illinois.edu, mperry@illinois.edu  
University of Illinois Urbana-Champaign

**Abstract:** Discussion forums in college courses play a vital role in enabling students to seek academic help from each other and from instructors, especially in online courses. However, it can also be difficult for instructors and researchers interested in help-seeking to identify when it occurs in large, active forums. This study examined large language models as measurement tools for automatically coding students' help-seeking forum posts, resulting in Spearman's *rho* up to .711 for the correlation between models' help-seeking ratings versus manually coding via an established schema. The large language model approach requires no problem-specific training data, enabling the help-seeking model to be constructed with minimal manual coding of forum posts compared to traditional machine learning approaches. Moreover, the models in this study are offline models, run on a typical laptop, which preserve students' privacy by eliminating the need to transmit forum data to a third-party service.

## Introduction

Many college courses include an online discussion forum component, which can serve to foster a sense of community and support peer learning among students (Fiock, 2020), particularly in fully online courses where students would otherwise have limited opportunities to discuss the course with each other. One of the critical activities that occurs on course discussion forums is academic help-seeking—i.e., when students indicate to the course community that they would like assistance with something relevant to the course (Fong et al., 2023). Previous work has studied the benefits of help-seeking in course discussion forums via strategies like manually coding forum posts according to the type of help-seeking they exhibit (Jay et al., 2020), which is valuable but time consuming and limited in terms of application to practice (i.e., instructors manually coding their forums). Other constructs, such as emotions, have been studied with the help of artificial intelligence (AI) tools that automate the coding process (Crossley et al., 2017). In this paper, we explore AI-driven analysis of help-seeking as well: specifically, recent large language models (LLMs). If successful, LLMs could provide opportunities to study and adapt to help-seeking at a much larger scale than manual efforts in course forum contexts.

LLMs work by learning to predict what word (or part of a word) is most likely to come next in a piece of text, given the preceding context. They are then often further trained (i.e., *fine-tuned*) specifically to follow instructions or to answer questions. Doing so enables more straightforward interaction with users and has led to promising results on tasks like automatically grading students' short-answer responses (Kortemeyer, 2023). However, substantial issues can plague applying some common LLMs for such tasks, especially related to privacy concerns that are salient in education. Students, parents, instructors, and other stakeholders may object to transmitting student work (especially open-ended forum discussions that often include identifying and personal information) to online LLM services, even given assurances that the data will not be harvested. Moreover, regulations may restrict the transmission of such data (e.g., the Family Educational Rights and Privacy Act [FERPA] in the United States). Consequently, in this research we focus on freely available LLMs that can be applied without special computing resources, and, more importantly, that are offline—i.e., they do not require transmitting students' discussion forum text anywhere outside of the researcher or instructor's computer.

Given the potential opportunities to study help-seeking in course discussion forums via scalable automatic methods, we investigate the question *how well do current offline LLMs work for automatically detecting help-seeking?* We do so with minimal customization required for the help-seeking construct specifically, with the intention of informing future efforts to measure other constructs without requiring customized methods and models for each one. Simplicity is especially important given the time and monetary costs involved with developing bespoke machine learning models for tasks in educational data (Hollands & Bakir, 2015). The efforts required for detecting help-seeking using LLMs are relatively lower but still substantial, given that results must be measured against a validated coding schema, although this can be postponed until after initial results have face validity (as described next).

## Method

In this study, we detected help-seeking events from students' discussion forum posts in an online, introductory-level natural science course at a large public university in the United States. Students in this course participated through a learning management system called LON-CAPA (Learning Online Network with a Computer-Assisted Personalization Approach), which delivers educational content, auto-graded exercises, and discussion forum functionality (Kortemeyer et al., 2003). Forum content included hierarchically threaded conversations between students and, in some cases, the instructor. For the purposes of this study, we consider only the top-level (i.e., initial, non-reply) forum posts from students. Forum posts constituted 5% of students' final grade in the class, which led to 860 posts from 82 students (and a further 70 posts from the instructor).

Many different LLMs, of varying complexity, have been trained on different text corpora; thus, we compared several to understand how much they might vary in their usefulness for measuring help-seeking. Here, we used 5 different fine-tuned offline LLMs to analyze 297 top-level student forum posts. These models were fine-tuned from two different "foundation" models, which are the LLMs trained to predict the next word in large text corpora before fine-tuning for instruction-following and question-answering. Specifically, we used models based on LLaMA 2 and Mistral (Jiang et al., 2023; Touvron et al., 2023), each of which was published with a fine-tuned version that was the one we used in this study. We also evaluated a publicly available alternative ("B" version) fine-tuned LLaMA 2 model and two ("B" and "C") alternatives based on Mistral. These alternatives were fine-tuned on additional text intended to improve their instruction-following properties.

We provided a prompt to each model that included a very short definition of help-seeking in the forum context (i.e., "*A request for help can consist of an explicit question or implicit indication that help is needed.*") and brief instructions to rate the forum post text that followed. These instructions were dramatically shortened and rewritten versus the instructions human raters followed, after we observed that the instructions for human raters were followed poorly by the LLMs. Specifically, LLMs tended to take an overly expansive interpretation of each instruction given, and to confuse long instructions with the forum post itself (despite appropriate delimiters). Hence, a less detailed prompt was more successful.

The prompt included instructions to "*Rate on a 0-9 scale if the following text contains a request for help, where 9 means that it definitely includes a request for help.*" We selected 0-9 because these LLMs generate numbers one digit at a time; if we had used a 1-10 scale, for example, the probability of a "1" and "10" rating would be more difficult to distinguish because a "10" rating would first require generating the "1". We also observed that the distribution of rating probabilities differed substantially, such that in some cases it was unimodal (i.e., one very likely rating) and in others multimodal (i.e., two or more similarly likely but dissimilar ratings). For example, if 3 is the most probable rating but only by a small margin over 7, 8, and 9, then perhaps 3 is an outlier and a higher rating is more appropriate. We rated each post nine times and computed the median as the final rating to avoid such outliers. LLMs were constrained to generate only values in the rating scale via the *Guidance* Python library (Lundberg, 2023). We also generated a brief explanation (approximately 100 words) from the LLMs for each rating. These are not a required part of the measurement process, and do not necessarily represent the actual reason why a rating was given. However, in developing the prompt for help-seeking detection, we found explanations useful for improving the prompt; there is at least some connection between the rating and the explanation, so clear mistakes in the explanation helped inform the prompt text (most of all, leading us to use a very short definition of help-seeking).

Finally, to determine how well the LLMs rated help-seeking, we compared LLM ratings to human expert ratings via an existing coding schema for help-seeking (Jay et al., 2020), which includes four levels: 1) no question or request for help; 2) question asked but no request for help; 3) implicit request for help; and 4) explicit request for help. We compared the models' ratings to human ratings via Spearman's *rho*, and compared models to each other with a *z*-test (Myers & Sirois, 2006). Note that while we tested a few prompt variations to produce reasonable-seeming explanations, we did so on only a few of the posts—moreover, we only compared LLM and human ratings once at the end of the process to avoid cherry-picking a prompt variation that works well on the measure of convergent validity. This also mirrors a typical use case where a researcher might adjust a prompt several times to achieve promising results and only spend time manually coding data if it seems like the LLM measure is promising.

We were particularly interested in LLMs that can run on typical consumer hardware. Hence, all experiments were conducted on a laptop with 16GB memory and an Intel Core i7-1165G7 processor (a four-core, 2.8GHz processor released in 2020). Mistral-based models have 7 billion parameters, while LLaMA-based models exist in several sizes (we used the 13 billion parameter size); both are too large fit in system memory, but can be *quantized* (i.e., reducing the precision of parameters while minimizing loss of generation quality) to fit. All of the code, prompts, links to quantized models, and instructions necessary for running our help-seeking experiments are available (<https://github.com/pnb/llm-measurement>).

## Results and discussion

Help-seeking predictions from the original LLaMA 2 model correlated  $\rho = .650$  with human ratings, indicating substantial accuracy for the model. This and all correlations were significantly greater than 0;  $p < .001$ . We compared the other models to this model, which all had accuracy that was lower or statistically indistinguishable (Table 1). Only two models, one fine-tuned LLaMA 2 model and one fine-tuned Mistral-based model, had significantly lower correlations with the human ratings, with  $\rho = .506$  ( $z = 2.653, p = .008$ ) and  $\rho = .529$  ( $z = 2.269, p = .023$ ) respectively. Hence, all models produced ratings substantially better than chance, but choice of model did make a difference in some cases.

We also examined the pairwise correlations between the individual models to understand whether models were wrong in a similar way (which may suggest a common cause, such as a misleading element of the prompt) versus wrong in different ways (which might suggest errors were more due to the models). The highest correlation between all pairs of models was  $\rho = .731$ , which was between the original LLaMA 2 model and the original Mistral model. The lowest was  $\rho = .415$ , between the LLaMA 2 B version and the Mistral B version, and the mean of all correlations was  $\rho = .566$ . Thus, models' ratings of help-seeking were only correlated with each other to approximately the same degree as those ratings correlated with human ratings. Hence, there may be opportunities for improvement in the LLMs themselves.

One consequence of having several measures that are not strongly related to each other is that, if averaged together, their errors cancel each other out to some extent (as opposed to correlated errors, which remain after averaging). Thus, we also analyzed a model consisting of the average rating of all five LLMs. The result ( $\rho = .711$ ) was not significantly better than the LLaMA 2 model ( $z = 1.361, p = .173$ ), but certainly at least as accurate as the best model and suggestive of improvements that could be made with more (rather than only better) models.

Finally, we also analyzed explanations from incorrect predictions the LLaMA 2 model made as examples of limitations in the method that could inform future improvements. Specifically, we selected the forum posts with the highest LLM help-seeking rating given the lowest human rating, and the lowest LLM rating given the highest human rating. Two forum posts satisfied the first criteria; these posts represent "false positive" cases where the model identified help-seeking that was not there. In the first such post, the student outlined how to solve a problem, describing it as "tricky," which the LLM explanation identified as an indication that the student might need help (though they did not). In the second false positive case, a student again outlined how to solve a problem and ended with "what am I doing wrong?", which the LLM explanation identified as an indicator of help-seeking. This latter case may indicate an improvement that could be made to the help-seeking prompt, which does indicate that a question qualifies as help-seeking, whereas the coding schema from Authors (2020) differentiates between asking a question with vs. without recognizing the community (i.e., peers in the discussion forum) as an essential ingredient in help-seeking. In contrast, there was one forum post that had the lowest LLM help-seeking rating given the highest human rating; this "false negative" case involved the student recognizing the community by asking if anyone could verify the correctness of their understanding, which the LLM explanation incorrectly indicated was "asking for confirmation, rather than seeking assistance." Improvements to the help-seeking prompt may help in this case, but may also require newer, more capable LLMs to process prompts more accurately—in this study, when given the entire coding schema from Authors (2020), models consistently confused parts of the coding schema with the forum text despite delimiters (including delimiters officially supported by the models).

**Table 1**  
*Results of Help-Seeking Detection by Different LLMs*

Model	Spearman's $\rho$	$z$	$p$ -value
LLaMA 2	.650		
LLaMA 2 (fine-tuned version B)	.506	2.653	.008
Mistral	.644	0.128	.898
Mistral (fine-tuned version B)	.570	1.562	.118
Mistral (fine-tuned version C)	.529	2.269	.023

*Note:* The  $z$  and  $p$ -values indicate difference in  $\rho$  versus the first model.

## Conclusion

In this study, we sought to determine how well current offline LLMs work for detecting help-seeking in an online college course discussion forum. The results indicated that it was indeed feasible, with a medium-to-strong association between LLM ratings and human ratings of help-seeking. Moreover, it was possible to make these automatic ratings without transmitting students' forum posts to online LLM services, and using only typical laptop



hardware. Our approach also offered advantages over traditional machine learning approaches, which require large amounts of manually coded data to train a model and more to test it, whereas the approach in this study requires only data to test the model—and even then, only if the model appears worth testing. The advantages of this method thus open up new use cases for automatic analysis of course discussion forums with respect to help-seeking and perhaps other constructs for which effective prompts can be written. For example, automatic analysis of help-seeking in forums could provide statistical power to detect the size of effects that may be expected (Fong et al., 2023), power to analyze proportionally small groups in large courses, and opportunities for instructors of large courses to find forum posts where their input may be the most helpful.

There remains work to be done with this approach as well. LLMs are trained on vast amounts of text that represent the cultures, languages, and linguistic styles of some students much more than others; hence, research is needed to determine and counteract potential systematic biases in the help-seeking ratings generated by LLMs. Additionally, as the landscape of LLMs shifts rapidly, new offline LLMs will need to be compared to determine whether they are better able to follow prompts, and especially to implement an entire coding schema given as part of the prompt—something that larger models, such as GPT-4, can perhaps do more effectively. However, current results are already promising and applicable, yielding immediate opportunities for the study and improvement of online course discussion forums.

## References

- Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, 49(3), 803–821. <https://doi.org/10.3758/s13428-016-0743-z>
- Fiock, H. (2020). Designing a community of inquiry in online courses. *The International Review of Research in Open and Distributed Learning*, 21(1), 135–153. <https://doi.org/10.19173/irrodl.v20i5.3985>
- Fong, C. J., Gonzales, C., Hill-Troglin Cox, C., & Shinn, H. B. (2023). Academic help-seeking and achievement of postsecondary students: A meta-analytic investigation. *Journal of Educational Psychology*, 115(1), 1–21. <https://doi.org/10.1037/edu0000725>
- Hollands, F., & Bakir, I. (2015). *Efficiency of automated detectors of learner engagement and affect compared with traditional observation methods* (pp. 1–37). Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. <https://repository.upenn.edu/abcse/4>
- Jay, V., Henricks, G. M., Anderson, C. J., Angrave, L., Bosch, N., Williams-Dobosz, D., Shaik, N., Bhat, S., & Perry, M. (2020). Online discussion forum help-seeking behaviors of students underrepresented in STEM. *Proceedings of the 14th International Conference on Learning Sciences (ICLS 2020)*, 809–810.
- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. de las, Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., & Sayed, W. E. (2023). *Mistral 7B*. <https://doi.org/10.48550/arXiv.2310.06825>
- Kortemeyer, G. (2023). *Performance of the pre-trained large language model GPT-4 on automated short answer grading* (arXiv:2309.09338). arXiv. <https://doi.org/10.48550/arXiv.2309.09338>
- Kortemeyer, G., Albertelli, G., Bauer, W., Berryman, F., Bowers, J., Hall, M., Kashy, E., Kashy, D., Keefe, H., Behrouz, M.-B., Punch, W. F., Sakharuk, A., & Speier, C. (2003). The learning online network with computer-assisted personalized approach (LON-CAPA). *Computer Based Learning in Science (CBLIS 2003)*, 119–130.
- Lundberg, S. M. (2023). *A guidance language for controlling large language models* [Python]. Microsoft. <https://github.com/microsoft/guidance>
- Myers, L., & Sirois, M. J. (2006). Spearman correlation coefficients, differences between. In *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/0471667196.ess5050.pub2>
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., ... Scialom, T. (2023). *Llama 2: Open foundation and fine-tuned chat models*. <https://doi.org/10.48550/arXiv.2307.09288>

## Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R305A180211 to the Board of Trustees of the University of Illinois. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education

# Examining Collaborative Clinical Reasoning within Synchronous Computer-Supported Collaborative Learning: A Cross-Cultural Comparison of Dutch and Finnish Medicine Students

Chrissa Manoli, Hannie Gijlers, Chandan Dasgupta  
c.m.manoli@utwente.nl, a.h.gijlers@utwente.nl, c.dasgupta@utwente.nl,  
University of Twente

Bas de Leng, University of Muenster, bdeleng@uni-muenster.de

**Abstract:** Mastering Collaborative Clinical Reasoning (CCR) can be challenging for medical students as they are required reach collaborative treatment decisions through effective argumentation. Computer-supported collaborative learning (CSCL) contexts are effective for facilitating CCR. Internationalization in higher education introduces additional challenges for CCR, such as communication barriers, process losses, and conflicts. To enhance CCR instruction via CSCL approaches, understanding cultural differences in argumentation differences is vital. This study examined CCR argumentation differences between Dutch and Finnish undergraduate medicine student groups. Despite both cultures being individualistic, Finnish students displayed more group-oriented, and data-supported arguments, while Dutch students were more pragmatic and assertive. Future research should focus on integrating scaffolding interventions in CSCL to explicitly teach clinical argumentation and intercultural competence.

## Introduction

Global health education incorporates clinical reasoning skills into medical curricula to equip future health professionals for diagnostic decision-making. Collaborative clinical reasoning (CCR) among physicians aims to establish effective treatment plans, enhance patient safety and improve team performance (Gordon et al., 2012). Despite its benefits, challenges such as weak evidence and hindered information-sharing persist (de Leng & Gijlers, 2015), partly due to individualistic teaching approaches (Kiesewetter et al., 2022), highlighting the need for collaborative learning. Globalization has increased cultural diversity in healthcare teams, affecting CCR quality due to communication challenges and conflicts (Gonzalez-Caminal & Kangasperko, 2023), therefore posing challenges for collaborative learning. Despite multicultural teams being common, there's insufficient focus on collaborative argumentation and intercultural competencies.

To tackle these issues, computer-supported collaborative learning (CSCL) using tailored tools can be employed. In multicultural contexts, CSCL creates a supportive environment promoting equitable participation, and enhancing intercultural awareness (Popov et al., 2014). In medical education, CSCL tools scaffold students' argumentative CCR by facilitating evidence-based hypotheses and diagnostic decisions, by highlighting appropriate argumentative components (Noroozi et al., 2013). Developing argumentative knowledge is crucial for CCR and can serve as evidence of students' CCR quality (Ju & Choi, 2018; Si et al., 2018). Nonetheless, the existing body of research focuses on asynchronous collaboration, providing limited insights into efficacious CSCL pedagogical strategies for enhancing synchronous CCR practice (Blondon et al., 2017; Gordon et al., 2012).

This paper investigates how the quality of the clinical argumentation between two culturally individualistic student populations – Finland and the Netherlands – compares. The study seeks to enhance intercultural competence in medical education curricula and integrate argumentation scaffolding tools for CCR in synchronous CSCL. It attempts to answer the following research question - Do the culturally homogeneous Dutch and Finnish medicine student groups display the same level of CCR argumentative quality?

## Theoretical framework

### Collaborative Clinical Reasoning (CCR) in Computer-Supported Collaborative Learning (CSCL)

Collaborative Clinical Reasoning (CCR) is the process whereby medical experts work together to analyze, justify, and negotiate diagnostic, prognostic, and therapeutic issues to establish a mutual understanding of the patient's condition and the approach to its treatment (Blondon et al., 2017; Kiesewetter et al., 2022). Clinical reasoning is a core skill in medical education, aiming at the development of students' and practitioners' diagnostic competence – the capacity to synthesize biomedical and clinical to generate effective therapeutic solutions (Hege et al., 2018).

In interprofessional healthcare contexts, the complexity of medical cases, the demands of specialized care, and the rigorous patient safety regulations often require collaboration among multiple practitioners. Physicians must therefore not only exhibit clinical expertise for effective problem-solving, but also adeptness in interpersonal competences encompassing information-sharing, elicitation, negotiation, and coordination in complex, time-pressing scenarios (Kiesewetter et al., 2022), alongside proficiency in cross-disciplinary and intercultural collaboration (Radkowitz et al., 2021). Computer-supported collaborative learning (CSCL) environments provide controlled practice settings for learners to enhance clinical reasoning at their own speed, gaining insights from diagnostic errors while avoiding patient harm (Hege et al., 2018). In CSCL medical education, digital tools, including visualization and argumentation tools, facilitate knowledge co-construction (Fischer et al., 2014). Visualization tools like web microscopes, interactive diagrams, and virtual patient simulations aid comprehension of complex medical topics, such as pathology and radiology (de Leng & Gijlers, 2015; Si et al., 2018).

## CCR argumentation dimensions

Argumentation plays a pivotal role in scientific reasoning and disciplinary knowledge acquisition (Weinberger et al., 2007). Its quality not only facilitates knowledge recognition but also serves as a benchmark for assessing CCR proficiency. Within the context of CSCL among medical students, understanding argument construction and its social dimension is paramount for knowledge acquisition. Arguments are analyzed as single components or in sequences (Weinberger & Fischer, 2006). In CCR, argument soundness requires three fundamental components: *claim(s)*: the position advanced, ground with *data*: facts and *warrants*: reasons justifying the transition from the data to the claim. Arguments might possess *qualifiers* limiting the claim's validity, and *rebuttals*, invalidating the claim. High quality argumentative sequences consist of *counterarguments* challenging the initial *argument* and *integrations* synthesizing the initial argument and the counterargument. Discourse can also be characterized by *non-argumentative moves* with no propositional value like questions, coordinating moves, and meta-statements. Balancing these, promotes deeper thinking and knowledge construction in complex problem resolution.

In social contexts, arguments' link to peer contributions influences knowledge acquisition. On the lowest level, arguments can be *externalizations* thought articulation, *elicitations* (response stimulation or questioning), or might serve for *quick-consensus building*. Conversely, *Conflict* and *Integration-oriented consensus-building* enhance argumentative quality by encouraging adaptive beliefs, and critical thinking through peer reasoning.

## Cross-cultural differences in CCR

Cultural diversity shapes learners' argumentation and learning abilities (Weinberger et al., 2007). Despite assertions of universal reasoning norms (Mercier, 2011), cultural factors impact engagement, argument types, and discourse quality (Hornikx & ter Haar, 2013). Cultural differences, therefore, significantly impact group dynamics, including trust-building and conflict resolution, requiring careful consideration in CSCL.

Comparisons between culturally proximate homogeneous groups can unveil cultural effects on argumentation patterns (Hornikx & ter Haar, 2013; Hornikx & Hoeken, 2007). Understanding these patterns can guide the development of effective scaffolds for intercultural collaboration, enhancing thus intercultural competence. Hofstede's (1991) framework presents six cultural dimensions as dual continua, enhancing cultural understanding and enabling cross-cultural assessments. The Individualist-Collectivist dimension entails that individualists prioritize personal goals, while collectivists emphasize group success, identity, norms, and commitments. Popov et al., 2014, found that learners from individualistic cultures, outperformed those from collectivist background.

Cross-cultural studies show Dutch participants favor direct communication and decentralized team management (Labrie et al., 2020), while Finnish participants prefer indirect communication and emphasize collaboration (Kim & Bonk, 2002; Marttunen & Laurinen, 2002). However, there's limited research on how cultural homogeneity-heterogeneity affects argumentative quality, underlining the necessity to comprehend learners' cultural backgrounds for adaptive learning environments (Hornikx & ter Haar, 2013).

## Methods

### Design and Instrumentation

This study which is part of a larger quasi-experimental research, compares two distinct cohorts of medical students hailing from universities in the Netherlands and Finland. In this study, we narrow our focus to examine the outcomes within the homogeneous Dutch and Finnish groups to get a deeper understanding of the characteristic differences in CCR quality. All groups worked on a CSCL scenario in microscopic pathology classes, incorporating an open-source web microscope, VQuest assessment tool, and PRISMA dashboard for collective visualization.

## Respondents and procedure

In a study with  $n=67$  students, including  $n=31$  second-year Dutch and  $n=36$  third-year Finnish students, all recently completing a microscopic pathology course, participants were divided into eleven groups: five culturally diverse, three Dutch, and three Finnish, each with around six students. We focus on the homogenous groups: D1 ( $n=7$ ), D2 ( $n=6$ ), D3 ( $n=7$ ), F1 ( $n=6$ ), F2 ( $n=6$ ), F3 ( $n=7$ ), with four female and two male students in each group. Ethical approval was obtained, and students provided informed consent. The study involved a 4-hour session with two phases introducing VQuest for solving medical cases, including group tasks, breakout sessions, and plenary discussions. In plenary sessions, instructors guided discussions for decision-sharing, feedback, and reflective learning. Cameras were often off for confidentiality during breakout sessions, impacting communication.

## Analysis

We assessed students' Critical Clinical Reasoning (CCR) argumentative discourse using Weinberger and Fischer's (2006) framework. Dialogues were transcribed and coded on the micro-level (claim, warrant, data, rebuttal, qualifier) and macro levels (argument-ARG, counterargument-CARG, integration-INT) of argumentation, including non-argumentative moves (Non-ARG). Social co-construction modes were coded as Externalization-EXT, Elicitation-ELC, quick consensus-building –CNS, Integration-oriented consensus-building – Int-CNS, and Conflict-oriented consensus-building –Con-CNS. Inter-coder reliability ( $\alpha = 0.923$ ) was high.

## Results and discussion

Our research delved into the argumentation styles of Dutch and Finnish students, revealing significant cultural variations consistent with Hofstede's findings (1991). While past intercultural studies in CSCL primarily focused on asynchronous contexts (van der Meijden, 2005), our work contributes by examining synchronous processes. Participants' argumentation styles mirrored their respective cultures, with Finnish students displaying a tendency to initiate more arguments and provide extensive data to support their claims. This contrasts with Dutch students, who adopted a more pragmatic, albeit less substantiated approach. Analysis revealed distinctions between Dutch and Finnish groups. Finnish students leaned towards theoretical underpinnings, possibly due to Finland's higher uncertainty avoidance tendencies (Labrie et al., 2020), while Dutch students incorporated more elicitations reflecting their communication norms. Finnish groups favored consensus-building with qualifiers, mirroring their communication style, while Dutch groups emphasized unanimity through elicitations, aligning with Dutch communication tendencies (Hornikx & Hoeken, 2007). Finnish students showed deeper engagement in CCR, reflecting a theory-oriented approach, in line with Hofstede's uncertainty avoidance index, while Dutch students exhibited pragmatism, reflecting cultural individualism.

## Limitations and future research

Our study, though limited by an exploratory approach and confidentiality constraints, highlights the urgent need for future research to employ experimental designs and larger sample sizes. This will allow for a deeper understanding of collaborative clinical reasoning (CCR) in synchronous contexts, ultimately optimizing learning environments. Moreover, leveraging frameworks such as Ju and Choi's (2018) can aid in the development of culturally tailored interventions, promoting the construction of robust clinical claims. Exploring the influence of cultural differences on CCR further, and implementing supportive tools like collaborative diagrams and argumentation maps, will enhance argument formation during collaborative CCR stages. Future studies should also integrate multilevel analyses to untangle the complexity of data structures and explore the effects of various cultural dimensions on CCR quality, as suggested by Janssen et al. (2013). In conclusion, our research sheds light on the intricate interplay between culture and argumentation styles in CCR, providing valuable insights for educators and researchers alike.

## References

- Blondon, K. S., Maître, F., Muller-Juge, V., Bochatay, N., Cullati, S., Hudelson, P., Vu, N. V., Savoldelli, G. L., & Nendaz, M. R. (2017). Interprofessional collaborative reasoning by residents and nurses in internal medicine: Evidence from a simulation study. *Medical Teacher*, 39(4), 360–367. <https://doi.org/10.1080/0142159x.2017.1286309>
- de Leng, B., & Gijlers, H. (2015). Collaborative diagramming during problem based learning in medical education: Do computerized diagrams support basic science knowledge construction? *Medical Teacher*, 37(5), 450–456. <https://doi.org/10.3109/0142159X.2014.956053>

- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dörner, B., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., Eberle, J., & Maximilians, L. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, 5, 28–45. <https://doi.org/10.14786/flr.v2i3.96>
- Gonzalez-Caminal, G., Kangasperko, M. (2023). Co-creating in International Environments. A Clinical Reasoning Course for Physiotherapy Students. *The Educational Review, USA*, 7(4), 443-448. <https://doi.org/10.26855/er.2023.04.009>
- Gordon, M., Darbyshire, D., & Baker, P. (2012). *Medical Education*, 46(11), 1042–1054. <https://doi.org/10.1111/j.1365-2923.2012.04343.x>
- Hege, I., Kononowicz, A. A., Kiesewetter, J., & Foster-Johnson, L. (2018). Uncovering the relation between clinical reasoning and diagnostic accuracy – An analysis of learner’s clinical reasoning processes in virtual patients. *PLOS ONE*, 13(10). <https://doi.org/10.1371/journal.pone.0204900>
- Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind*. London: McGraw-Hill.
- Hornikx, J., & Hoeken, H. (2007). Cultural Differences in the Persuasiveness of Evidence Types and Evidence Quality. *Communication Monographs*, 74(4), 443–463. <https://doi.org/10.1080/03637750701716578>
- Hornikx, J., & ter Haar, M. (2013). Evidence Quality and Persuasiveness: Germans Are Not Sensitive to the Quality of Statistical Evidence. *Journal of Cognition and Culture*, 13(5), 483–501. <https://doi.org/10.1163/15685373-12342105>
- Janssen, J. J. H. M., Cress, U., Erkens, G., & Kirschner, P. A. (2013). Multilevel analysis for the analysis of collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (pp. 112-125). Routledge.
- Ju, H., & Choi, I. (2018). The role of argumentation in hypothetico-deductive reasoning during problem-based learning in medical education: A conceptual framework. *Interdisciplinary Journal of Problem-Based Learning*, 12(1), 11–14. <https://doi.org/10.7771/1541-5015.1638>
- Kim, K.-J., & Bonk, C. J. (2002). Cross-cultural Comparisons of Online Collaboration. *Journal of Computer-Mediated Communication*, 8(1). <https://doi.org/10.1111/j.1083-6101.2002.tb00163.x>
- Kiesewetter, J., Hege, I., Sailer, M., Bauer, E., Schulz, C., Platz, M., & Adler, M. (2022). Implementing Remote Collaboration in a Virtual Patient Platform: Usability Study. *JMIR Medical Education*, 8(3), e24306. <https://doi.org/10.2196/24306>
- Labrie, N., Aranka Akkermans, & Hample, D. (2020). A Dutch dose of dissent. *Journal of Argumentation in Context*, 9(2), 219–242. <https://doi.org/10.1075/jaic.19009.lab>
- Marttunen, M., & Laurinen, L. (2002). *Learning Environments Research*, 5(1), 99–123. <https://doi.org/10.1023/a:1015666724692>
- Mercier, H. (2011). On the Universality of Argumentative Reasoning. *Journal of Cognition and Culture*, 11(1-2), 85–113. <https://doi.org/10.1163/156853711x568707>
- Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers & Education*, 61, 59–76. <https://doi.org/10.1016/j.compedu.2012.08.013>
- Popov, V., Noroozi, O., Barrett, J. B., Biemans, H. J. A., Teasley, S. D., Slof, B., & Mulder, M. (2014). Perceptions and experiences of, and outcomes for, university students in culturally diversified dyads in a computer-supported collaborative learning environment. *Computers in Human Behavior*, 32, 186–200. <https://doi.org/10.1016/j.chb.2013.12.008>
- Radkowsitch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, 75. <https://doi.org/10.1016/j.learninstruc.2021.101487>
- Si, J., Kong, H. H., & Lee, S. H. (2018). Developing clinical reasoning skills through argumentation with the concept map method in medical problem-based learning. *Interdisciplinary Journal of Problem-Based Learning*, 13(1), 1–11. <https://doi.org/10.7771/1541-5015.1776>
- Weinberger, A., Clark, D. B., Häkkinen, P., Tamura, Y., & Fischer, F. (2007). Argumentative Knowledge Construction in Online Learning Environments in and across Different Cultures: A Collaboration Script Perspective. *Research in Comparative and International Education*, 2(1), 68–79. <https://doi.org/10.2304/rcie.2007.2.1.68>
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95. <https://doi.org/10.1016/j.compedu.2005.04.003>
- van der Meijden, H. (2005). *Knowledge Construction through CSCL: Student Elaborations in Syn-chronous, Asynchronous, and Three-Dimensional Learning Environments*. n.a

## Convergent Group Understanding: Indicator of Knowledge Integration?

Xuesong Cang, University of Wisconsin-Madison, xcang@wisc.edu  
Litong Zeng, University of Illinois, Urbana-Champaign, litongz2@illinois.edu  
Mike Tissenbaum, University of Illinois, Urbana-Champaign, miketiss@illinois.edu  
Dana Gnesdilow, University of Wisconsin, gnesdilow@wisc.edu  
Sadhana Puntambekar, University of Wisconsin, puntambekar@wisc.edu

**Abstract:** This study investigates small group collaborative learning with a technology-supported environment. We aim to reveal key aspects of collaborative learning by examining variations in interaction, the influence of small group collaboration on science knowledge integration, and the implications for individual knowledge mastery. Results underscore the importance of high-quality science discourse and user-friendly tools. The study also highlights that group-level negotiations may not always affect individual understanding. Overall, this research offers insights into the complexities of collaboration and its impact on science learning.

### Introduction

Computer-supported collaborative learning is influenced by multifaceted factors, such as the alignment of an individual's own understanding and the coordination of divergent ideas between group members (Järvelä et al., 2015). It reflects the intertwined nature of an individual's own learning and that of others in same small group (Stahl, 2015). Examining the dynamics of interactions in small groups being supported by distributed technological tools, such as actions, goals, and knowledge convergence toward shared understanding, can offer insights on the collaborative process (Damşa, 2014). Moreover, shared understanding does not merely result from aggregating individual ideas and actions in group collaboration (Roschelle, 2012). It raises the question of whether shared understanding and knowledge co-construction implies a more complex comprehension of science knowledge and phenomena, both for individual learning and for the products of group learning.

This study aims at (1) unfolding the various types of interactions among students engaged in small group collaboration, (2) investigating the extent to which students collectively integrated knowledge within groups, and (3) exploring how individual learning was influenced by the knowledge integration contributed by group members. We used the Knowledge Integration (KI) framework to examine knowledge components and their relationships (Liu et al., 2008). Our research questions were: (1) What were the key variations in collaborative approaches utilized by different groups in understanding scientific concepts and relationships during a design-based unit? (2) How does small group collaboration impact students' science knowledge integration? (3) What implications does group knowledge integration hold for an individual's learning?

### Method

#### Context and participants

Eighteen 8th grade students from the midwestern United States engaged in a 13-day (50 minutes a day) curriculum unit entitled, "*Make Your Own Compost*". In the unit, students worked on a compost design challenge by building a physical decomposition bioreactor, conducting virtual experiments, and writing a final design recommendation. Their collaborative efforts were facilitated by various tools, including (1) a virtual compost simulation; (2) a digital interactive concept map (VidyaMap) for biology concepts, and (3) the Idea Wall that is interactive allowing students to see, move, vote, and combine notes. All tools were encapsulated within a digital Science Notebook to support individual's and group's scientific reasoning. Seven students from two groups ( $N_1=3$ ;  $N_2=4$ ) in one class were videotaped during four activities supported by these technological tools.

#### Data resources and analysis

*Video Data* was transcribed for analysis by turns of talk. The teacher's talk was excluded, resulting in 418 turns of talk ( $N_1=120$ ;  $N_2=298$ ). We utilized inductive and deductive techniques to code and capture interactions (Derry et al., 2010). A coding scheme (Table 1) with three themes emerging from eight codes was developed to capture how students (1) created joint attention & awareness that refers to the shared focus of individuals on the same object, (2) constructed shared science understanding that contributed to collective comprehension of knowledge, and (3) goal adaptations between individual and group goals to generate collective learning products (Damşa, 2014; Tissenbaum et al., 2017). Two raters achieved a Kappa of .80 for inter-rater reliability across 30% of the

data. Disagreements were resolved and the rest of the data were coded by the two raters independently. A mixed approach that quantified the qualitative science discourse in the videos was used, creating an overview of the types of verbal actions in the interactions. The percentage of each code was calculated.

**Table 1**  
*Science Discourse Coding Scheme (SD-Codes)*

Interactions	Actions	Descriptions	Examples
<b>Joint attention &amp; awareness</b>	Narration	Share information from learning resources	"209 grams of water" (Read notes)
	Procedural clarification	Make inquiry about the procedural tasks in the activity and the responses to these inquiries	"Will you guys look through the prediction?"
	Tools Engagement	Talk about the use the tools (e.g., drag and relocate Ideas, note combination or deletion)	"Yeah, put that into the combine zone." Or "Control deletes."
<b>Shared Science Understanding</b>	Idea generation	Bring new ideas (e.g., hypothesis, predictions)	"I want to say like aeration"
	Idea negotiation	Ask peers to explain their ideas, suggest next steps for group actions or disagree with peers	"No, ... the amount of oxygen and hydrogen affects water."
	Ideas taken-up	Agree or accept peers' input or explain their own ideas being accepted to group ideas	"Yeah, that's material. I'll put down that."
<b>Goals adaptation</b>	Individual goal	Describe, ask, or adjust their individual goal	"I'm going to do abio-factors."
	Group goal	Inquiry or report on the group learning process or group learning products	"How much do you guys get done right now?"

The Idea Wall log data from was retrieved, encoded, and organized at the unit of individual notes. Each note entry consisted of information on (1) *Group Name*; (2) *Note Content*; (3) *Vote History*, recording the manipulations of notes to "Yup", "No" or "Combine Zone"; and (4) *Note Combination*, records reflecting if notes were combined to generate a new note, such as the consolidation of two notes, "Moisture" and "dirt", into a single note, labeled "Dead frog". Forty-three entries were generated by two groups when they collectively reasoning about the factors influencing decomposition before (Day 3) and after (Day 10) compost investigations. The Knowledge Integration coding scheme (KI scheme), which uses a 0-5 score scale across six categories (*5-complex link*; *4-full link*; *3-partial link*; *2-no link*; *1-off task*; *0-no answer*) to gauge the complexity of Knowledge Integration (Liu et al., 2008), was used to analyze the videos and note data from Idea Wall sessions, to reveal KI within group learning.

*Application of the KI scheme to Idea Wall:* *5-complex link* indicates scientific explanations of how two more ideas are related influencing decomposition; *4-full link* is elaboration of a complete idea by synthesizing fragmented ideas on notes. We did not identify examples identified in our study for *5-complex* and *4-full link*. *3-partial link* is new ideas by combing notes with fragmented similar ideas (e.g., Note "Carbon: Nitrogen ratio" is generated by combining "Carbon to nitrogen" & "Green Brown"); *2-no link* means merging notes with identical ideas (e.g., Note "Moisture" is generated by merging "moisture" & "Moisture"); *1-off task* indicates irrelevant ideas (e.g., Note "Car" is generated by combing "hello" & "Fisher"); *0-no answer* is not applicable to our data.

The compost design report provided students with a chance to explain factors (e.g., Moisture, Particle Sizes) influence decomposition and their design. Design reports written by students in Group 2 were analyzed. One student in Group 1 did not write the design report and was excluded due to the incomplete data. The KI scheme was also used to understand individual students' knowledge integration levels in written Compost reports.

*Application of the KI scheme to Compost Reports:* *5-complex link* indicates scientific explanations of one decomposition factor relates to other factors in affecting compost; *4-full link* is the elaboration of one decomposition factor influences compost by specifying ideal range; *3-partial link* means implicit elaborations of one decomposition factor without giving ideal range; *2-no link* means stating one factor is related to decomposition without explaining how it relates; *1-off task* means unrelated content. *0-no answer* is no response.

## Results

### Overview of interactions between two groups being supported by distributed tools

We observed that Group 2 ( $N_2=298$ ) generated a higher number of science discourse turns in comparison to Group 1 ( $N_1=120$ ). This disparity may be attributed to Group 2 having one additional student, potentially leading to more extensive discussions. However, upon closer examination of their interactions, we identified notable similarities between the two groups. Both groups engaged in numerous instances of *Narrations* ( $N_1=10.00\%$ ;  $N_2=10.40\%$ ) and *Tool Engagement* ( $N_1=13.33\%$ ;  $N_2=11.44\%$ ) to establish joint attention and awareness. Additionally, we observed similar patterns in which students tried to get an alignment of individual ( $N_1=7.05\%$ ;  $N_2=9.06\%$ ) and

group goals ( $N_1=16.67\%$ ;  $N_2=15.10\%$ ). Much higher percentage of science discourses centered around group goals compared to individual goal potentially indicated the endeavors that students put in achieving shared group goals. Another commonality was the shared science understanding, where both groups contributed a similar percentage of science discourse in Ideas Generation ( $N_1=12.50\%$ ;  $N_2=10.74\%$ ) & Ideas Negotiation ( $N_1=25.00\%$ ;  $N_2=23.15$ ), indicating the active engagement in discourse to co-construct knowledge from both groups.

The main differences were in (1) Procedural Clarifications ( $N_1=14.17\%$ ;  $N_2=25.17\%$ ) and (2) Ideas Taken-up ( $N_1=13.33\%$ ;  $N_2=5.37\%$ ). The additional student in Group 2 likely increased focus on understanding procedural stages and clarification. This complexity also heightened challenges in achieving consensus on divergent ideas, resulting in fewer individual ideas being integrated into shared science understanding.

### Group-level knowledge integration using the Idea Wall

Eight out of forty-three note entries from the Idea Wall were created by merging existing notes. This combining of notes predominantly occurred during the initial use of the Idea Wall (Day 3). However, all the notes were coded at the "2-No link" KI coding level. Students demonstrated "3-Partial Link" connections during the second Idea Wall episode (Day 10) using the combine zone feature. We did not find any other levels of connection.

We further examined students' talk during group discussions to illustrate how the verbal negotiations and actions in the Idea Wall related to the instances of "No link" or "Partial Link" to gain insights on how group interactions were mediated by the Idea Wall. Table 2 shows how students negotiated what to keep when seeing identical ideas across notes as well as the verbal Science Discourse Code (SD-code). Table 3 shows students' talk as they created a note "Carbon: Nitrogen (Ratio)" by merging notes with similar ideas "Browns : Greens" & "Carbon/Nitrogen." Group members negotiated their differing ideas regarding whether "Greens to Browns ratio" and "Carbon and Nitrogen ratio" were the same. Students didn't confirm if two notes were the same, but the recorded combined action on the Idea Wall suggests agreement. Both exemplars were from Group 2.

**Table 2**

*Exemplar Science Discourse for 2-No Link (Merge identical notes)*

Student	Transcript	SD-Codes
A	We have three different moisture notes	Narration
B	Okay, I might delete it. Can you delete it?	Procedural Inquiry
C	No, combine this moisture and this moisture?	Ideas Negotiation; Tool Engagement; Group goal
A	Yeah.	Ideas Taken-up; Group goal

**Table 3**

*Exemplar science discourse for 3-Partial Link (Merge two notes with similar science ideas)*

C	Carbon to nitrogen is the same thing as Green to Browns	Ideas Generation
A	Brown and greens are facts, but carbon and nitrogen are factors.	Ideas Negotiation
	So we are saying H <sub>2</sub> O and water are different things?	Ideas Negotiation
A	No, it's like saying that the amount of oxygen and hydrogen, like affects the amount of water.	Ideas Negotiation

### Knowledge integration in an individual's compost design from Group 2

We noticed varying levels of individual's understanding among students indicating different levels of KI. Student B provided explanations for all four factors, integrating information from the Idea Wall with "4-Full link" (Carbon-Nitrogen Ratio, Particle Size & Turning) or "5-Complex link" (Moisture). For example, student B explained that "it's good to have ...40% to 55% moisture. If the compost is too wet, there will not be enough air flow. If it is too dry it's hard for the decomposers to live in." This explanation explicitly stated the ideal range of moisture and how moisture level influences decomposers that directly affect decomposition. Student B was also the only student who explained the factor turning by writing, "turn or mix the compost every 1 to 7 days, anything more than 7 will make a slow decomposition." Student A explained three factors (all except Turning), at the "3-Partial Link" level. Taking one of the implicit explanations (Particle Size) from Student A as an example, "a medium particle size... would create the fastest decomposing compost system", student did not give ideal range of particle size and how it influences decomposition rate. Student C also explained three factors, with one at the "3-Partial Link" (Particle Size) level and the other two (Carbon-Nitrogen Ratio & Moisture) at the "2-No link" level. An example of the "no link" response can be seen in Student C's vague explanation that the Carbon-Nitrogen Ratio affects decomposition without providing any explanation about how, by writing that "amount of carbon and nitrogen...it affects how fast we can compost." Student D included information about how Moisture impacts compost at a "Partial Link" level and Carbon-Nitrogen Ratio at a "No link" level.



We also observed that the converged group understanding was not always exhibited at the individual level". For instance, Students A & C actively discussed whether "Green and Brown" and "Carbon and Nitrogen" represented the same concept and collectively decided to merge these terms into "Carbon and Nitrogen" on the Idea Wall. However, Student C employed the concept of "Carbon and Nitrogen" ("*amount of carbon and nitrogen...affects how fast we can compost*") to the final report while she initially thought "Green to Brown" is the same as "Carbon and Nitrogen" in the discussion. Student A wrote "*slightly more browns than greens would breast the fastest decomposing*" while he reasoned that these two concepts are different in the Ideal Wall session.

## Discussion and conclusions

We explored interactions that provide insights into the essential aspects of collaborative learning (Damşa, 2014; Järvelä et al., 2015) when students collaborated in technology-enriched learning environment (Stahl, 2015). Our findings show that despite differed length of turns of talk being generated, both groups exhibited similar patterns of discussion, with identical percentages of productive interactions. It implies that merely urging students to speak more shouldn't be the primary goal to foster productive interactions, particularly in a learning environment using various technological tools. Offering activities that help students to successfully negotiate conflicts, build upon others' ideas, and align personal actions to achieve common goals is important for collaboration and learning.

Our study also explores KI during collaborative interactions using the Idea Wall. Notably, initial usage revealed a higher number of entries for creating notes and combining actions, but all at the low level of KI, indicating superficial technology use. This could be due to students' limited familiarity with the tools or scientific concepts early in the unit. Providing support in comprehending the tools, through activities that introduce students to their functions or even allowing them to practice using the tools before engaging in group collaboration, can undoubtedly enhance interactions and potentially foster stronger integration of knowledge at the group level.

While students did not score highly in KI on their combined notes in Idea Wall, some students' final explanations scored much higher, indicating that they did benefit from the collaboration. However, our observations revealed an interesting nuance: traces of group science discourse were evident in the explanations of compost design for some individual students but not for others. This variance implies that group idea negotiations and collective decisions, intended to share meaning and knowledge, do not consistently translate into changes or manifestations in the individual understanding of students. This signifies the need for providing support for science understanding, whether through group-level activities or individual mastery of scientific concepts. As collaborative learning evolves, with a shift towards emphasizing group understanding over individual mastery from group work, it's vital to recognize how group-level learning outcomes affect KI knowledge across levels. This interconnectedness underscores the need to recognize the complex and multifaceted factors influencing collaborative learning, which do not neatly delineate the boundaries between individual and group learning.

## References

- Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning*, 9, 247-281.
- Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., ... & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *The Journal of the Learning Sciences*, 19(1), 3-53.
- Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., ... & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development*, 63, 125-142.
- Liu, O. L., Lee, H. S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment*, 13(1), 33-55.
- Roschelle, J. (1992). Learning by Collaborating: Convergent Conceptual change. *Journal of the Learning Sciences*, 2(3), 235-276. [https://doi.org/10.1207/s15327809jls0203\\_1](https://doi.org/10.1207/s15327809jls0203_1)
- Stahl, G. (2015). A decade of CSCL. *International Journal of Computer-Supported Collaborative Learning*, 10, 337-344.
- Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning*, 12, 35-64.

## Acknowledgments

We thank the students and the teacher who participated in this study. The research reported in this study is supported by a grant from the National Science Foundation (2010357) to the third and last authors.

# Inferential Detectors to Automate Feedback on Social Skills in the Wild: The Case of the Advice Hour Graphic Novel

Seth Corrigan, University of California, Irvine, scorrig1@uci.edu

**Abstract:** While youth may benefit from coaching on complex social skills, human-provided solutions may not always be available or scalable. Automating detection of and feedback on social skills, however, requires development of systems that can detect social cues and provide feedback in potentially noisy environments. Advice Hour is an online, interactive graphic novel that is capable of making inferences about social cues of unscripted speakers in uncontrolled environments, i.e., *in-the-wild*. The Advice Hour system is an example of an inferential detector developed through use of a straightforward three-phase approach that utilizes measurement and machine-learning processes. Importantly, its architecture is flexible, allowing for multiple models that can be localized, or even personalized to reflect the likely inferences of multiple groups or single individuals. This paper describes the Advice Hour experience, the data and process used to develop its inferential detectors, and their level of performance to date.

## Introduction

The present study seeks to describe a general and established approach to automating detection of social cues through description of a single example - Advice Hour. Advice Hour is an interactive online graphic novel in which readers assume the role of a podcast host responding to callers' questions about how to handle specific communication dilemmas in their personal and professional lives. The structure of the Advice Hour storyline allows for presentation of multiple caller scenarios. Each of the scenarios ends with a request prompting speaker-participants to record their own examples of how the given caller should respond. The Advice Hour architecture allows for flexible development and implementation of additional graphic novels that can be used to elicit and detect multiple types of social cues in order to provide practice and feedback.

The Advice Hour system records, transcribes and stores participants' responses through use of existing, secure cloud-based applications. In near-real time however, the system is capable of inferring the presence of one or more social cues through use of cloud-based inferential detectors developed by the research team. Importantly, the Advice Hour system was designed to operate *in-the-wild* using unscripted talk produced in naturally occurring settings such as participants' homes. This paper provides evidence such a system is feasible, describing the process used to develop the inferential detectors that automate detection of social cues and providing data describing their performances. Performance of the Advice Hour inferential detectors is compared to an existing benchmark that originates with the INTERSPEECH competitions (Schuller et al., 2012).

## Background

Inferential detectors, also referred to as virtual sensors, utilize a set of measurement processes that allow for quantification of properties usually determined by human perception or judgment (see Vallejo et al., 2019). There exist several examples of such inferential detectors (IDs). They have been developed to infer human impressions of objects and environments such as taste (Jiang et al., 2018), sound (Dal Palu et al., 2014), touch (Gee et al., 2005), glossiness (Leloup et al., 2014), and color and texture (Eugene, 2008). Inferential detectors have also been created to infer perceptions and judgements of others and their behavior, such as individuals' affective states (e.g., D'Mello & Graesser, 2010), instances of deceit (Bhaskaran et al., 2011; Ludwig et al., 2016) and flirtatiousness (Ranganath et al., 2009). The inferential detectors informing inferences for the Advice Hour experience utilize acoustic features of participants' speech such as pitch and amplitude; they also utilize the lexical content of the speech.

Development of inferential detectors generally requires three scopes of work: 1) selection and measurement of physical properties of a target object, agent, process, or environment (OAPE); 2) measurement of human perceptions or judgements of those properties; and 3) development of one or more models capable of creating a mapping between the two resulting sets of measured values. When all three scopes of work are carried out successfully, the resulting inferential detector(s) can be used to justify claims about how the target OAPE *would be* perceived or judged by a human observer, given the target's physical properties, even in the absence of such an observer.

## Methods

Data was collected under a cross-sectional design from expert and non-expert speakers who responded extemporaneously to prompts in naturalistic, uncontrolled environments. A stratified random sample of five-second audio-clips was drawn from the larger pool of participants' recordings and subsequently delivered for review and scoring by human raters. After raters' scores were modeled, the audio-clips judged to exhibit the highest (4<sup>th</sup> quartile) and those judged to exhibit the lowest (1<sup>st</sup> Quartile) levels of the two speech types were selected and used for development of a set of machine-learned models. The resulting models utilize acoustic and/or lexical features of speech to infer raters' judgements.

## Participants, recordings and ratings

The stratified random selection of five-second audio clips used for this study were drawn from recordings by expert ( $n = 101$ ) and non-expert ( $n = 154$ ) speakers. Nonexpert speakers reported that they had never trained in or engaged in professional voice-over acting; they were recruited using the Amazon Mechanical Turk system. Expert speakers reported some level of training or experience in acting or voice-over work, and were recruited using online markets Fiverr (<http://fiverr.com>) and Upwork (<http://Upwork.com>). Participants were evenly distributed with regard to reported gender. A total of 101 adult speakers (48 (47.5%) female and 53 (52.5%) male), 20-65 years of age ( $M = 40.6$ ,  $SD = 12.6$ ), participated in the study. Of the 154 non-expert speakers participating in the study, 67 (43.5%) were female and 87 (56.5%) were male, exhibiting a range of 18-55 years of age ( $M = 23.24$ ,  $SD = 3.94$ ).

Speaker participants were provided access to the Advice Hour's Online Recording System (ORS). This cloud-based recording tool presented speaking prompts embedded in an online graphic novel. The ORS requests access to the microphone on the user's computer and leads them through a series of checks to ensure the microphone is working correctly and that the noise level of the user environment is acceptable. Users then complete a series of recording prompts that are embedded in the Advice Hour graphic novel. After each recording, the Advice Hour ORS gives speakers a chance to review their recording and either accept or revise it.

Participants' recordings were reviewed by members of the research team for evidence of on-task performances. Recordings for each task were then segmented into five-second clips and indexed by a unique identifier indicating the speaker, speaker-type (expert versus non-expert), task, task type, and the window rank of the clip (a cardinal value denoting the position of the clip within the full recording).

A pool of three hundred and fifty-four ( $n = 354$ ) audio clips was reviewed and rated by all eight raters. The team was composed of four females and four males. Training followed a three-stage process focused on scoring a subset of data and adjudicating their scores in cases where there were disagreements, until achieving a minimal targeted level of agreement of 0.70 for each of the rating scales, using Cohen's weighted kappa (Cohen, 1968). Raters recorded their impressions of each audio clip by responding to a series of eight rating scales. The scales presented raters with four possible levels of response: from 0 (*Not at all*) to 3 (*Very much*).

## Data structure and data processing

Two sets of files were created and used for the eventual analyses presented here: one for estimation of measured values from raters' scores, using the faceted rating scale model (FRSM), and a second for development and testing of the planned machine learning models for classification of speech types—competence-focused and likability-focused speech. Acoustic features for each clip were extracted using the ComParE feature set within the OpenSMILE 3.0 library (Eyben et al., 2010). Lexical feature extraction followed Lee & Naratanan (2002).

## Outcome definition

Measured values of the targeted speech type were estimated using the faceted rating scale (Linacre, 1989). For detection of high and low levels of competence-focused speech, a single binary outcome label was created from the resulting measured values in order to indicate whether a given five-second audio clip was in the fourth quartile of measurements (high-level) or in the first quartile of measurements (low-level) for the given speech type. The same approach was used for detection of likability-focused speech.

## Sampling procedures

A total of 4,713 audio clips were gathered as a part of a larger study. Subsequently, a stratified random sample of  $n = 400$  audio clips was drawn from the full set for review and a final set of  $n = 356$  audio clips was used for rating and analysis. The sex of speakers at birth and their expertise level (non-experts versus actors) were used as strata to ensure balance across these two demographic characteristics. As noted above, members of the study team rated the resulting three hundred and fifty-six clips, indicating the extent to which they thought speakers emphasized

competence-focused and likability-focused speech. The rater scores were then converted to an interval level variable for both constructs through use of the faceted rating scale mode, and binned via quartiles. This resulted in two data files for further analyses: one for audio-clips exhibiting the highest or the lowest levels of competence-focused speech, and a second for audio-clips exhibiting the highest or the lowest levels of likability-focused speech.

## Data partitions

Data was partitioned using a 70:30 train-test split, with random selections made at the speaker level to avoid leakage of information between the resulting train and test data sets. Audio clips from a total of 53 speaker participants make up the file for competence-focused speech analyses, and audio-clips from 55 speakers make up the file for likability-focused speech analyses.

The training set for competence-focused speech represented a total of 112 audio clips; 51 (46%) clips are from the fourth quartile of scores; the remaining 65 (54%) of the audio clips are from the first quartile of scores. The testing set represented a total of 56 audio clips, with 31 (55%) from the fourth quartile of scores and 25 (45%) from the first quartile of scores. Labels were *high* and *low*.

The training set for likability-focused speech represented a total of 116 audio clips, 51 (44%) labeled high-level and 65 (56%) labeled low-level. The testing set represented a total of 50 audio clips, 33 (66%) labeled high-level and 17 (34%) labeled low-level. The samples were well balanced with regard to sex at birth but exhibited less balance with regard to speaker status—i.e., expert vs non-expert speakers.

## Modeling approach

The L1-logistic regression, support vector classifier, and support vector machine were each investigated as candidate approaches to create the desired mapping between the human rater impressions of the study's audio clips and their acoustic and lexical features. The models were developed within a ten-fold cross validation framework. Unweighted average recall (UAR) and the Area Under the Receiver Operator Curve (AUC) were used to evaluate performance.

## Results

**Table 1a**  
*Results for Classification of High/Low Competence-Focused Speech*

	Acoustic Only				Lexical Only				Acoustic + Lexical			
	sens	spec	uar	auc	sens	spec	uar	auc	sens	spec	uar	auc
L1	0.839	0.625	0.746	0.820	0.677	0.625	0.655	0.669	0.742	0.625	0.691	0.766
SVC	0.742	0.667	0.709	0.728	0.548	0.667	0.600	0.676	0.839	0.630	0.746	0.788
SVM	0.615	0.880	0.719	0.862	0.583	0.645	0.618	0.687	0.774	0.667	0.727	0.852

*Legend.* sens: sensitivity; spec: specificity; uar: unweighted average recall; AUC: area under the curve.

**Table 1b**  
*Results for Classification of High/Low Likability-Focused Speech*

	Acoustic Only				Lexical Only				Acoustic + Lexical			
	sens	spec	uar	auc	sens	spec	uar	auc	sens	spec	uar	auc
L1	0.636	0.529	0.600	0.521	0.758	0.412	0.640	0.683	0.606	0.647	0.620	0.740
SVC	0.515	0.647	0.560	0.726	0.576	0.529	0.560	0.668	0.667	0.588	0.640	0.672
SVM	0.576	0.706	0.620	0.668	0.529	0.576	0.560	0.672	0.636	0.941	0.740	0.850

Test set results of the three different modeling approaches for classification are summarized in Table 1a and Table 1b. The best performances were achieved by the support vector classifier and the support vector machine with radial kernel, using both acoustic and lexical features. In the competence-focused speech classification task, using both the acoustic and the lexical features, the SVC had an unweighted average recall of 0.746 and an AUC of 0.788. Sensitivity was 0.839 and specificity was 0.630. In the likability-focused speech classification task, the SVM had a UAR of 0.740 and an AUC of 0.850. The sensitivity was 0.636 and specificity was 0.941. The unweighted average recall of the models was improved through use of acoustic *and* lexical features in most cases, but not all.

## Discussion

Using the Advice Hour speech data and human ratings presented in the current study, inclusion of lexical and acoustic features of speech as inputs to the study's classification models results in performances that favorably compare to Schuller et al. (2012) benchmarks mentioned above. Performance of the Advice Hour's inferential detectors provides an existence proof for automated detection of social cues *in-the-wild*, i.e., in contexts in which ad hoc speech is gathered in naturally occurring settings.

## References

- Bhaskaran, N., Nwogu, I., Frank, M. G., & Govindaraju, V. (2011, March). Lie to me: Deceit detection via online behavioral learning. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 24-29). IEEE. <https://doi.org/10.1109/fg.2011.5771407>
- Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, *70*(4), 213. <https://doi.org/10.1037/h0026256>
- D'mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, *20*(2), 147-187. <https://doi.org/10.1007/s11257-010-9074-4>
- Dal Palù, D., De Giorgi, C., Astolfi, A., Lerma, B., & Buiatti, E. (2014). SounBe, a toolkit for designers dealing with sound projects. In *DS 77: Proceedings of the DESIGN 2014 13th International Design Conference* (pp. 2011-2022). <https://www.designsociety.org/publication/35342/SOUNBE%2C+A+TOOLKIT+FOR+DESIGNERS+DEALING+WITH+SOUND+PROJECTS>
- Eugène, C. (2008, September). Measurement of "total visual appearance": a CIE challenge of soft metrology. In *12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement* (pp. 61-65). Eyben, F., Batliner, A., & Schuller, B. (2010, April). Towards a standard set of acoustic features for the processing of emotion in speech. In *Proceedings of Meetings on Acoustics 159ASA* (Vol. 9, No. 1, p. 060006). Acoustical Society of America. <http://dx.doi.org/10.1121/1.4739483>
- Gee, M. G., Tomlins, P., Calver, A., Darling, R. H., & Rides, M. (2005). A new friction measurement system for the frictional component of touch. *Wear*, *259*(7-12), 1437-1442. doi:10.1016/j.wear.2005.02.053
- Jiang, H., Zhang, M., Bhandari, B., & Adhikari, B. (2018). Application of electronic tongue for fresh foods quality evaluation: A review. *Food Reviews International*, *34*(8), 746-769.
- Lee, C. M., Narayanan, S. S., & Pieraccini, R. (2002, September). Combining acoustic and language information for emotion recognition. In *INTERSPEECH*.
- Leloup, F. B., Obein, G., Pointer, M. R., & Hanselaer, P. (2014). Toward the soft metrology of surface gloss: A review. *Color Research & Application*, *39*(6), 559-570.
- Linacre, J. M. (1989). *Many-faceted Rasch measurement* (Doctoral dissertation, The University of Chicago).
- Ludwig, S., Van Laer, T., De Ruyter, K., & Friedman, M. (2016). Untangling a web of lies: Exploring automated detection of deception in computer-mediated communication. *Journal of Management Information Systems*, *33*(2), 511-541. <https://doi.org/10.1080/07421222.2016.1205927>
- Ranganath, R., Jurafsky, D., & McFarland, D. A. (2013). Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates. *Computer Speech & Language*, *27*(1), 89-115. <https://doi.org/10.1016/j.csl.2012.01.005>
- Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., van Son, R., Weninger, F., Eyben, F., Bocklet, T., Mohammadi, G., & Weiss, B. (2012). The INTERSPEECH 2012 speaker trait challenge. In *INTERSPEECH 2012, Portland, OR, USA*. <https://doi.org/10.21437/interspeech.2012-86>
- Vallejo, M., De La Espriella, C., Gómez-Santamaría, J., Ramírez-Barrera, A. F., & Delgado-Trejos, E. (2019). Soft metrology based on machine learning: a review. *Measurement Science and Technology*, *31*(3), 032001. <https://doi.org/10.1088/1361-6501/ab4b39>

## Acknowledgments

This work was funded by the William and Flora Hewlett Foundation. Special thanks and recognition to Aysegül Askan for preparing the data frames for this project and Jordan Suhr for bringing his excellent project management skills to this work; thanks also to the larger project team for their time and effort reviewing and rating the recordings for this project.

# Behavior Sequence Patterns in Collaborative Problem Solving: A Multiple-perspective Analysis

Mengxiao Zhu, University of Science and Technology of China, mxzhu@ustc.edu.cn  
Xin Wang, University of Science and Technology of China, wangx804@mail.ustc.edu.cn  
Qizhi Xu, University of Science and Technology of China, fred@ustc.edu.cn  
Jessica Andrews-Todd, Educational Testing Service, jandrewstodd@ets.org

**Abstract:** Collaborative problem solving (CPS), as a complex but important competence, needs to be explored in depth. However, a few studies conducting cluster analysis were mainly based on behavior frequencies, which cannot capture the temporal and dynamic sequence process of CPS. In this study, sequence-based cluster analysis was used to divide 119 triad teams into two clusters. Then, process mining analyses, including the first-order Markov model and epistemic network analysis, were conducted to explore the behavior characteristics of the two clusters. The results showed that teams with active actions, such as enacting plans, communicating, or reminding teammates, had greater performance than teams showing passive actions, such as role identification and task clarification.

## Introduction

Collaborative problem solving (CPS) has been recognized as a pivotal skill, serving as a catalyst for successful job performance and active learning. The acknowledgment of CPS's significance has spurred researchers to delve into its assessment, analysis, and development. Regarding analysis, some existing studies have used machine learning techniques, such as cluster analysis, to explore the behavior characteristics of students in CPS activities based on process data (e.g., Andrews-Todd et al., 2018). However, most of these studies used cluster analysis based on behavior frequencies and ignored the timing and sequences of behaviors. Such analysis fails to capture the interactive and dynamic characteristics during CPS activities (Ouyang et al., 2023). Additionally, some research considered the sequence information and conducted process mining analysis on pre-defined subgroups with different performances (e.g., Chang et al., 2017). Nevertheless, such comparisons were constrained by the subgroup boundaries and did not allow exploration of process patterns over all samples. To bridge the gap, this study employs a sequence-based cluster method over the whole dataset and to mine behavior patterns in different clusters. Comparisons of the clusters revealed significant differences in both behavior patterns and performances.

## Literature review

Collaborative problem solving (CPS) competence is important for achieving career and academic success (Andrews-Todd et al., 2018). CPS-related activities entail rich information on individual behaviors as well as interactions among team members, leading to increased analytical complexity. Therefore, researchers pay attention to analyzing CPS process data, shifting their focus from determining "what the result is" to understanding "how the result is produced". In online CPS platforms, the logging systems record process data, including action behaviors on the platform and communications with teammates generated during the task with timestamps.

Since process data are usually unstructured, researchers tend to first code process data based on certain coding schema to transform them into structured data. Then, data-driven methods, such as machine learning or process mining, are employed to explore how students collaborate to solve problems. Regarding machine learning, cluster analysis is the most commonly used method in existing research. For instance, Andrews-Todd et al. (2018) identified four solver profiles based on behavior frequencies. As another example, Zhu et al. (2023) conducted cluster analysis based on behavior frequencies and found two types of collaborative patterns. Moreover, they explored the interactive impact of domain knowledge and CPS subskills on performance. Nevertheless, frequency-based cluster analysis cannot capture the behavior dynamics during the task.

To consider the sequence information in process data, researchers adopted process mining analysis, an event-based data analysis method that extracted sequential, associative, and temporal insights from process data. For example, Chang et al. (2017) employed lag sequential analysis to identify problem solving strategies that can promote team performance. Zhang et al. (2022) used epistemic network analysis to reveal relationships between different skills. However, these studies mainly focused on comparing high and low performing groups and ignored important diversity within groups (Saqr & López-Pernas, 2023). A combination of cluster analysis with process mining analysis may help us to better discover behavior patterns without enforcing pre-defined groups.

The fusion of cluster analysis and process mining methods has found application to process data analysis on other topics. Regarding the topic of learning analytics, Matcha et al. (2019) employed this integrated approach to discern students' learning strategies during online pre-class activities in a flipped classroom setting and further delved into the correlation between these strategies and academic performance. Similarly, Saqr et al. (2023) utilized cluster analysis to categorize students into distinct four groups during online problem-based learning activities. By employing process mining, they compared interaction patterns among these groups, elucidating how different group dynamics influence learning outcomes. These studies leverage cluster analysis, rather than pre-defined categorizations, to unveil nuanced insights into interaction sequences that may predict higher achievement.

To investigate whether and what kind of behavior patterns contribute to better team performance in CPS activities, this study employs sequence-based cluster analysis and process mining to reveal the behavior patterns exhibited by different clusters in CPS tasks. Regarding process mining, we use the first-order Markov model, a commonly used method in sequence analysis that leverages time and probability metrics, to uncover potential transition paths among behaviors. Additionally, we conduct epistemic network analysis to elucidate the interactivity dynamics inherent among behaviors. The two analyses take into account directed and undirected relations between behaviors, respectively, revealing behavior patterns from different perspectives. In summary, this study applies multiple analytical approaches to address the following two critical research questions.

- Q1. How many clusters with different behavior sequences can be identified from the process data in the CPS task?*  
*Q2. How do these identified clusters differ in their behavior patterns and performances?*

## Methodology

### Tasks and participants

This study utilized the three-resistor task as the CPS activity. The task required a team of three individuals to complete within a given class period. The task involved adjusting the resistance value of the respective resistor in a series circuit to reach the target voltage. Thus, effective communication and collaboration among teammates were crucial for performing the task. Additionally, the task featured four different levels, with increasing levels of difficulty. Each team was encouraged to complete as many levels as possible.

A total of 378 participants were recruited from universities and community colleges in the USA, aged between 18 and 20. They were randomly assembled into 126 teams. Eventually, 110 teams completed Level 1 and proceeded to Level 2, 87 teams participated in Level 3, and 25 teams participated in Level 4.

### Dataset

In total, the logging system recorded 51,805 pieces of data, comprising 34,867 action data records (e.g., person A changes his resistor from  $3.50\Omega$  to  $3.63\Omega$  via keypress) and 15,950 communication data records (e.g., person A types that "we could experiment with the higher resistances"). To get structured data, two trained coders manually encoded the collected process data based on the PISA 2015 framework (OECD, 2017). The framework consisted of two core dimensions (i.e., problem solving and collaborative), with the two dimensions interacting to form a matrix format of 12 subskills. The problem solving dimension included four sub-dimensions, (A) *exploring and understanding*, (B) *representing and formulating*, (C) *planning and executing*, and (D) *monitoring and reflecting*. The collaborative dimension included three sub-dimensions, (1) *establishing and maintaining a shared understanding*, (2) *taking appropriate action to solve the problem*, and (3) *establishing and maintaining team organization*. One of the interacting subskills was (C2) *Enacting plans*. Additionally, 7 teams were excluded due to loss of coding, resulting in a dataset of 119 teams for subsequent analysis.

### Analysis methods

We employed a multiple-perspective analysis to gain insights into characteristics of behavior patterns based on the coded process data. To address the first question, we employed cluster analysis to categorize teams' behavior sequences. Specifically, to assess the similarity of behavior sequences among different teams, we utilized the optimal matching (OM) algorithm, which incorporates the Levenshtein distance as a measure of edit distance. Then, we employed Ward's clustering (WC) to cluster. The algorithm was executed using the R package *cluster*.

Following the clustering process, we conducted quantitative analysis and process mining to explore the behavior patterns of various clusters to answer the second question. In quantitative analysis, we employed t-tests to compare the performance and prior knowledge of different clusters. Regarding process mining, we used the first-order Markov model (FOMM) and epistemic network analysis (ENA) to reveal behavior dynamics during the task. We implemented FOMM using the *pMineR* in R package and conducted ENA through the use of an ENA Webkit available at [epistemicnetwork.org](http://epistemicnetwork.org).

## Results

In this section, we presented the cluster results derived from the behavior sequences of 119 teams, and the results of quantitative analysis and process mining analysis.

### Cluster and basic quantitative analysis

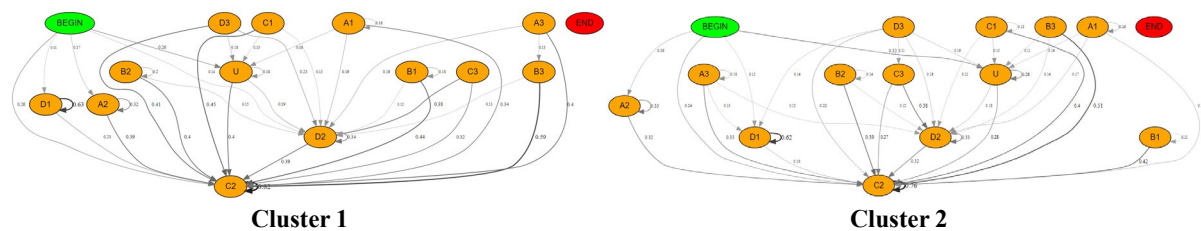
The cluster results led to the classification of teams into two distinct types, referred to as Cluster 1 and Cluster 2. Cluster 1 comprised 34 teams, while Cluster 2 encompassed 85 teams. The independent-sample *t*-test revealed that Cluster 1 ( $M = 2.22, SD = 0.91$ ) outperformed Cluster 2 ( $M = 1.83, SD = 1.02$ ) in completed task levels, with  $t = 1.99, p = .049$ , and effect size  $d = 0.21$ . Regarding the domain knowledge related to the circuit concept, no significant difference was observed between the two clusters ( $t = -0.07, p = .95$ ; Cluster 1,  $M = 15.28, SD = 3.34$ ; Cluster 2,  $M = 15.32, SD = 3.74$ ). Therefore, the performance disparities between the two clusters were more likely attributable to differences in their behavior patterns during the task-solving processes.

### Process mining analysis

Figure 1 shows the transition probabilities with more than 0.1 between subskills in each cluster based on the first-order Markov model. In this representation, each node corresponds to a subskill, and a directed edge signifies the transition from one subskill to another. The results revealed that Cluster 1 exhibited higher transition probabilities from other subskills (e.g., A3, understanding roles to solve problems to C2, enacting plans). Conversely, Cluster 2 displayed a higher transition probability from A3 to A3. These findings suggested that Cluster 1 can be characterized as active actionists due to their frequent transitions to enact plans, indicating a proactive approach in communicating, gathering task clues, and taking action to solve the problem. In contrast, Cluster 2 tends to be passive actionists due to their dependency on teammates' roles, more relying on teammates' actions and suggestions, and not taking the initiative in acting. In summary, the first-order Markov analysis displays the distinct state transition patterns exhibited by the two clusters in the CPS task.

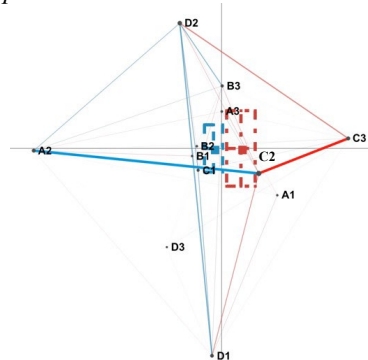
**Figure 1**

*The First-Order Markov Model Results in the Two Clusters*



**Figure 2**

*The Difference in the Two Clusters of the Epistemic Network*



The results of epistemic network analysis are shown in Figure 2, which demonstrates the difference in ENA structure between the two clusters. In the network, nodes refer to CPS subskills and undirected edges represent the relative frequency of co-occurrence of two subskills in adjacent actions or communications (Zhang et al., 2022). Additionally, the red edges represent the more frequent co-occurrence of subskills in Cluster 1 than Cluster 2, and the blue edges represent the more frequent co-occurrence of subskills in Cluster 2 than Cluster 1. Furthermore, the red square represents the centroid (i.e., the mean position of all projection nodes) of Cluster 1,



and the blue square corresponds to the centroid of Cluster 2. The difference of the two centroids can be represented by the difference in the horizontal and vertical directions. A Mann-Whitney U test indicates a significant difference ( $t = 3.27, p < .001, d = 0.66$ ) between the two clusters on the horizontal axis (Cluster 1,  $M = 0.25, SD = 0.53$ ; Cluster 2,  $M = -0.10, SD = 0.54$ ). Overall, the results showed that Cluster 1 exhibited stronger connections between subskills C2 (enacting plans) and C3 (following rules of engagement, e.g., reminding other members to complete their tasks), whereas Cluster 2 displayed stronger connections between A2 (discovering the type of collaborative interaction required and establishing goals) and C2. This further demonstrated that Cluster 1 comprised active actionists who communicate with teammates and take actions actively to solve problems, whereas Cluster 2 was more dependent on teammates to passively take part in tasks.

## Discussion

This study clusters students' behavior sequences during a CPS task by incorporating the temporal factor and considering the dynamic nature of behavior patterns. It employs process mining methods, including the first-order Markov model and epistemic network analysis, to investigate the differences between the identified clusters on the transition probabilities and co-occurrences between CPS subskills by modeling the connections of the subskills as either directed or undirected networks. It was found that teams in Cluster 1 engaged in more proactive communication and action, contributing to better performance in the task. In contrast, teams in Cluster 2 were more passive and relied on the roles of teammates, resulting in slightly poorer performance. Different from previous studies (e.g., Ouyang et al., 2023), this study not only illustrates the role of action and communication in CPS tasks, but also underscores the importance of proactive actions.

This study also has some pedagogical and research implications. Regarding pedagogy, it is necessary to encourage students to actively participate in collaborative processes to ensure more smoother problem solving processes and better performance. Active communication is beneficial for organizing teams, and active action can prompt team interactions toward higher performance. In terms of research implications, this study used multiple-perspective analysis methods to understand teams' behavior patterns in CPS tasks, which can better extract and process the complex and dynamic structure of CPS activities, and increase the interpretability of clustering results.

## References

- Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior, 104*, 105759.
- Chang, C. J., Chang, M. H., Chiu, B. C., Liu, C. C., Chiang, S. H. F., Wen, C. T., Hwang, F. K., Wu, Y. T., Chao, P. Y., Lai, C. H., Wu, S. W., Chang, C. K., & Chen, W. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education, 114*, 222-235.
- Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pp. 461-470.
- OECD. (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving. *OECD Publishing*, Paris.
- Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 39-66.
- Saqr, M., & López-Pernas, S. (2023). The temporal dynamics of online problem-based learning: Why and when sequence matters. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 11-37.
- Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education, 183*, 104485.
- Zhu, M., Xu, Q., Dou, Z., Zhang, L., & Andrews-Todd, J. (2023). Exploring the interplay between domain knowledge and collaborative problem solving skills through user profiling. In *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning-CSCL 2023*, pp. 400-401.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (62177044).

# The Relationships between Students' Epistemological Beliefs and Social Interactions and Epistemic Actions in Collaborative Inquiry Practice

Xiaojie Zhang, The University of Hong Kong & Southern University of Science and Technology,  
zxjie@connect.hku.hk

Shihui Feng, The University of Hong Kong, shihuiife@hku.hk

Carol Chan, The University of Hong Kong, ckkchan@hku.hk

Jianhua Zhao, Southern University of Science and Technology, zhaojh@sustech.edu.cn

**Abstract:** Students' epistemological belief about how ideas develop is crucial as it helps them better understand how they construct knowledge in collaborative learning. This study investigated how primary students' epistemological beliefs relate to collaborative inquiry activities in an online knowledge-building environment using data analytics approaches. Findings based on social network analysis and epistemic network analysis on students' online discourse suggested that students' epistemological beliefs influenced their knowledge-building activities. Students with higher-level epistemological beliefs engaged more in productive and effective community inquiry activities; they are more active in reading others' ideas and proposing diverse ideas for idea improvement in a collaborative community. The study contributes to understanding students' views on the nature of knowledge building and the influences on students' social interactions and epistemic actions in collaborative inquiry online environments.

## Introduction and background

Advancing technology and global challenges underscore the need for education systems that foster young people to engage in the civilization of knowledge creation, moving beyond mere knowledge acquisition in classrooms. Emphasizing computer-supported collaborative inquiry learning, students are enabled to participate in high-level knowledge creation activities, enhancing sustained high-quality knowledge development. Research indicates that students' epistemological beliefs — understandings of how knowledge creation developed — affect their collaborative inquiry processes and learning outcomes (Bråten, 2016). However, most studies in learning science employ interviews and questionnaires to assess epistemological beliefs (Gok, 2018), and very few studies assess epistemological beliefs by non-self-report approaches.

According to Popper (1979), there are three worlds: the physical/material world (world 1), the psychological world existing in the human mind (world 2), and the reality primarily composed of ideas (world 3). World 3 consists of ideas created by knowledge workers such as engineers, scientists, designers, and architects. These ideas, once generated, take on a tangible form within the social realm and can be further explored, modified, and experimented with by other knowledge agents. From a World 3 perspective, ideas are considered tentative knowledge claims and should undergo continuous critical examination, clarification, and potential falsification for further development. Knowledge building (KB) as a pedagogical approach is influenced by Popper's three-world epistemological position of constructivism. *Knowledge Building* is defined as a community-based collective learning process highlighting sustained theory generation and idea improvement (Scardamalia & Bereiter, 2016). To foster a thriving knowledge-building community, students need to recognize the role of ideas as World 3 objects that can be improved collectively for advancing knowledge. This means perceiving ideas as malleable entities open to scrutiny and refinement within the community to drive progress (Scardamalia, 2002). In KB, Students' understanding of 'ideas' has been categorized as epistemological beliefs using the framework of World 3 (Hong, 2016). Students who think of ideas as mental representations in individual minds were identified as having low epistemological beliefs (World 2), and students who see ideas as collective creations generated from group discussion were identified as having high epistemological beliefs (World 3). The present study will use the framework to characterize students' epistemological beliefs.

Instructional design has been proven to influence student epistemological beliefs (Ding & Mollohan, 2015). In KB, students' epistemological beliefs and epistemic discourse understanding can be improved by engaging them in a constructivist way of collaborative knowledge building (Hong et al., 2009; Tong & Chan, 2023). However, few studies investigated the relationship between students' epistemological beliefs and social-cognitive practice in knowledge-building environments. Learning analysis approaches have been used to unpack social-cognitive practice in KB. Previous studies have used the social networks analysis (SNA)' indicators of "who

*read who*" and "*whose notes built on whose notes*" to reflect knowledge diffusions, such as connection network density, centralization, and average degrees of separation (Oshima et al., 2012, p. 910). Epistemic network analysis (ENA) can capture, visualize, and quantitatively compare patterns of learning activities across conditions and be used with smaller datasets (Shaffer et al., 2016). Previous studies showed that ENA is a promising approach to assessing students' epistemic actions by providing more profound, valid quantitative ethnographies in knowledge-building discourse (Ma et al., 2019). Therefore, this study aims to explore students' different views about ideas depicting the three-world framework and investigate how this epistemological belief relates to students' collaborative inquiry on Knowledge Forum, illustrating collective social interaction and epistemic actions assessed with SNA and ENA. Two specific research questions are:

RQ1: What are the relationships between students' epistemological beliefs and social interactions in collaborative inquiry learning?

RQ2: What are the relationships between students' epistemological beliefs and epistemic action patterns in collaborative inquiry learning?

## Method

Based on knowledge-building pedagogy, thirty 5th-grade students from a primary school in China participated in a collaborative inquiry activity about *Light*. As an online platform, Knowledge Forum® (<http://kf6.edu.hku.hk>) was used to support the knowledge-building processes (Scardamalia & Bereiter, 2014). Knowledge Forum allows students to publish their ideas as *notes*, "*build on*" existing notes to post their further questions or improve ideas, and reflect on the progress of ideas improvement. The data contains epistemological belief writing, KF notes, KF log data, and interviews.

Students' writing about what they thought about 'ideas' (e.g., "how can ideas be improved") was characterized as epistemological belief, according to Popper's World 3 theory, based on the coding framework in a previous study (Hong, 2016, p. 38). Nine students who think of ideas as mental representations in individual minds were identified as having low-epistemological beliefs (World 2), and the other 21 students who see ideas as collective creations generated from group discussion were identified as having high-epistemological beliefs (World 3). Students wrote 217 notes, which were analyzed using a coding scheme adapted from a knowledge-building coding framework (van Aalst, 2009) (Table 2). The coding scheme was adjusted to reflect students' epistemic actions in this study, containing *My Question (Q)*, *My Idea (I)*, *Further Question (FurQ)*, *Improved Idea (ImpI)*, *My Argument (Arg)*, and *Extraneous information (Ext)*.

For Research Question 1, the matrix of "*who read who*" and "*whose notes built on whose notes*" explored from the Knowledge Forum log data was transferred to social networks. In the networks, the nodes represent students, and the edges represent which students read or build on which students' notes. In SNA, some crucial node measurements can capture individual participant social interactions – weighted degree, closeness centrality, betweenness centrality, and local clustering coefficient (Knoke & Yang, 2020). Weighted degree centrality accounts for the weight of a node's edges in the network. In this study, weighted in-degree means how many times one student's notes were read/built on by others, and weighted outdegree means how many times one student read/built on others' notes. Closeness centrality and betweenness centrality are related to the potential of control over communication. In this study, closeness and betweenness centrality denotes the distance from one student to other students in the network. The local clustering coefficient measures the degree to which neighboring nodes of a node are also interconnected. The larger the local clustering coefficient, the more cliquish the network and the less likely it is to diffuse information. To examine the social interactions of individual students within communities, each student's SNA node measurements in the two networks were calculated by Gephi 0.10. Then, Mann Whitney U was employed to compare the SNA measurements of the two groups, depicting different patterns of epistemological understanding.

For Research Question 2, encoded KB epistemic actions data was analyzed using the ENA web tool (<http://app.Epistemic.network.org>). Students' ID and epistemological category were used as *units* to construct ENA networks. Students' notes under the same discussion topics were identified as *conversations* to analyze the co-occurrence of epistemic actions. The co-occurrences of KB epistemic actions in conversations were converted into adjacency matrices. Then, cumulative adjacency matrices were expressed as adjacency vectors, which were spherically normalized. The spherically normalized adjacency vectors were reduced into several dimensions via singular value decomposition for the visualization.

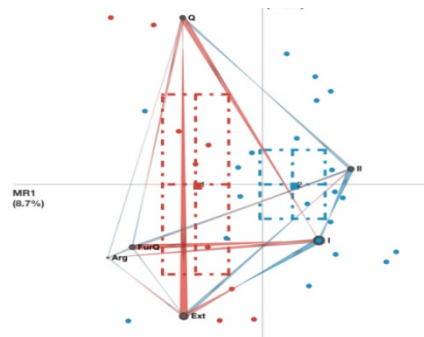
## Results

Mann Whitney-U test was employed to compare "*who read who*" network measurements among high/low epistemological belief students. The results suggest that students with higher epistemological beliefs have higher mean weighted outdegrees than those with lower epistemological beliefs ( $n = 21$ , mean = 41.29;  $n = 9$ , mean =

28.78). Moreover, students with higher epistemological beliefs have a statistically significantly lower local clustering coefficient than those with lower epistemological beliefs ( $n = 21$ , mean = 0.55;  $n = 9$ , mean = 0.57,  $U = 45.5$ ,  $p = 0.027$ ). That means students with higher epistemological beliefs read more notes and tended to read notes from different individuals. In other words, those students were aware to read different ideas from the class-level community.

Mann Whitney-U test was employed to compare “*who build-on who*” network measurements among high/low epistemological belief students. The results suggest that students with higher epistemological beliefs have lower weighted outdegrees than those with lower epistemological beliefs ( $n = 21$ , mean = 3.90;  $n = 9$ , mean = 6.67). Moreover, students with higher epistemological beliefs have lower closeness centrality ( $n = 21$ , mean = 0.30;  $n = 9$ , mean = 0.41) and betweenness centrality ( $n = 21$ , mean = 30.29;  $n = 9$ , mean = 63.89) than students with lower epistemological beliefs. That means students with higher epistemological beliefs build on less, but they tend to build on notes from different individuals. In other words, those students are aware of the need to contribute different ideas from the class-level community.

**Figure 1**  
*The ENA Results*



*Note. Q: My Question, I: My Idea, FurQ: Further Question, ImpI: Improved Idea, Arg: My Argument, Ext: Extraneous information. Group 1 and Group 2 were shown in blue and red separately. The squares represent the ENA network's centroids (mean position of the projected points) in each phase. The dots mean students' centroids in the ENA space. The edges in the ENA networks indicate the co-occurrence of epistemic actions in four temporally adjacent notes. The thicker the edge, the more times the KB processes co-occur. The arrows indicate the transition direction. These findings suggest that students with World 3 views about ideas also tend to engage in more sophisticated epistemic actions in collaborative inquiry on Knowledge Forum.*

Significant differences existed between the two groups' ENA networks (see Figure 1). Along the X axis, a two-sample t-test assuming unequal variance showed 1 (mean=-0.48, SD=0.31, N=9) was statistically significantly different at the  $\alpha=0.05$  level from 2 (mean=0.22, SD=0.50, N=20;  $t(23.96) = 4.57$ ,  $p=0.00$ , Cohen's  $d=1.54$ ). The ENA results demonstrated that students with lower epistemological beliefs tended to have co-occurrence from "My Idea" to "My Question" and from "My Question" to "Extraneous information." They did not read other's notes, so they proposed many repetitive questions, which other students answered. Moreover, they posted more content unrelated to the inquiry than students in group 2. Students with higher epistemological beliefs posted more "Improved Idea" and "Argument" after "My Question" and "My idea." They tended to improve the existing ideas with a more detailed explanation: "Light does indeed propagate in curves, but its curves are so small that they cannot be seen." Moreover, they challenge others' ideas by posting, "If light propagates in a curve, there will be no shadows," and "If you do not believe in microwaves, then do not use microwaves."

## Discussion and conclusion

The social interactions and ENA results revealed differences between the students' epistemic views toward the notions of ideas in knowledge building. Students with high epistemological belief engaged more in reading others' ideas, contributing diverse ideas, and improving existing ideas, while students with low epistemological belief read fewer others' ideas, contributed more repeat ideas, and tended to some extraneous epistemic actions. In other words, high-epistemological students made the community's ideas more diverse and profound. In contrast, low-epistemological students tended to express their ideas but ignored the improvement of community ideas. The results suggest that high epistemological students have more of a Popper's World 3 perspective and community consciousness (Popper, 1979). They believe ideas are tentative knowledge claims and should undergo continuous

critical examination, clarification, and potential falsification for further development. A student said in the interview, "Some of my ideas have already been posted by others, so I did not want to repeat them." "I just thought I could post more diverse ideas and share the ideas and knowledge with the community."

Previous research has shown that students' epistemological beliefs impact learning, thinking, and understanding (Atman & Yildiz, 2022). However, the previous studies mostly used interviews and survey data to present students' epistemological beliefs. This study contributes to broader ideas of epistemology, including examining students' epistemic views about ideas and using data analytics methods to show how different knowledge-building processes influenced and co-evolved with epistemic views of ideas. Reading ideas from different participants in the community and contributing diverse ideas are the typical features of students with community awareness while engaging in collaborative inquiry. Through the ENA analysis of students with different epistemic stances, this paper also dynamically shows the epistemic actions and details the relationship between epistemic actions. The transition from individually held and existing ideas to collectively generated improved ideas and arguments in public space is crucial to sustained idea improvement in collaborative inquiry and knowledge building. Future research will analyze the dynamic changes in students' epistemological views and their influence on social networks and epistemic actions in connection with profound conceptual changes in collective knowledge building.

## References

- Atman Uslu, N., & Yildiz Durak, H. (2022). The relationships between university students' information-seeking strategies, social-media specific epistemological beliefs, information literacy, and personality traits. *Library & Information Science Research*, 44(2), 101155. <https://doi.org/10.1016/j.lisr.2022.101155>
- Bråten, J. A. G., William A. Sandoval, Ivar (Ed.). (2016). *Handbook of Epistemic Cognition*. Routledge. <https://doi.org/10.4324/9781315795225>
- Ding, L., & Mollohan, K. N. (2015). How College-Level Introductory Instruction Can Impact Student Epistemological Beliefs. *Journal of College Science Teaching*, 44(4), 19–27.
- Gok, T. (2018). The Evaluation of Conceptual Learning and Epistemological Beliefs on Physics Learning by Think-Pair-Share. *Journal of Education in Science Environment and Health*, 4(1), Article 1.
- Hong, H.-Y., Chai, C. S., & Lin, S.-P. (2009, November). *Effects of knowledge-building on teacher-education students' epistemological belief change*. <https://repository.nie.edu.sg/handle/10497/14496>
- Knoke, D., & Yang, S. (2020). *Social Network Analysis*. SAGE Publications, Inc. <https://doi.org/10.4135/9781506389332>
- Ma, L., Matsuzawa, Y., & Scardamalia, M. (2019). *Using Epistemic Network Analysis to Explore Ways of Contributing to Knowledge Building Discourse*. <https://repository.isls.org/handle/1/1710>
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60(5), 903–921. <https://doi.org/10.1007/s11423-012-9265-2>
- Popper, K. R. (1979). *Objective knowledge: An evolutionary approach*. Clarendon press.
- Scardamalia, M. (2002). Collective Cognitive Responsibility for the Advancement of Knowledge. In B. Smith & C. Bereiter (Eds.), *Liberal education in a knowledge society* (pp. 67–98).
- Scardamalia, M., & Bereiter, C. (2014). Knowledge Building and Knowledge Creation: Theory, Pedagogy, and Technology. In R. K. Sawyer, *Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). Cambridge University Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), Article 3. <https://doi.org/10.18608/jla.2016.33.3>
- Tong, Y., & Chan, C. K. K. (2023). Promoting knowledge building through meta-discourse and epistemic discourse understanding. *International Journal of Computer-Supported Collaborative Learning*, 18(3), 353–391. <https://doi.org/10.1007/s11412-023-09403-2>
- van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 259–287. <https://doi.org/10.1007/s11412-009-9069-5>

## Acknowledgment

This study is supported by the National Social Science Foundation of China (Grant No. BCA200090).

# Supporting Pre-Service Teachers' Evidence-Informed Reasoning Through Peer-Feedback: Effects of Feedback Provision and Feedback Integration Scaffolds

Julia Hornstein, Martin Greisel, Ingo Kollar  
julia.hornstein@uni-a.de, martin.greisel@uni-a.de, ingo.kollar@uni-a.de  
University of Augsburg

**Abstract:** Having pre-service teachers provide and receive peer-feedback on their analyses of authentic classroom problems may help them acquire evidence-informed reasoning skills. However, without instructional guidance, students may struggle to provide high-quality feedback and to integrate multiple feedback messages. We investigated the impact of feedback provision and feedback integration scaffolds and their combination on (a) pre-service teachers' improvement from draft to revision, and (b) their acquisition of evidence-informed reasoning skills.  $N = 254$  pre-service teachers analyzed a classroom case vignette by aid of educational theories, provided feedback to two peers, and revised their initial analyses based on the feedback they received. Neither the feedback provision scaffold nor the feedback integration scaffold had a significant effect on the improvement. For the acquisition of evidence-informed reasoning skills, there was a significant negative interaction effect. This suggests that the scaffolds need to be better synchronized so that their combination yields additional effects.

## Aims of the study

Teachers are increasingly required to solve teaching problems in accordance with educational theories and findings (Dekker & Meeter, 2022). Yet, studies revealed that pre-service teachers often struggle with evidence-informed reasoning when given the task to solve authentic classroom problems (Kierner & Kollar, 2021). One possible way to support pre-service teachers' acquisition of evidence-informed reasoning skills might be to engage them in mutual peer-feedback on their analyses of authentic classroom cases. By providing feedback to each other, students are actively involved in knowledge construction processes (e.g., Double et al., 2020). Further, integrating the feedback they receive from their peers into their own analysis can be considered an important prerequisite for knowledge and skill acquisition (Chi & Wylie, 2014). Nonetheless, prior studies demonstrated that pre-service teachers require support in delivering high-quality feedback and integrating multiple feedback messages. The current study looks into ways how to best instructionally scaffold the peer-feedback process, with a specific emphasis on the effects of a feedback provision scaffold and a feedback integration scaffold on the extent to which pre-service teachers increase the quality of their initial analyses, and on the acquisition of evidence-informed reasoning skills.

## The peer-feedback process and its potential for learning

Peer-feedback is a reciprocal process in which learners evaluate each other's performance (e.g., Hattie & Timperley, 2007). It includes four phases: First, in the *task performance phase*, the learners, usually working individually, carry out an assignment focused on a particular subject. Subsequently, in the *feedback provision phase*, learner A assesses the quality of B's performance, and vice versa. The ensuing *feedback reception phase* encompasses both learners receiving and assimilating feedback from each other. Finally, in the *revision phase*, the learners revise their task solutions based on the feedback they have received (Kollar & Fischer, 2010).

The potential of peer-feedback to support learning lies in the fact that students take on an active role in evaluating, assessing, and monitoring their own learning (e.g., Double et al., 2020). In fact, empirical research has shown that peer-feedback may be beneficial for learning, both for the feedback provider and the feedback recipient (Li et al., 2020). Through feedback provision, students assess their own work by comparing it to their peers' and gain insights into their performance. Upon feedback reception, students explore, compare, and weigh alternative task approaches and develop skills in handling various feedback types, enabling them to extract the maximum benefit from it (Nicol et al., 2014). Yet, studies that look at the effects of peer-feedback in the context of pre-service teachers' evidence-informed reasoning are rare.

## Augmenting peer-feedback with a feedback provision scaffold and a feedback integration scaffold

Even though peer-feedback has a strong potential to support student learning, this potential is not always used. Students often have difficulty to (a) provide high-quality feedback and to (b) process the feedback they receive effectively (e.g., Carless & Boud, 2018). With respect to (a), high-quality feedback should include information that prompts the recipient to recall the task (feed up), assess their performance in relation to it (feed back), and provide recommendations for improvement (feed forward, Hattie & Timperley, 2007). However, without guidance, students rarely provide such high-quality feedback (Alemdag & Yildirim, 2022). The same is true for (b) processing received feedback effectively (e.g., Lui & Andrade, 2022), especially when multiple feedback messages need to be integrated with one another. In other words, students often struggle when they receive feedback on their initial task solutions from more than one peer. In this context, integration is defined as the active involvement with multiple texts in which significant content is interconnected to serve as a foundation for revision (e.g., Barzilai et al., 2018). To benefit from such multiple feedback messages, however, mindful processing by the recipient is necessary, but rarely achieved by learners (e.g., Berndt et al., 2018). To promote pre-service teachers' evidence-informed reasoning, both providing and receiving feedback should thus be augmented with appropriate scaffolds, such as rubrics or prompts (e.g., Prins et al., 2005). Although there is a lack of research on the use of peer-feedback to promote evidence-informed reasoning, studies in other contexts, such as self-regulated coherence construction, indicated that a combination of prompts and worked examples holds significant promise for feedback provision (e.g., Graichen et al., 2019). However, to the best of our knowledge, no studies exist on how appropriate scaffolds should be designed for the integration of multiple feedback messages in order to achieve positive effects on performance improvement in the context of analyzing authentic teaching cases and on pre-service teachers' acquisition of evidence-informed reasoning skills.

## Research questions and hypotheses

This study examines the effects of feedback provision and feedback integration scaffolds and their combination on pre-service teachers' (a) improvement of their analyses of authentic teaching problems from initial draft to revision, and (b) their acquisition of evidence-informed reasoning skills. We hypothesized that both kinds of scaffolds would lead to an improvement from draft to revision (H1), and an improved acquisition of evidence-informed reasoning skills (H2).

## Methods

### Sample and design

To test our hypotheses, we ran an experimental study with a 2x2 factorial between-subjects design with the independent variables "feedback provision scaffold" (available vs. not available) and "feedback integration scaffold" (available vs. not available).  $N = 254$  pre-service teachers participated in the context of a regular higher education course, with a mean age of  $M = 22.56$  ( $SD = 4.30$ ) and 77.95% of them being female on average in the middle of their studies ( $M = 4.61$ ,  $SD = 1.20$ ). The study was conducted using an online tool that can be used to structure the peer feedback process.

### Procedure

After a pretest that measured demographic variables, the learning phase corresponded to the four phases of the peer-feedback process described above (task performance phase, feedback provision phase, feedback reception phase, revision phase). The students had one week to complete each phase. Students were asked to individually analyze a written case vignette describing a problematic classroom situation. Each case vignette included six problems, and participants were asked to structure their analysis of each of these problems in five steps: problem identification, problem description, problem explanation, goal setting, and deciding for action (see Greisel et al., 2022). To support the analysis, they were given a summary of two educational theories and related empirical evidence, namely (a) Cognitive Load Theory (Sweller, 1999) and (b) the ICAP model (Chi & Wylie, 2014). Then, the students provided feedback on the problem analyses of two peers. After having received feedback from two peers, the students were supposed to integrate the two feedback messages for the revision of their original analysis. One week after the revision, students completed a posttest to measure their evidence-informed reasoning skills.

### Independent variables

During the *feedback provision* phase, we varied whether or not students received three kinds of prompts and a specific example of how these prompts can be used in practice: to (a) explain the assignment to their classmates once more (feed up, e.g.: "Recall the task and describe it briefly in your own words."), to (b) assess the degree to

which their peers engaged in constructive and critical problem-solving (feedback, e.g.: “Explain how the fellow student completed the task.”), and to (c) offer suggestions for revision (feedback, e.g.: “Finally, make specific suggestions for how your peer can improve his or her problem analysis.”). Regarding *feedback reception*, students received either prompts and an example that explicitly guided them in integrating the feedback they received from the two peers, or nonspecific prompts, depending on the condition. The scaffold acted as a systematic approach to integrating the feedback messages. First, students were directed to read both feedback messages and then compare them. Next, they were asked to highlight points of agreement in green, complementary points in yellow and points of contradiction in red. In the third step, students were tasked with revising their draft using the feedback messages and indicating the changes using the same color-coded system from the previous step.

## Instruments

To evaluate the *quality of the initial draft* and the *revision*, we used a rubric with four levels ranging from 0 (poor quality) to 3 (high quality). Two raters, unaware of the conditions, coded 10 % of the data independently. They achieved excellent interrater agreement (Gwet’s  $AC1 = .99$ ). Subsequently, the remaining 90 % of the data was evenly divided between the two coders. The difference scores between the quality of the draft and the quality of the revision were used as an indicator for improvement.

To measure *skills in evidence-informed reasoning*, we presented students three exemplary analyses of a new case vignette describing problematic classroom situations according to the Cognitive Load Theory (Sweller, 1999) and the ICAP-Model (Chi & Wiley, 2014). These analyses varied concerning the framework used for evidence-informed problem analysis. In the best problem analysis, all five steps were executed, while in the moderate or worst problem analyses, steps were either combined or disregarded entirely. The students had the task to sort the three analyses from best to worst.

## Results

We fitted two linear models to predict the improvement from initial draft to revision and to predict the acquisition of evidence-informed reasoning skills, using R version 4.2.2.

The results revealed no significant effects for *the improvement from initial draft to revision*, neither for the feedback provision scaffold ( $b = .07, p = .58$ ) nor for the feedback integration scaffold ( $b = .14, p = .22$ ). Furthermore, the interaction effect was not significant either ( $b = -.27, p = .13$ ).

For the *acquisition of evidence-informed reasoning skills* (H2), neither main effects of the feedback provision scaffold ( $b = .17, p = .12$ ) nor of the feedback integration scaffold ( $b = .17, p = .06$ ) were found. However, there was a significant interaction effect ( $b = -.38, p = .004$ ): the feedback provision scaffold only had a favorable impact when no additional feedback integration scaffold was presented and vice versa.

## Discussion

Contrary to our expectations, we found no effects for the improvement from initial draft to revision, neither for the feedback provision scaffold nor for the feedback integration scaffold nor for their interaction. This is surprising, since theoretically it could be expected that students would improve their performance through peer-feedback as they should actively engage with the material and read their peers’ analyses, which should in turn provide them with different perspectives that they have to reconcile with their own solution (e.g., Nicol et al., 2014). Thus, it seems that at least when it comes to help pre-service teachers improve upon their evidence-informed analyses of authentic classroom cases, scaffolding approaches that have been shown beneficial in other settings cannot simply be transferred to this context. Future research is necessary to figure out how to best design the peer-feedback process to actually achieve such improvements.

The negative interaction effect of the two scaffolds on students’ acquisition of evidence-informed reasoning skills indicated that the scaffolds work on their own, but not in combination. One explanation might be the different focus of the scaffolds: While the feedback provision scaffold referred to a distinct structure of how to provide feedback, the integration scaffold was concerned with connecting the feedback messages to one another, regardless of the structure of the feedback. As a result, the feedback integration scaffold might have diverted students’ attention away from the required analysis method. It might thus be promising to look into how feedback provision scaffolds on the one hand and feedback integration scaffolds on the other can be better synchronized to avoid mutual interference effects (Prins et al., 2005). This should be investigated in future studies.

## References



- Alemdag, E., & Yildirim, Z. (2022). Effectiveness of online regulation scaffolds on peer-feedback provision and uptake: A mixed methods study. *Computers & Education, 188*, 104574. <https://doi.org/10.1016/j.compedu.2022.104574>
- Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting Integration of Multiple Texts: A Review of Instructional Approaches and Practices. *Educational Psychology Review, 30*(3), 973–999. <https://doi.org/10.1007/s10648-018-9436-8>
- Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education, 33*(1), 31–49. <https://doi.org/10.1007/s10212-017-0343-z>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist, 49*(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Dekker, I., & Meeter, M. (2022). Evidence-based education: Objections and future directions. *Frontiers in Education, 7*, Article 941410. <https://doi.org/10.3389/educ.2022.941410>
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review, 32*(2), 481–509. <https://doi.org/10.1007/s10648-019-09510-3>
- Graichen, M., Wegner, E., & Nückles, M. (2019). Wie können Lehramtsstudierende beim Lernen durch Schreiben von Lernprotokollen unterstützt werden, dass die Kohärenz und Anwendbarkeit des erworbenen Professionswissens verbessert wird? *Unterrichtswissenschaft, 47*(1), 7–28. <https://doi.org/10.1007/s42010-019-00042-x>
- Greisel, M., Wekerle, C., Wilkes, T., Stark, R., & Kollar, I. (2022). Pre-service Teachers' Evidence-Informed Reasoning: Do Attitudes, Subjective Norms, and Self-Efficacy Facilitate the Use of Scientific Theories to Analyze Teaching Problems? *Psychology Learning & Teaching, 22*(1), 1–19. <https://doi.org/10.1177/14757257221113942>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Kiemer, K., & Kollar, I. (2021). Source selection and source use as a basis for evidence-informed teaching: Do pre-service teachers' beliefs regarding the utility of (non-)scientific information sources matter? *Zeitschrift für Pädagogische Psychologie, 35*(2-3), 127–141. <https://doi.org/10.1024/1010-0652/a000302>
- Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344–348. <https://doi.org/10.1016/j.learninstruc.2009.08.005>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193–211. <https://doi.org/10.1080/02602938.2019.1620679>
- Lui, A. M., & Andrade, H. L. (2022). The Next Black Box of Formative Assessment: A Model of the Internal Mechanisms of Feedback Processing. *Frontiers in Education, 7*, Article 751548. <https://doi.org/10.3389/educ.2022.751548>
- Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2005). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education, 11*(3), 289–303. <https://doi.org/10.1007/s10459-005-3250-z>
- Sweller, J. (1999). *Instructional design in technical areas*. ACER Press. <https://doi.org/10.3389/educ.2022.976791>

## Acknowledgments

This work was supported by the Stiftung Innovation in der Hochschullehre (project “Facilitating Competence Development through Authentic, Digital, and Feedback-Based Teaching-Learning Scenarios”, grant number FBM2020).

# Unveiling Collaborative Learning Dynamics: Insights from Remote Jigsaw Instruction on Knowledge Creation Perspectives

Ayano Ohsaki, Shinshu University, aohsaki@ohsaki-lab.net  
Toshinobu Kasai, Okayama University, kasai@okayama-u.ac.jp

**Abstract:** This study examined generative tasks for jigsaw instruction to discuss the design and support of collaborative problem-solving from knowledge creation (KC) perspectives. By analyzing the outcomes and discourse data in relation to KC community, idea improvement, and using knowledge from educational materials, we confirmed the differences between high- and low-outcome groups. Even though the data size is small, the results of this study suggest the potential of predicting outcomes and supporting students based on the process data.

## Background and research question

Facilitating learning from aspects of knowledge creation (KC) has gained increasing attention in the learning sciences field. Jigsaw instruction (Miyake & Kirschner, 2014) has received particular attention for increasing students' engagement in collaborative learning in a group as a community. There are two types of collaborative learning in jigsaw instruction: expert group activities (learners in a group are given the same material to study collaboratively) and jigsaw group activities (each learner shares information and uses it for collaborative problem-solving). Hence, jigsaw instruction has a good structure to support construction of knowledge in a group because learners are assumed to practice KC by collaboratively constructing knowledge objects in learning as KC (Paavola & Hakkarainen, 2005; Scardamalia & Bereiter, 2022).

However, few studies have discussed designing jigsaw instruction elements from the KC perspectives. The latest research about designing jigsaw instruction from the KC perspectives, Ohsaki and Oshima (2021), pointed out that most studies on jigsaw instruction have analyzed collaborative learning as knowledge integration in the belief mode. Moreover, the research discussed designing jigsaw instructional elements from the KC perspectives, using educational materials with both science and engineering components. To illustrate the collective nature of knowledge as clusters of words representing ideas used in discourse and how a word's network structure changes over time by comparing high learning-outcome and low learning-outcome groups, they used temporal socio semantic network analysis (tSSNA) as an advanced analytics algorithm of network analysis for KC (Oshima et al., 2012; Ohsaki & Oshima 2021; Scardamalia & Bereiter, 2022). Ohsaki and Oshima (2021) concluded that the main challenge in jigsaw instruction for smooth transitions to the generative task might prompt students to create a shared understanding and transition to generative collaborative actions.

Besides, when designing the jigsaw method, teacher intervention is an issue that should be considered. Despite not being discussed previously, this is a crucial aspect not only in the design of jigsaw instruction but also in the design of group activity in remote classrooms, which has recently revitalized discussions regarding infection control and learner convenience. For example, prior research on remote classrooms discussed students being required to self-management studying because the teacher is not physically present (Raes et al., 2020). Furthermore, Bower et al. (2015) reported that while participating in class remotely could be helpful in managing the demands of time and place, some students felt that the learning experience was compromised in a remote classroom. Hence, examining how learners engage in a main challenge without teachers' intervention in a remote classroom can provide critical insights for redesigning main challenges and educational material and designing support by teachers or teaching assistants.

Given the above, this study examines the design of jigsaw instruction with generative tasks in remote classrooms from KC perspectives. The research question is "How do students engage in a jigsaw instruction environment to solve generative challenges in a remote classroom?"

## Method

### Data collection

This study used two types of datasets: individual outcome and discourse data during the jigsaw group activity. A total of 34 students from several department types, such as Liberal Arts, Data Science, and Science, were divided into 11 groups of three or four. All students were studying for the teachers' certificate, but the subjects differed. Furthermore, six groups attended a remote classroom, and five groups attended an in-person classroom. This study focused on two groups of students pursuing science teacher licensure who participated in the remote classroom,

allowing for a comparative analysis under equivalent conditions. Their activities were recorded and transcribed. The average number of utterance lines was 471.5 (SD = 136.47).

## Class design

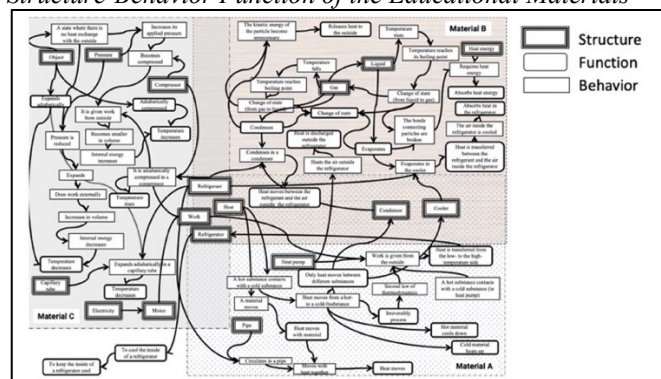
The teacher designed a 90-minute jigsaw instruction class as part of the specialized subject for teaching practice. The main challenge of the jigsaw activity was, “In a vending machine that offers cold soda and hot coffee, how should the refrigerator be used to control the temperature of the drinks efficiently?” This challenge was designed as a task to facilitate students to not only integrate science and engineering knowledge but also generate their ideas. Before and after the group work, students were required to describe their ideas about the main challenge on a worksheet individually. In the expert group activity, students studied the three different documents, on (A) theory of heat, (B) phase change and thermal energy, and (C) pressure and temperature, in each group. Next, members from different groups created a new group for the jigsaw group activity for 30 min. For this group work, the teacher set up three phases to support students: Phase 1 for sharing information from educational materials; Phase 2 for creating shared answers to the main challenge; and Phase 3 for improving their answers.

The class was held in a remote classroom because of the classroom capacity. The teacher instructed from the next classroom via an online video meeting tool Zoom (n.d.), and students discussed in the group in-person. Two assistants in the classroom helped with the distribution of educational materials and worksheets. However, they did not intervene in solving the problem.

## Analysis procedure

This study analyzed the two datasets from three perspectives: KC community, idea improvement, and collaborative problem-solving. First, the outcome data was used for evaluating group activities as a KC community, referring to prior study (Ohsaki & Ohima, 2021). We analyzed this data using the structure-behavior function (SBF) framework (Hmelo-Silver & Pfeffer, 2004) to assess individual knowledge integration. We confirmed the results at a group level to determine whether KC was taking place as a community, defining two groups: high-SBF (all members integrated knowledge) and low-SBF (some members integrated knowledge). Figure 1 shows the SBF we created to evaluate student outcomes. This figure illustrates 88 SBF components and which educational materials provide information.

**Figure 1**  
*Structure-Behavior Function of the Educational Materials*



Second, we analyzed discourse data from the aspect of idea improvement. Following previous research (Ohsaki & Oshima 2021), this study applied the tSSNA method, which is advanced analytical method that adds a temporality perspective to socio-semantic network analysis (SSNA), a well-known analytical method for improving ideas in KC. SSNA and tSSNA calculate the density of key phrases (the sum of the degree centrality of key phrases) as a representation of participants' ideas (Oshima et al., 2012; Ohsaki & Oshima, 2021). Regarding the setting of tSSNA, the window size was four and the network lifetime was two, because the minimum number of members was three and the activity duration was short. Although the previous study visualized data using timestamp information, our dataset did not include timestamps, so we present the data as a simplified tSSNA, based on the temporal Phases 1, 2, and 3 of the classes. The number of key phrases was 33, which came from educational materials and class design by the teacher.

Third, discourse data was also analyzed to capture how students used information in educational materials for collaborative problem-solving. For this analysis, we used epistemic network analysis (ENA), which can visualize the epistemic framework (Shaffer, 2017). The coding category was created deductively to show how

students use the information from the educational materials. These 13 codes are shown in Table 1. The settings for the ENA were as follows: window size = 4; unit = Level, Phase, and Speaker; conversation = Level and Phase.

**Table 1**

*Coding Category*

Code	Definition
Question	Talking about the main challenge
A, B, C	Talking about each educational material (e.g., “A” means the information from the educational material A)
A.B, B.C, C.A	Talking about the relationship between two educational materials (e.g., “A.B” means that students compared/related to educational materials A and B)
AppA, AppB, AppC	Solving the main challenge using each educational material (e.g., “App.B” means information from the educational material B was used to solve the main challenge.)
AppA.AppB, AppB.AppC, AppC.AppA	Talking about the relationship between two educational materials as applied to the main challenge (e.g., “AppC.AppA” means both using information from educational materials C and A to solve the main challenge and comparing/relating to educational materials C and A.)

## Result and discussion

This study conducted three data analyses to answer the research question, “How do students engage in a jigsaw instruction environment to solve generative challenges in a remote classroom?” The first analysis assessed success of KC community based on students’ final challenge description. The first and second authors independently assessed SBF inclusion in the descriptions of all six students in both groups. The inter-rater agreement was moderate ( $\kappa = 0.77$ , Cohen’s kappa). While all participants in one group incorporated all materials, only one participant in the other group did. Therefore, we defined the former as high-SBF and the latter as low-SBF.

The second analysis aspect was idea improvement. The first row in Figure 2 illustrates the transitions of total degree centrality during the jigsaw group activities. The X-axes show line order, and the Y-axes show the total degree centrality score. Moreover, to enhance the changes in both groups, the Y-axis range of the three graphs in this row differs in each graph. The analysis results from idea improvement describe that both groups’ activity was similar to typical Jigsaw instructions in in-person classrooms, such as they marked higher scores in the early phase than in the late phase. Moreover, the graphs show that the high-SBF group engaged in tasks for the whole activity duration, but the low-SBF group talked regardless of the task. Specifically, the scores of the high-SBF group were changed through all phases. By contrast, the low-SBF group has low vibrations of the scores, and there were many marks of zero despite having more lines than the high-SBF group.

The third analysis compared the high- and low-SBF groups using ENA to examine how students used information to address the main challenge. The first and second authors coded the discourse data using social moderation (Herrenkohl & Cornelius, 2013) for the high-SBF group, with subsequent independent analysis for the low-SBF group showing a code agreement rate of 0.89 (Cohen’s kappa). The subtracted ENA graphs in Figure 2(2) showed that the high-SBF group talked more about the relationship among all educational materials in Phase 1 and then used the information from these materials to solve the main challenge in Phase 2 compared to the low-SBF group. Interestingly, in Phase 3, the low-SBF group talked more about the relationship or comparison of applied knowledge from educational materials than the high-SBF group. The students in the low-SBF group confirmed that they comprehensively utilization of all the provided educational materials to complete the main challenge. Hence, Figure 2(2c) suggests that the low-SBF group set the goal of using all educational materials instead of improving their ideas. Mann–Whitney tests on WebENA indicate the high-SBF group (Mdn = -0.36, N = 9) significantly differs ( $U = 6.00$ ,  $p = 0.00$ ,  $r = 0.85$ ) from the low-SBF group (Mdn = 0.28, N = 9) along the X-axis. Conversely, along the Y-axis, no significant difference was found between the high- (Mdn = 0.33, N = 9) and low-SBF groups (Mdn = 0.27, N = 9,  $U = 35.00$ ,  $p = 0.67$ ,  $r = 0.14$ ).

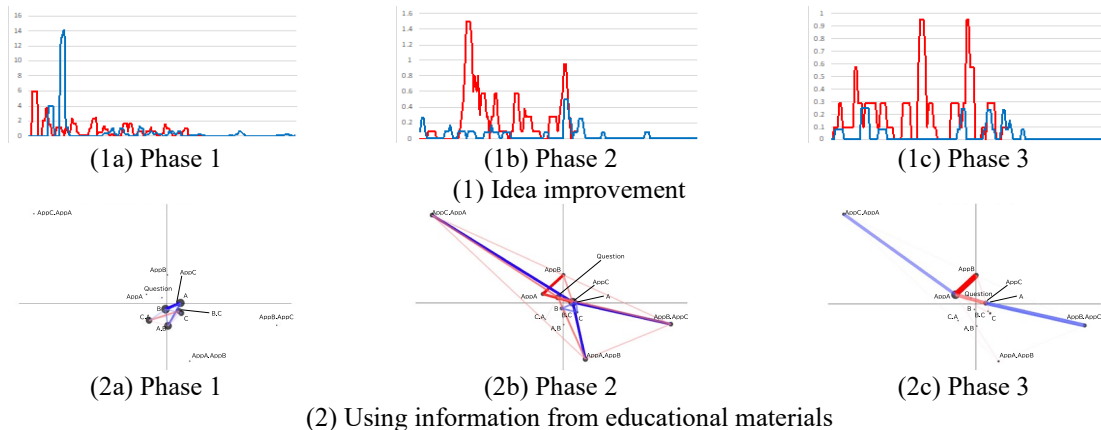
Our discussion is limited because this study used only two groups’ data. To focus on the design of the jigsaw instruction, we gathered data from students studying for teaching certification in science, who did group activities in a remote classroom without teacher intervention. Analysis of data from students studying for teaching certification in other subjects and students in an in-person classroom and comparison with the results of this study will provide additional valuable insights into teacher support and learning environment design for collaborative learning.

Despite its limitations, this study contributes to the further design and support of collaborative learning. We confirmed that students in a high-outcome community engaged in continuous idea improvement and

generative tasks without teacher intervention. These findings align with prior research results and theories (e.g., Ohsaki & Oshima, 2021). Furthermore, our analysis approach has distinguished the high- and low-KC communities by the combination of SBF, simple tSSNA, and ENA. In other words, this study suggests that this research could be extended to predict community outcomes from the learning process and support students during their studies to achieve better outcomes.

**Figure 2**

*Results of Analysis for Idea Improvement and Using Information from Educational Materials*



Note: high-SBF group is shown in red; low-SBF group is shown in Blue

## References

- Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In de Corte, E. E., Verschaffel, L. E., Entwistle, N. E., & Van Merriënboer, J. E. (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 55-68). Pergamon/Elsevier Science Ltd.
- Bower, M., Dalgarno, B., Kennedy, G. E., Lee, M. J., & Kenney, J. (2015). Design and implementation factors in blended synchronous learning environments: Outcomes from a cross-case analysis. *Computers & Education*, 86:1-17.
- Herrenkohl, L., R. & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *Journal of the Learning Sciences*, 22(3),413–461
- Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions, *Cognitive Science*, 28(1), 127-138.
- Miyake, N., & Kirschner, P. A. (2014). The social and interactive dimensions of collaborative learning. In Sawyer, K. (Ed.), *The Cambridge handbook of the learning sciences*, 2nd ed. (pp. 418–438). NY: Cambridge University Press.
- Ohsaki, A., & Oshima, J. (2021). Knowledge Creation Analytics for Jigsaw Instruction: Temporal Socio-Semantic Network Analysis. In *Proceedings of the 15<sup>th</sup> ICLS* (pp. 537-540), Online: ISLS
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60(5), 903-921.
- Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor—An emergent epistemological entapproach to learning. *Science Education*, 14(6), 535-557.
- Raes, A., Detienne, L., Windey, I., & Depaepe, F. (2020). A systematic literature review on synchronous hybrid learning: gaps identified, *Learning Environments Research*, 23:269–290
- Scardamalia, M., & Bereiter, C. (2022). Knowledge Building: Advancing the State of Community Knowledge. In K. Sawyer (Ed.) *The Cambridge handbook of the learning sciences: Third edition* (pp.261-279). Cambridge University Press.
- Shaffer, D.W. (2017), *Quantitative Ethnography*, Cathcart, Madison
- Zoom (n.d.), <https://zoom.us/> (last accessed 2023-11-16)

## Acknowledgments

We extend our gratitude to all the students who participated in the research endeavor and to everyone who contributed to the class's execution. This research was supported by JSPS KAKENHI (JP22H01043, and JP23K11357).

## Who Wants to Hear Bad News? How the Epistemic Perspective Determines the Perception of Peer Feedback

Martin Greisel, Julia Hornstein, Anna Weidenbacher, Johanna Ott, Ingo Kollar  
martin.greisel@uni-a.de, julia.hornstein@uni-a.de, anna.weidenbacher@gmx.de, johanna.l.ott@uni-a.de,  
ingo.kollar@uni-a.de  
University of Augsburg, Germany

**Abstract:** Processing feedback from peers is an essential part of learning through peer feedback. However, if a feedback message is critical about the students' initial task solution, students might perceive it as inadequate and not process it further. Based on multiple document research, we assume that epistemic perspectives (i.e., absolutism, multiplism, and evaluativism) determine as how adequate students perceive feedback in case it conflicts with their initial solution. We asked 254 pre-service teachers to analyze a classroom case vignette, provide feedback to each other, and revise their case analysis. Linear mixed models indicated that the lower students' absolutism or evaluativism was, and the more their feedback contained criticism, the less adequate they perceive it. Multiplism did not interact with criticism. We conclude that the effect of absolutism might depend on the identification with one's initial solution, and that evaluativism helps to value criticism for it containing new information.

### Rationale of this paper

Peer feedback is an effective method to foster learning, on average even more powerful than feedback from teachers, though its effects vary in recent meta-analyses, ranging from large negative to large positive effects (e.g., Double et al., 2020). In this paper, we address one potential source of this variance: the students' epistemic perspective, that is, the way how students think about the nature of knowledge and its justification (Barzilai & Weinstock, 2015). This is the rationale: Feedback comprises potentially valuable new information, especially if it is critical. However, criticism is most likely to conflict with students' original viewpoints. Thus, students might tend to reject the feedback as inadequate, which should depend on their epistemic perspective as it determines how learners handle conflicting information (Bråten et al., 2013). To sum up, we argue that peer feedback might not work for everyone to the same extent because individuals differ in how they process the feedback they receive (Lui & Andrade, 2022). The way they process that feedback might be influenced by their epistemic perspectives.

### The individual in the peer feedback process

A learning scenario which contains peer feedback typically includes the following steps (Bauer et al., 2023): First, students create an initial solution to a task. Second, the initial solutions are distributed among peers and peers provide each other with feedback. Third, students receive feedback from their peers and process it. Fourth, they revise their initial solution based on the feedback they received.

Thus, to benefit from it, processing the feedback from peers thoroughly is crucial. Yet, if students deem the feedback inadequate (Strijbos et al., 2021), it is unlikely that they are motivated to revise their initial solution and consider the feedback content thereby. For this reason, it is important that students appraise the feedback they receive as objectively correct and only dismiss it if it does not contain any helpful comment at all. However, students' perception of the feedback they receive is likely filtered through their individual epistemic perspectives (Barzilai & Weinstock, 2015). As feedback comprises information intended to be integrated into one's knowledge, epistemic perspectives are influential because they determine how one thinks about knowledge and its justification (Barzilai & Weinstock, 2015). Following Barzilai and Weinstock (2015), there are absolutist, multiplist, and evaluativist perspectives. An absolutist perspective sees the truth as certain and objectively available in the external world, whereas a multiplist perspective views the truth as uncertain and subjective. Holding an evaluativist perspective, finally, should lead the learner to acknowledge that truth is not easily objectively available but has to be concluded by interpretation of data based on criteria (Barzilai & Weinstock, 2015).

### Epistemic perspectives and feedback

In previous studies, epistemic beliefs' relation with peer feedback behavior varied (e.g., Noroozi, 2023). However, these studies did not consider how critical peers' feedback was. To make assumptions on how students handle critical feedback they receive, we can borrow from research on multiple documents. There, evaluativism yields the best integration and comprehension of conflicting texts, whereas multiplism performs worst (e.g., Bråten et

al., 2013). In peer feedback, students have to deal with multiple documents as well: their own initial task solution and the corresponding feedback message. These documents might agree or take conflicting viewpoints.

Consequently, the epistemic perspective should matter especially when the feedback does not agree with one's initial solution (Barzilai & Eshet-Alkalai, 2015): As an absolutist stance assumes that only one of two conflicting viewpoints can be true, and as it is likely that the own perspective is preferred, critical feedback should be less accepted. In contrast, as a multiplist perspective entails that it is impossible to find the objective truth, every viewpoint is a viable opinion. Then, it should not matter whether feedback favors one's solution or not. Third, as an evaluativist perspective means that data should be interpreted based on criteria, students should appraise feedback based on the quality of its reasoning independent from whether it is critical or not. In the following, we test whether these interactions between epistemic perspectives and the extent of criticism in feedback messages predict how adequate students perceive the feedback in a peer feedback environment.

## Method

### Sample

A sample of 254 pre-service teachers (mean age = 22.56,  $SD = 4.30$ ; 77.95% female) in their  $M = 4.61$  semester ( $SD = 1.20$ , about the middle of studies) from majors in elementary school education (55.12%) and different variants of secondary school education participated in the study. The peer feedback scenario was a mandatory part of a course on educational psychology. However, participation in the scientific data collection was voluntary.

### Procedure

Participants studied over three weeks within a digital learning environment that was designed to help them acquire the skill to reason about teaching problems in an evidence-informed manner (Greisel et al., 2022). In Week 1, the students answered a questionnaire regarding their epistemic perspectives and analyzed a case vignette of a lesson in which a teacher has problems such as students not talking to each other during collaborative learning. Their task was to identify and describe each problem, explain the problem with matching theoretical concepts, derive a goal, and develop a plan for teacher action. The students were supported with summaries of educational theories, which matched the problems in the case, and a description and a worked example explaining the steps they were asked to perform in their analysis. In Week 2, students rated the quality of the case analyses from two randomly assigned peers and produced written feedback messages. In Week 3, the feedback messages were anonymously delivered, students rated and processed their feedback, and then revised their case analysis.

### Instruments

To assess epistemic perspectives, we used items from the Epistemic Thinking Assessment (Barzilai & Weinstock, 2015), which comprises 7 questions with 3 items each which represent *absolutism*,  $\alpha = .69$ , *multiplism*,  $\alpha = .60$ , and *evaluativism*,  $\alpha = .69$ , answered on a Likert-scale from 1 = *do not agree at all* to 6 = *completely agree*. A sample question was “Is there an answer to how problematic teaching situations can be solved?” accompanied by these items “Eventually there will be one right answer” (absolutism), “In principle, it is impossible to know the right answer” (multiplism), “There may be multiple right answers but they are not equally right” (evaluativism).

The extent of criticism in a feedback message was measured indirectly. Feedback providers assessed the quality of the initial solution by indicating on five items if the peer was able to analyze the case in an evidence-informed manner,  $\alpha = .89$ , using the item stem, “Overall, my fellow student succeeded, using the ICAP model and cognitive load theory, in...” (sample item: “explaining instructional problems correctly”). The less participants agreed on a Likert-scale from 1 = *not at all true* to 5 = *completely true*, the more likely it is that their feedback would contain criticism to a larger extent. Raw data were reversed to ease interpretation (higher values equal higher criticism). These values were used twice to measure (a) the criticism participants received and (b) the criticism participants provided to their peers. The latter was only used as covariate in all subsequent regressions to control for the quality which the other case analyses had which a student provided feedback for.

Perceived adequacy of feedback was measured with the Feedback Perception Questionnaire (Strijbos et al., 2021). We computed a total scale value based on the dimensions *fairness* (“I would consider this feedback fair”), *usefulness* (“I would consider this feedback useful”), and *acceptance* (“I would accept this feedback”) with three Likert-scaled items with  $\alpha = .93$ .

## Results

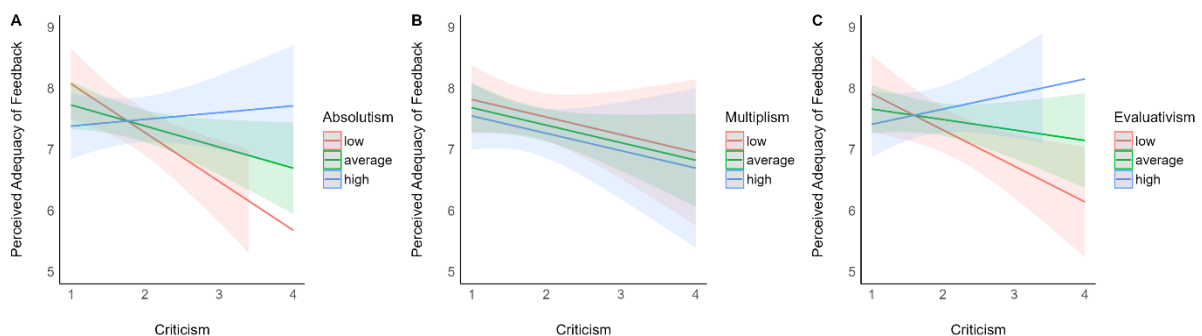
Descriptives showed that students held mostly evaluativist perspectives,  $M = 4.31$  ( $SD = 0.64$ ), compared to absolutism,  $M = 2.74$  ( $SD = 0.72$ ) and multiplism,  $M = 2.93$  ( $SD = 0.63$ ), and that they perceived feedback they

received as rather adequate,  $M = 7.29$  ( $SD = 1.25$ ). Perceived adequacy of feedback was not associated with criticism,  $M = 2.03$  ( $SD = 0.75$ ), independent from whether calculated as bivariate Pearson correlation with the mean of both feedback perceptions,  $r = .08$ ,  $p = .327$ , or calculated as multilevel regression,  $\beta = -0.14$ ,  $p = .104$ . Regarding the main hypotheses, we calculated multilevel regressions as each participant received up to two feedback messages. The interaction effects of feedback criticism with absolutism,  $\beta = -0.23$ ,  $p = 0.008$ , and evaluativism,  $\beta = -0.21$ ,  $p = 0.015$ , were statistically significant and negative, whereas criticism and multiplism did not interact,  $\beta = -0.001$ ,  $p = 0.993$  (see Table 1 and Figure 1). That is, the less absolutistic students' perspectives were, the less adequate they perceived critical feedback. The same is true for evaluativism. In contrast, students perceived critical feedback independently from their multiplistic perspective.

**Table 1**  
*Multilevel Regression of Perceived Adequacy of Feedback on Epistemic Perspectives and Feedback Criticism*

Predictors	Model 1		Model 2		Model 3	
	<i>b</i> ( <i>SE</i> )	<i>p</i>	<i>b</i> ( <i>SE</i> )	<i>p</i>	<i>b</i> ( <i>SE</i> )	<i>p</i>
(intercept)	11.69 (1.47)	<0.001	9.14 (1.94)	<0.001	13.42 (2.91)	<0.001
criticism of peers' solutions	-0.33 (0.20)	0.098	-0.26 (0.20)	0.203	-0.30 (0.20)	0.127
criticism	-2.03 (0.67)	<b>0.003</b>	-0.30 (1.03)	0.774	-3.30 (1.26)	<b>0.010</b>
absolutism	-1.08 (0.47)	<b>0.021</b>				
absolutism * criticism	0.62 (0.23)	<b>0.008</b>				
multiplism			-0.23 (0.64)	0.722		
multiplism * criticism			0.00 (0.35)	0.993		
evaluativism					-1.13 (0.63)	0.074
evaluativism * criticism					0.71 (0.29)	<b>0.015</b>
<i>Random Effects</i>						
$\sigma^2$	1.38		1.48		1.49	
$\tau_{00}$	0.74 <small>t11fdn</small>		0.73 <small>t11fdn</small>		0.60 <small>t11fdn</small>	
ICC	0.35		0.33		0.29	
N	104 <small>t11fdn</small>		104 <small>t11fdn</small>		104 <small>t11fdn</small>	
Observations	146		146		146	
Marginal $R^2$ / Conditional $R^2$	0.076 / 0.400		0.034 / 0.351		0.080 / 0.345	

**Figure 1**  
*Perceived Adequacy of Feedback Predicted by an Interaction of Epistemic Perspectives and Criticism*



## Discussion

In order to foster learning, feedback from peers needs to be processed thoroughly (Lui & Andrade, 2022). This is only likely if students consider the feedback they receive as adequate (Strijbos et al., 2021). However, we assumed that this depends on the extent of criticism the feedback contains, moderated by learners' epistemic perspectives.

First, we hypothesized that an absolutist perspective should make learners prone to favor their own perspective when viewpoints are conflicting (Barzilai & Eshet-Alkalai, 2015) because if only one viewpoint can be true, then favoring the own viewpoint is more consistent and takes less effort. However, we found the opposite, that is, only students low on absolutism seem to reject critical feedback. This indicates that favoring the own viewpoint in case of conflicting viewpoints is not generally true. More specifically, more able students provide



more critical feedback regarding substantive issues (Patchan et al., 2013). Therefore, the more critical feedback might also have been more elaborated and better justified, thus might have seemed more trustworthy. For this reason, absolutistic students might have picked the external viewpoint instead of their own when pressured to decide because of conflicting positions. Second, we hypothesized that a multiplist perspective would help students to ignore whether the feedback they receive is critical of their work (Barzilai & Eshet-Alkalai, 2015). However, as criticism was not associated with perceived adequacy even when multiplism was not considered, the potential of a multiplist perspective to qualify other opinions was obsolete. Third, we hypothesized that an evaluativist perspective would render students' feedback appraisal independent of the amount of criticism contained in the feedback because such a perspective should emphasize evaluation based on criteria (Barzilai & Eshet-Alkalai, 2015), that is, the quality of argumentation should matter more than the valuation itself. Yet, we found that the more evaluativistic students' perspective was, the more they favored critical feedback. This also makes sense from an evaluativistic standpoint, as agreement with one's own position might not provide new insights, whereas critical and conflicting positions might provide the feedback recipient with new information. As evaluativists base their perceptions of true and false on data, new information might be considered beneficial in its own right.

## Limitations and conclusions

Of course, this study is not without limitations. First, the factorial validity of the measurement of epistemic perspectives was not satisfying. Second, we did not directly measure the extent of criticism in the feedback messages. Thus, some meaningful covariation of true feedback content and epistemic perspectives might be lost. However, both limitations would typically yield lower correlations and obscure effects. Therefore, the reported effect sizes might constitute lower boundaries of the real effects rather than exaggerations. Third, self-reports might be biased towards social desirability masking negative reactions to criticism.

Theoretically, our results imply that theory regarding the effects of an absolutist perspective should integrate how strongly one identifies with one's own viewpoint or prior solution as a moderator (Barzilai & Eshet-Alkalai, 2015), which constitutes a mechanism that has not been considered before in this field (e.g., Noroozi, 2023). Practically, students often hold back criticism because they worry that critical arguments would not be valued by the feedback recipients (Vanderhoven et al., 2015). However, our results could be used to encourage students to be thorough and critical. Teachers could scaffold high-quality criticism, for example, by providing evaluation criteria or sample solutions which students can use to compare their peers' work with.

## References

- Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learning and Instruction*, 36, 86–103. <https://doi.org/10.1016/j.learninstruc.2014.12.003>
- Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology*, 42, 141–158.
- Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2013). Justification beliefs and multiple-documents comprehension. *European Journal of Psychology of Education*, 28(3), 879–902.
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review*, 32(2), 481–509.
- Greisel, M., Wekerle, C., Wilkes, T., Stark, R., & Kollar, I. (2022). Pre-service teachers' evidence-informed reasoning: Do attitudes, subjective norms, and self-efficacy facilitate the use of scientific theories to analyze teaching problems? *Psychology Learning & Teaching*, 22(1), 20–38.
- Lui, A. M., & Andrade, H. L. (2022). The Next Black Box of Formative Assessment: A Model of the Internal Mechanisms of Feedback Processing. *Frontiers in Education*, 7.
- Noroozi, O. (2023). The role of students' epistemic beliefs for their argumentation performance in higher education. *Innovations in Education and Teaching International*, 60(4), 501–512.
- Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, 41(2), 381–405. <https://doi.org/10.1007/s11251-012-9236-3>
- Strijbos, J.-W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, 68, 100980.
- Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, 81, 123–132.

## Acknowledgments

This work was supported by the Stiftung Innovation in der Hochschullehre; FBM-2020.

# Learning Under Stress: Enhancing Team-Based Simulation Training with Multimodal Data

Daniel Sánchez, University of Oslo, [daniel.sanchez@iped.uio.no](mailto:daniel.sanchez@iped.uio.no)

**Abstract:** High-stakes healthcare environments present unique stress challenges impacting individual performance and collaborative dynamics. This study harnesses multimodal learning analytics to investigate stress patterns in team-based nursing simulations, using medical-grade sensors to map physiological stress responses to key collaborative events, like decision-making during patient health crises. Analysis revealed that participants' stress levels not only peaked during high-responsibility tasks but also showed elevated baselines, suggesting pre-simulation stress. Integrating these findings into reflective debriefing sessions can enhance stress management and support deeper learning. This research highlights the value of multimodal data in CSCL environments to support both individual and collective learning processes, with broader implications for designing educational technologies that prepare students for the realities of clinical practice.

## Introduction

Computer-Supported Collaborative Learning (CSCL) is a domain that actively investigates the nexus between technology, collaboration, and learning outcomes in diverse educational settings. The advent of sophisticated simulation technologies, especially in high-stakes fields like healthcare education, presents a unique opportunity to examine the influence of stress on collaborative learning environments (Issenberg et al., 2005). High-fidelity simulations in nursing education are not only instrumental in imparting technical skills but also in refining how learners cope with stress within team-based scenarios, a critical component of effective professional practice (Fraser et al., 2012).

Building on this foundation, the current study expands the scope of CSCL by incorporating multimodal learning analytics (MMLA) to monitor physiological stress responses during simulated clinical exercises. The relationship between physiological markers and the cognitive, affective, and social dimensions of learning underscores the potential of MMLA in enhancing the fidelity of simulations and, consequently, the attainment of learning outcomes (Blikstein & Worsley, 2016; Yan et al., 2023).

The body of research on this specific topic is relatively small but growing. For instance, (Zhao et al., 2022) integrated positional and audio data in nursing simulations to differentiate simultaneous team communications, contributing to understanding the complexity of team interaction in stressful conditions. Similarly, Yan et al. (2023), using physiological data and surveys, found that physiological synchrony and arousal provide insights into how learners' stress levels can directly influence team dynamics and individual performance. Their study underscored the role these factors play in high-stress healthcare simulations, calling for a deeper exploration of such physiological measures in educational practices.

Despite these advancements, defining clear criteria for assessing student performance with MMLA remains challenging. This study adopts the approach proposed by (Sánchez et al., 2024) that merges multimodal data with nursing expertise. This approach requires the identification of specific student actions within the learning scenario, deemed relevant by the experts for their practical and educational significance. These actions, once identified, are then assessed from the perspective of the experts to ensure their alignment with both the practice and the intended learning outcomes of the activity. In doing so, the data and sensor selection are directly informed by and relevant to the nuances of professional nursing practice. This integration of technological capabilities with expert-driven perspective is a key feature of the present study.

Drawing upon the theoretical framework of collaborative learning (Dillenbourg, 1999), this research investigates how physiological stress indicators, such as heart rate, correlate with team dynamics and individual performance in high-pressure simulated healthcare settings. These physiological measures are crucial, as they offer objective insights into learners' emotional states, often obscured in traditional assessment methods (D'Mello & Graesser, 2012).

This work contributes to the existing body of CSCL literature by providing empirical evidence on the efficacy of integrating MMLA into collaborative learning environments. By doing so, it seeks to bridge the gap between the subjective experience of stress and its observable impact on collaborative learning processes. This research aligns with the CSCL community's increasing focus on leveraging advanced technologies to support and enhance learning in complex, real-world tasks (Wise & Schwarz, 2017).

## Methods

### Participant and setting

This study was conducted with students enrolled in a master-level nursing course at a Norwegian university. Participation was voluntary, with an additional simulation session offered as an incentive. Thirteen students enrolled to participate in the study. They were divided into three groups based on their availability to ensure optimal participation and engagement. The present study focuses on one of these groups, comprising five students. Each participant was assigned a specific role such as Leading Nurse, responsible for overall patient care coordination; Supporting Nurses, assisting in procedures; Handover Responsible, in charge of information transfer; and Observer, tasked with monitoring and feedback. Ethical approval and informed consent were obtained in accordance with the university's research guidelines.

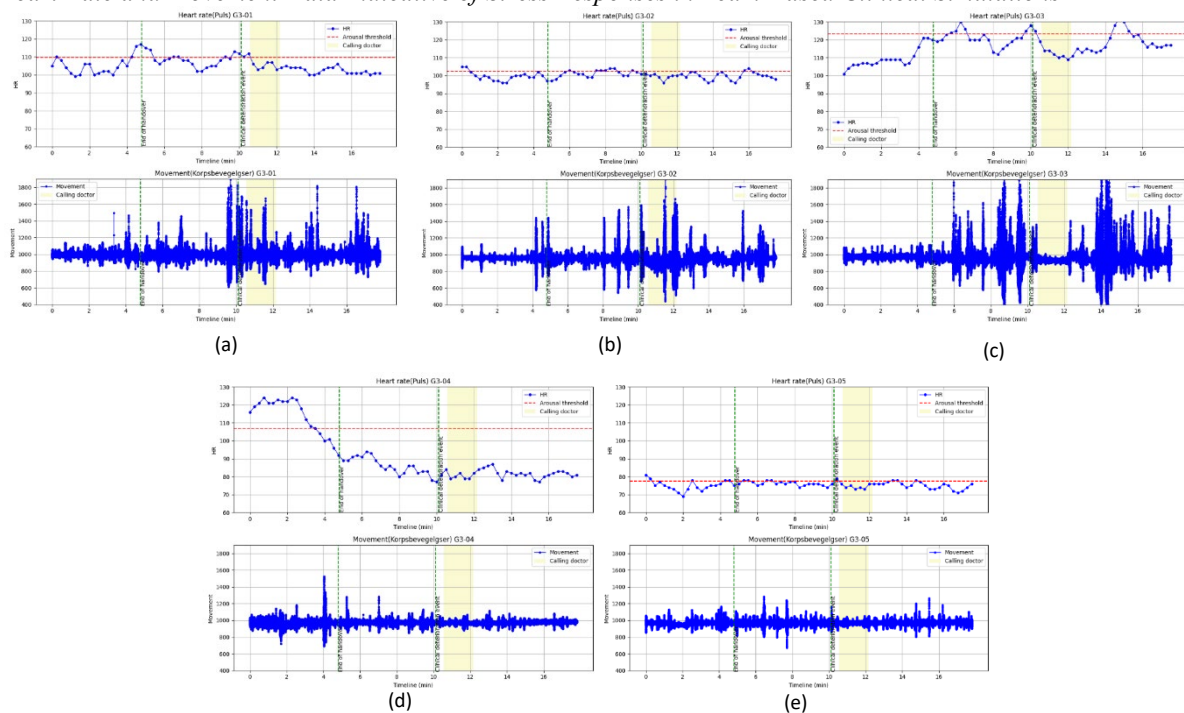
### Simulation design and phases

The simulation was designed to mimic the high-stakes environment of an ICU, with an emphasis on teamwork, communication, and stress management skills. In this simulation, nurse teams were tasked with receiving and attending to a patient, represented by a high-fidelity manikin, who had been involved in a motorbike accident. This included managing critical health deteriorations such as addressing a sudden drop in blood pressure while contending with bleeding. Simulation was structured into three distinct phases to emulate the varying levels of stress and complexity: (1) 'Handover Phase,' where nurses were briefed with patient information, (2) 'Stabilization Phase,' involving the application of care procedures, and (3) 'Post-Deterioration Phase,' where the teams navigated a scenario involving the decline of the patient's condition.

### Data collection and processing

Multimodal data were collected using eq02+ LifeMonitor sensors, worn by participants throughout the simulation. These sensors provided continuous heart rate (HR) monitoring for arousal level assessment and included triaxial accelerometers to track body movement. Net acceleration was calculated based on measurements across the three axes, offering a comprehensive view of physical activity. Being medical-grade devices, these sensors ensured the accuracy of the measurements. Baseline HR levels were recorded 15 minutes prior to the simulation, establishing the baseline as the mean of these pre simulation readings plus one standard deviation. This approach provided a reference for evaluating stress responses during the simulation. Significant collaborative events, such as calling a physician, were identified and recorded to align physiological responses with specific team dynamics.

**Figure 1**  
*Heart Rate and Movement Data Indicative of Stress Responses in Team-Based Clinical Simulations*



## Results

The multimodal data analysis captured during the ICU simulation provided compelling insights into the participants' stress responses, as depicted in Figure 1. Each panel corresponds to a different participant, indicated by their roles: G3-01 (Leading Nurse), G3-02 and G3-03 (Supporting Nurses), G3-04 (Handover Responsible), and G3-05 (Observer). Each panel in Figure 1 is divided into two graphs: The upper graph displays heart rate (HR) data, with the baseline represented by a horizontal threshold (red-dotted) line. When HR values rise above this line, it indicates an arousal peak suggestive of increased stress. Vertical (green-dotted) lines delineate the simulation phases, while the yellow area highlights the critical event of a student calling the doctor for patient updates and guidance.

The lower graph shows movement data, expressed as net acceleration, indirectly measuring physical activity. The combined examination of HR and movement data allows for a detailed examination of stress responses, distinguishing if the stress peaks are due related to physical exertion or mainly to psychological stress factors.

### Individual and team Stress Responses

- *Lead Nurse (G3-01)*: Exhibiting a sustained HR above the baseline threshold with episodic moments of body movement, Figure 1 (a) suggests continuous stress, with pronounced peaks during decision-making events, such as calling the doctor.
- *Supporting Nurses (G3-02 and G3-03)*: Displayed contrasting HR patterns; see Figure 1 (b) and (c). G3-02 maintained stable HR levels despite movement peaks. At the same time, G3-03 showed a significant HR rise in conjunction with physical activity and afterward, implying an interplay of psychological and physical stress.
- *Handover Responsible (G3-04)*: Initially, this participant's HR data, as seen in Figure 1(d), showed spikes that correlate with the period of transferring critical patient information. Notably, once the Handover Responsible's duties concluded, their HR returned to normal levels (60-90 beats per minute), signifying a release from the high-stress role.
- *Observer (G3-05)*: The Observer, depicted in Figure 1(e), demonstrated a relatively stable HR profile with occasional minor fluctuations, potentially indicative of empathetic stress or cognitive involvement.

The collective narrative unfolds during the high-stress event of patient deterioration, where an evident synchronized HR increase across all roles showcases a unified team response to the critical situation. The lack of significant movement data spikes during this phase accentuates the psychological nature of the stress response.

The visualized HR and movement data in Figure 1 encapsulate the study's core findings, illustrating the individual stress profiles within the team and their collective response to a high-stakes event. This figure serves not just as an analytical tool but as a narrative device that conveys the complex interplay between physiological stress, collaborative roles, and high-pressure healthcare environments. It underscores the significance of understanding stress management in CSCL settings, particularly for teams in demanding contexts.

## Discussion

The findings from this study offer a nuanced understanding of how stress is manifested and managed within a collaborative learning environment, particularly in high-stress healthcare simulations. The physiological metrics—heart rate and body motion—shows how physiological responses to stress are linked to specific simulation phases and team roles, echoing the multifaceted nature of stress reported by Fraser et al. (2012).

Notably, the role-specific stress responses, as highlighted by the Handover Responsible (G3-04), underscore the acute impact of high-responsibility tasks on physiological arousal. This finding aligns with Issenberg et al.'s (2005) emphasis on the need for high-fidelity simulations to closely mimic the stressors of real clinical environments to prepare students for actual healthcare settings. The fluctuating stress levels, modulated by the phase of the simulation and the immediacy of the tasks, further corroborate the complex relationship between task demands and stress responses.

Additionally, the collective increase in heart rate among participants during the patient deterioration scenario mirrors the collaborative dynamics emphasized in the CSCL framework by Dillenbourg (1999). This collective physiological response to a high-stress event within the simulation demonstrates the interdependence of team members' emotional states and the need for effective communication and collaboration skills in managing critical situations.

The study also uncovered that some participants, including G3-04, exhibited elevated baseline heart rates—some surpassing 110 beats per minute—before the simulation began, indicating pre-simulation stress. This observation suggests that the stress associated with anticipating the simulation tasks can significantly influence baseline HR measurements.

As suggested by the findings, integrating multimodal data into the debriefing process can significantly enhance the reflective learning experience. This approach is supported by D’Mello & Graesser (2012), who argue for the recognition and addressing of learners’ affective states to optimize learning. By providing objective, quantifiable data on stress responses, nursing teachers can tailor debriefing sessions to address both individual and team stress management strategies, potentially improving future performance in similar high-stress scenarios.

Furthermore, the study provides a starting point for future CSCL innovations prioritizing learners’ emotional and psychological well-being. The observed connection between physiological stress indicators and the learning environment illuminates new pathways for developing educational technologies and teaching methodologies, equipping students to effectively navigate the demanding realities of clinical practice.

## Future work

The promising results from the multimodal data analysis in this study suggest impactful future applications, particularly in enhancing debriefing sessions post-simulation with physiological feedback to provide targeted feedback and actionable strategies for stress management. Delayed debriefings before simulation could further solidify this learning, optimizing stress-coping mechanisms. Teachers can utilize these insights to tailor feedback and provide specific, practical guidance, fostering an improved learning environment that resonates with the CSCL community’s goal of applying research to enrich collaborative learning practices.

## References

- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), Article 2. <https://doi.org/10/ggkm7c>
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In *Collaborative learning: Cognitive and Computational Approaches* (pp. 1–19). Elsevier.
- D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. <https://doi.org/10.1016/j.learninstruc.2011.10.001>
- Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education*, 46(11), 1055–1062. <https://doi.org/10.1111/j.1365-2923.2012.04355.x>
- Issenberg, S. B., Mcgaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher*, 27(1), 10–28. <https://doi.org/10.1080/01421590500046924>
- Sánchez, D., Kaliisa, R., Esterhazy, R., & Damsa, C. (2024). Unpacking the Elicitation Process for Multimodal Learning Analytics based Feedback. *Companion Proceedings 14th International Conference on Learning Analytics & Knowledge (LAK24)*. [https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24\\_CompanionProceedings.pdf](https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24_CompanionProceedings.pdf)
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467. <https://doi.org/10.1007/s11412-017-9267-5>
- Yan, L., Martinez-Maldonado, R., Zhao, L., Li, X., & Gašević, D. (2023). Physiological Synchrony and Arousal as Indicators of Stress and Learning Performance in Embodied Collaborative Learning. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 602–614). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-36272-9\\_49](https://doi.org/10.1007/978-3-031-36272-9_49)
- Zhao, L., Yan, L., Gasevic, D., Dix, S., Jaggard, H., Wotherspoon, R., Alfredo, R., Li, X., & Martinez-Maldonado, R. (2022). Modelling Co-located Team Communication from Voice Detection and Positioning Data in Healthcare Simulation. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 370–380. <https://doi.org/10.1145/3506860.3506935>

## Acknowledgments

This work has been partially funded by the Research Council of Norway.

## Evaluating Metaspace as an Infrastructure for Teacher Cross-Community Collaboration in Knowledge Building

Guangji Yuan, Nanyang Technological University, [guangji.yuan@nie.edu.sg](mailto:guangji.yuan@nie.edu.sg)  
Bodong Chen, University of Pennsylvania, [cbd@upenn.edu](mailto:cbd@upenn.edu)  
Chew Lee Teo, Nanyang Technological University, [chewlee.teo@nie.edu.sg](mailto:chewlee.teo@nie.edu.sg)  
Carol K. K. Chan, University of Hong Kong, [ckkchan@hku.hk](mailto:ckkchan@hku.hk)

**Abstract:** In this paper, we analyze a case of teacher-cross-community collaboration and knowledge-building professional development in an online space named Metaspace. Teachers from multiple regions of two countries, China and Singapore, shared their knowledge-building practices and student artifacts in metaspace. The analysis of teacher interviews illustrated how the metaspace design enhanced the professional development of teachers as a collaborative infrastructure. Researchers identified salient themes to review a more conceptual understanding of teacher knowledge-building, professional development, and network-building as a multi-level dynamic process.

### Introduction

The trend toward teacher professional development (PD) is a testament to the dynamic and emergent nature of educational ecosystems, especially in knowledge-building communities. Scaling up such communities for teacher development is essential for continuous growth across multi-level education infrastructure (Chan et al., 2011; Law et al., 2021; Scardamalia et al., 2017). Knowledge Building (KB) pedagogy, which is guided by twelve principles and emphasizes collective idea improvement over individual learning (Scardamalia & Bereiter, 2014), has been adopted in many countries, reshaping the teaching and learning paradigm. KB teachers are crucial designers and facilitators who orchestrate emergent, dynamic learning processes that require a deep understanding of the KB principles (Chan et al., 2011, Zhang et al., 2017, Yuan et al., 2022). Thus, professional development in KB is necessary to prepare teachers to perform such roles. While localized teacher communities have made strides in spreading practices, the geographically dispersed nature of such communities has hindered the upscaling and expansion of collaborative systems. This underscores the necessity of adopting innovative designs to facilitate the creation of cross-community spaces for teacher PD. Additionally, it is imperative to understand the processes underlying teacher KB adoption to mitigate the ambiguity surrounding the manner in which such a new upscaling space affords teachers opportunities for innovative reforms. Learning in a cross-community space is vital, as it bridges the gaps between isolated KB practices and promotes cohesive and dynamic learning for professional development. A KB metaspace is a space that enables teachers to reflect on their teaching strategies, share their teaching stories, and communicate with other KB teacher fellows worldwide, reviewing analytics to support their interaction (Teo et al., 2022). Such innovative pedagogical designs showcase the metaspace as a model for CSCL for cross-country collaboration. Investigating the experiences of teachers from diverse regions of Singapore and China exemplifies the dynamic cross-community interactions that provide new practical knowledge and opportunities for upscaling professional development towards a broader KB network.

### Metaspace as a meso-level infrastructure for knowledge building

When could an online space become an infrastructure for teacher professional development in knowledge building? The traditional examples of infrastructures, such as power grids, the internet, and file folders are characterized by their transparent usage and multiple configurations (Star & Ruhleder, 1994). These infrastructures are put in place to support social activities and are transparent to use when functioning. Similarly, in the context of KB, the online metaspace is conceptualized to serve as an essential infrastructure for the spread and sustainment of KB practices across geographically dispersed communities. Creating such infrastructure involves the creation of metaspace for teachers' PD and KB practices, which focuses on the design efforts in creating conditions, space, time, participants, goals, contexts, and culture that support making novice designs and coordinate into the existing network infrastructure. There have been multiple networks built for improving knowledge building practices. For instance, the Institute for Knowledge Innovation and Technology (IKIT) comprises a multidisciplinary team of researchers, scientists, teachers, engineers, and policymakers worldwide (Scardamalia et al., 2017). Leveraging technological advancements, we have witnessed a multifaceted enhancement in the resources and tools available for KB teacher development. The Knowledge Building Teacher Network (KBTN) in Hong Kong was established through a teacher secondment program in 2006 and subsequently grew through the school-university partnership project, which received funding from the Education Bureau (the

equivalent of the Ministry of Education in Hong Kong; Chan et al., 2011). In addition to these KB teacher networks, Knowledge Building Connect Network Learning (KBCNL) was founded in Singapore to increase the connections among local KB teachers. However, the challenge persists in crafting a space that fosters and nurtures teacher networks across diverse communities, ensuring accessibility and relevance at a larger scale. Design and redesign a KB metaspace that assembles diverse stakeholders to tackle the intricacies is critical as part of an infrastructuring process that deals with the interrelations among the existing elements within the system (Chen, et al, 2022) in multiple dimensions: management of multiple stakeholders and policies, resources, analytics, research designs, and curriculum adjustment (Teo, 2017). A meso-level refers to a network-level of teacher network building, which supports teacher change towards classroom innovation. In contrast, previous KB teacher's PD studies mainly focus on the micro-level classroom design, which refers to using principles, pedagogy, and technology that are integrated at the classroom level (Chan, 2011). This conceptualization recognizes holistic multi-layered relations involved in educational change and opens up venues for designing and researching infrastructures for learning – by teachers and students. The metaspace in the Knowledge Forum platform (KF; Scardamalia & Bereiter, 2006) was designed as an open and organic space for cross-community collaboration and interaction. Stakeholders (school leaders, researchers, professors, and KB practitioners) can share and access KB stories based on different disciplinary content. The KB story contains critical information about the whole classroom's idea development over a learning unit, such as big wondering questions, main topics students initiated and developed, what they have learned as a community, and resources that students shared. Moreover, based on the KB stories, teachers shared their meta-cognitive reflection notes, with specific reflection scaffolds, to build on each other's thoughts and deepen their understanding of KB principles and practices. To further understand how teachers perceive the metaspace, this paper asks how the KB metaspace enhances teachers' professional development as a collaborative infrastructure and further deepens the knowledge-building discourse.

## Methodology

### Context and participations

Initiated in August 2021, this project aimed to foster collaboration across various communities using the KF. The principal participants were six teachers from China and three from Singapore, and together, they shared 8 lesson plans in the format of Knowledge Building Story (KB Story), including one co-taught lesson. Throughout the online sharing event, a diverse group of 141 individuals attended and shared their discussion, including teachers, researchers, administrators, principals, policymakers, and master students. Regular monthly design meetings were conducted within the research teams to design the metaspace before the event. The collaborative design, facilitated through the online platform KF, encouraged teachers to create artifacts to prepare for the workshop and engage in reflective practices and cross-community interactions on the KF metaspace (Figure 2) during and after the professional development (PD) workshop. Three virtual network meetings were organized to support synchronous communication, witnessing participation from additional school and network members in September 2022.

### Data resources and analysis

Data collected for this study comprised six post-event interviews conducted after September 2022. Each interview, spanning 40-60 minutes, revolved around four central themes: principles of knowledge construction, the metaspace for knowledge construction and cross-border learning, teaching and research activities in cross-border learning, and prospects. The interviews yielded approximately 300 minutes of data, with four conducted in Chinese and two in English. The Chinese interviews were transcribed and translated into English for analysis. Using Nvivo (2020), researchers applied a grounded-theory approach to systematically analyze, create open codes, and reorganize the data for primary themes to capture the salient patterns. This approach facilitated the development of distinct codes to encapsulate the prominent perspectives of teachers on the subject matter.

## Results: How did the KB metaspace enhance teachers' professional development as a collaborative infrastructure?

The interview data revealed five main themes illustrating how metaspace enhances KB practices and teachers' professional development.

### Theme 1: Facilitating communication and idea exchange

The meatspace provides a vital platform for sharing ideas, teaching practices, and KB strategies across international borders. As teachers reflected: "The KB metaspace is like a bridge, connecting us across countries...It is invaluable for our professional growth." Another teacher reflected: "It is equivalent to providing

such a platform, which allows teachers in China and Singapore to have a platform for Knowledge Building teaching and can carry out exchanges and cooperation... express some of our questions and ideas." These reflections suggest that the metaspaces enabled teachers from multiple geographically distanced regions to exchange their KB experiences through KB discussion, and KB, with both Zoom synchronous and KF asynchronous discussion, metaspaces contributed to the vibrant knowledge sharing environment and fostered a global KB community of practice.

### Theme 2: Inspiration for new knowledge building pedagogical approaches

The metaspaces inspired teachers to generate new ideas when sharing teaching strategies and organizational methods. Teachers could observe and learn from others' teaching examples, leading to awareness and adoption of innovative approaches. For instance, one teacher shared: "Seeing how a colleague integrated war history into a lively discussion was eye-opening. It pushed me to think beyond textbooks and bring real-life topics into my classroom." Another teacher reflected that: "I am more impressed by... class examples. I was thinking, and maybe we can combine topics with our life... open topics, like war, students may have more different, that is, more exciting views." These reflections highlighted how teachers were inspired by other peers' examples and strategies, indicating their willingness to adopt new approaches demonstrated in the KB metaspaces to create a more student-driven learning environment.

### Theme 3: Collaborative discourse through KB stories and reflection

The sharing of KB stories and teachers' reflections after the sharing sparked a new round of KB talk in the metaspaces. As the main design of the boundary object and knowledge artifact in metaspaces, teachers used these KB stories to share specific subject-based KB cases and teaching experiences, which were reflected as valid by other teachers to understand the procedure of KB teaching practices. For instance, one teacher illustrates that: "This kind of Knowledge Building story is actually based on the presentation of our class examples, which may affect each other, and then whether it is possible for us to propose such a topic, other teachers may see some of their own thinking, and then we see some examples of other teachers, whether it is their discipline or not, there will definitely be some ideas, and then it may be the exchange of this special case, which is beneficial for us to continue to carry out our Knowledge Building teaching." This also resonates with the design of the KB story as the boundary object with the same structure to make it easier for other newcomers to access the existing knowledge.

### Theme 4: Interdisciplinary interaction with themes

The metaspaces invited teachers from multiple subject domain areas, providing new opportunities to learn different strategies and increasing the potential for innovative cross-subject practices. For example, one teacher noted, "It will include teachers from various disciplines as well as professors... learn from one another about construction examples from various disciplines." The metaspaces promoted interdisciplinary understanding, allowing teachers to exchange insights from various fields and apply them based on their interests.

### Theme 5: Deepening pedagogical knowledge

Teachers can share their insights during the sharing, and advanced, experienced professors and researchers in the field also support them. Such leveraged understanding was crucial as they are the ones to guide their classroom practices in applying various KB principles and analytical tools effectively. For instance, a teacher mentioned: "Engaging with the metaspaces has deepened my understanding of KB principles. It is one thing to know the theory; it is another... to see it in action and apply it effectively in my teaching."

## Discussion

The creation of meso-layer infrastructure for cross-country teacher PD presented in this study is an innovative design in educational reform, as teachers are the core of such transformation. This paper investigated teachers' experiences and perception of the metaspaces as an infrastructure facilitating cross-community collaboration in knowledge building communities. This study explored new designs of collective teacher metaspaces with the designed knowledge artifacts like KB story; the metaspaces in this study functioned as a meso-layer infrastructure in supporting the unique and temporal connections among various spaces and stakeholders. Analysis of teachers' interviews revealed several salient patterns and themes in how metaspaces benefit teachers' professional growth. Metaspaces, as an intermediate space that supports teachers' idea exchange and build-on, disseminates the teaching practices and localized designs to a broader level. Such exchanged ideas further sparked new ideas for teachers to test in their home schools. At the same time, some teachers also expressed the need for more extended learning units so that teachers from different regions can have a standard co-design learning unit that enables students from



both ends to join the longer KB building to extend the KB learning. Taken as a whole, the results of this study revealed new insights into how KB metaspace facilitated teachers' PD through cross-community collaboration.

## Reference

- Chan, C., Law, N., Hui, D., & Fung, Y. H. (2011). Knowledge Building Teacher Network (KBTN) in Hong Kong: Sustaining and Scaling up Knowledge Building through Principle-Based Innovation.
- Chen, B. (2022). Toward a framework of infrastructuring for sustainable educational innovations.
- Law, N., Zhang, J., & Pepler, K. (2021). Sustainability and scalability of CSCL innovations. *International handbook of computer-supported collaborative learning*, 121-141.
- Lumivero (2020) Nvivo R1. Release 2020. Version 1.7.1 <https://support.qsrinternational.com/nvivo/s/>
- Scardamalia, M., & Bereiter, C. (2006). FCL and knowledge building: A continuing dialogue. *Institute for Knowledge Innovation and Technology*, University of Toronto.
- Scardamalia, M., Bereiter, C., Laferrière, T., Bielaczyc, K., Chai, S., Chan, C. K. K., Chen, B., Chen, M.-H., de Jong, F., del Castillo, F. D., Hakkarainen, K., Matsuzawa, Y., McAuley, A., Montané, M., Nunes, C., Reeve, R., Seitamaa-Hakkarainen, P., Oshima, J., Shirouzu, H., ... Zhang, J. (2017). Toward a multi-level knowledge building innovation network. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL*, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017 (Vol. 2, pp. 703–710). *International Society of the Learning Sciences*. <https://doi.org/10.22318/cscl2017.113>
- Star, S. L., & Ruhleder, K. (1994, October). Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 253-264).
- Teo, C. L. (2017). Symmetrical advancement: Teachers and students sustaining idea-centered collaborative practices. Keynote presented at the 12th *International Conference on Computer Supported Collaborative Learning*, June 18-22, Drexel University School of Education & University of Pennsylvania Graduate School of Education, Philadelphia, PA.
- Teo, C. L., Yuan, G., Feng, X., Ong, A., Tong, Y., Zhao, J., & Chan, C. (2022). Investigating teachers' reflections, roles of researchers, and collective inquiry in cross-community collaboration in KB Metaspace for teachers. Paper presented at the *Knowledge Building Summer Institute*, Toronto, Canada.
- Yates, JoAnne. 1989. *Control through Communication: The Rise of System in American Management*. Baltimore, MD: Johns Hopkins University Press
- Yuan, G., Zhang, J., & Chen, M. H. (2022). Cross-community knowledge building with idea thread mapper. *International Journal of Computer-Supported Collaborative Learning*, 17(2), 293-326.
- Zhang, J., Bogouslavsky, M., & Yuan, G. (2017). Cross-Community Interaction for Knowledge Building in Two Grade 5/6 Classrooms. In *Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL 2017)*. *International Society of the Learning Sciences*.

## Co-Design a Logic Model for Inclusive AI-Powered Learning Application with Primary School Teachers

Ariel Han, Seth Corrigan, Shenshen Han, Kylie Pepler  
hanjy3@uci.edu, scorrig1@uci.edu, shenshh@uci.edu, kpeppler@uci.edu  
University of California, Irvine

**Abstract:** This study explores the development of a logic model to guide the design and implementation of an AI-powered story-authoring platform tailored for young learners, particularly those facing literacy challenges, such as English language learners, students with disabilities (e.g., dyslexia), and people with low socioeconomic status (SES). Amidst the rise of Generative AI tools, there's a need to bridge the gap between AI advancements and their practical use in K-12 educational settings. To address this, the researchers collaborated with primary school teachers, and synthesized insights from the literacy education literature. We conducted a co-design process, involving six educators, to identify the platform's potential users, features, functionalities, and desired outcomes. The findings emphasized the importance of making writing more accessible, with a focus on engaging students and motivating them to write. The proposed logic model underscores strategies like gamified elements, culturally-responsive imagery, and AI-generated mentor texts tailored to individual student needs.

### Introduction

Used in the context of a backward design framework (Wiggins & McTighe, 2008), logic models can provide valuable guidance for the development of innovative educational technologies - clarifying a technology's target audience, functionality and desired impacts. The purpose of the current study is to report on development of a logic model that informs the design and development of a cutting-edge AI-powered story-authoring platform for young learners. The design process incorporated classroom teachers in order to understand their unique perspectives and needs. Furthermore, we incorporated current research on literacy education, in order to inform the final inputs and outputs of the resulting logic model (Graham et al., 2012). Lastly, we aimed to integrate inclusive design elements, utilizing diverse learner variability navigators provided by Digital Promise in order to meet the unique needs of individual learners (Digital Promise., 2023).

The recent launch of Generative AI (i.e., ChatGPT, Dall.E3) sparked educational discourse, raising both fears and hopes of adapting the cutting-edge platforms in education (Han et al., 2024; OpenAI, 2023; Kasneci et al., 2023). However, there is a gap in the field's understanding of how to best leverage AI technologies for improved student learning. We utilize a combined backward design and co-design approach in concert with logic model development in order to bridge this gap. A key part of our strategy involves collaborations with primary school teachers to build logic models in order to specify well the desired input, activities, outputs, and outcomes associated with the eventual technology. Through the process, teachers' expertise in existing methods for teaching and learning literacy skills is integrated with the development team's understanding of the affordances and limitations of current AI platforms.

We co-design a logic model as a collaborative effort of six educators (K-12 classroom teachers) to identify who might be users and what context these new systems can fit in, and what features and functionalities are needed for desired outputs for an AI-powered story-authoring platform. We recorded co-design sessions and analyzed the resulting qualitative data to elicit teachers' intended goals, needs, and their suggestions. We also drew the insights from synthesizing several lines of scholarly literature including literacy education to identify desirable outputs (what learners achieve) (Graham et al., 2012). Additionally, we leveraged the learner variability navigator as we developed inputs and activities (features and interventions) for inclusive design. As a result of our findings, we were able to identify the target users, desired behaviors and activities, what users learn, and the ultimate purpose of the tool (Digital Promise, 2023).

The contribution of the study to the learning science community is two-fold: first, we collaborated with classroom teachers (K-12) to understand their opinions, with the aim of developing a logic model to inform designing learning applications tailored to their specific needs and motivations. As a first step towards designing and developing a teacher-centered AI-powered educational platform that meets their unique needs. This collaboration underscores an opportunity to foster joint efforts between researchers, educators, AI, and NLP specialists to design a AI-powered platform for K-12 education. Second, we present the process of constructing a causal pathway, incorporating strategies to achieve desired outcomes, in the creation of innovative educational technology, focusing on the development of creative AI-powered learning tools—a logic model that offers

valuable insights to learning designers and researchers alike to find features and functionalities of integrating such emerging tools for educational purpose.

## Background

### Logic model

A logic model is a conceptual framework that specifically highlights the causal pathways leading to desired outcomes (Stegemann & Jaciw, 2018). The logic model provides guidance on the big picture of how key features are intended to achieve the goals of a given product or intervention. Logic models have been a long term feature in educational research and design (Stegemann & Jaciw, 2018). Their utility derives from their ability to systematically map out the expected functionality and impact of a product and guide design decisions. Logic models identify the expected context and users, inputs and activities (i.e., features and functionality), and describe how these influence the user's experience, the users' outputs (i.e., user experience data), and the expected outcomes (i.e., potential benefits of using the product).

### AI-powered literacy development

The rapid advancement of Generative Artificial Intelligence (GAI) models, such as Large Language Models (LLMs) and Text-to-Image (TTI), make it possible for them to learn patterns and structures from existing data and to generate new content. These breakthroughs have led to a new generation of storytelling systems that enable open-ended conversation and create pedagogically beneficial text and images (Han & Cai, 2023; OpenAI, 2023). In these ways, GAI models expand the ability of teachers to facilitate open-ended discussions with underserved youth (Han et al., 2024). This new technology opens up new possibilities for developing literacy education platforms built on top of the new technology capabilities - such as creating artificial intelligence agents to facilitate human-to-AI communication (Brusilovsky, 2023). To integrate such emerging technology effectively into educational settings, it is critical to understand the practical challenges and needs faced in classrooms and the complexities of teachers' roles, deeply rooted in the reality of teaching and learning (Han et al., 2024; Long et al., 2021).

## Methods

We conducted co-design sessions with six primary school teachers in the United States. The co-design sessions took place individually via video conferencing due to geographical distances, with an average length of approximately one hour and a half for three days (i.e., once a week) between July and September 2023. Teachers were recruited through snowball recruitment as part of researchers' personal networks, with the only criteria for eligibility being that they were either current or former K-12 teachers. The teachers were classroom teachers from 1st to 8th grades, and their experience averaged 15.8 years (min=2.3 years, max=31 years).

On the first day of the co-design session, we sought to elicit teachers' current practices, their struggles, and their motivations when they teach writing to their students. Afterward, in the second session, we introduced GAI platforms (i.e., features and functionalities). We asked about their experiences and opinions about adapting it in educational settings specific to writing activities with their students. Then, co-developing logic models and sharing feedback and suggestions was discussed. Participants were compensated \$25 for their time and effort. Our study was approved by the authors' institutions' institutional review boards (IRBs).

## Data sources and analysis

We analyzed the qualitative data with a deductive approach from the co-design sessions. First, the interview data was automatically transcribed (Otter.ai) using the original audio and aligned with the transcript. The transcript was transferred to qualitative data analysis software (Atlas.ai) for the first round of coding. We conducted an deductive approach to analyze interview data (Azungah., 2018) from the logic model structure to identify *context*, *inputs*, *output*, and *outcomes*. We read the transcripts and identified relevant themes of the text. We categorized codes into four high-level themes (i.e., context, inputs, outputs, and outcomes) and 27 codes under each theme (see Table 1).

## Findings

From the findings, we identified the context and users of the AI-powered story-authoring platform for struggling writers and low-level readers. The results revealed input stations in which what students do, what features are necessary in the platform encompass: providing playful literacy activities, text chats, selecting culturally responsive imagery, writing models, feedback, and translanguaging (Digital Promise, 2023). Drawing from teachers' interviews and synthesizing evidence-based research literature, we demonstrated effective strategies in

literacy instruction for youths as outputs of the logic model, which is about what students learn as follows, genre-specific writing strategies, fluent in spelling, typing, word processing, lexical diversity, and verbal reasoning (Graham et al., 2012). Lastly, we drew desired outcomes from interview data based on teachers' perspectives, such as improving students' self-efficacy and motivation in writing, as articulated by one teacher:

I think that as a writing teacher, a lot of us want the students to feel curious, feel confident, feel motivated, feel activated as writers are encouraged. I think a lot of writing teachers will say 'I can work with the student whose writing is weak', but getting them to start, getting them to continue and finish, and then getting them to revise. Those are the psychological hurdles, right?

A teacher's goal is to support a student's writing activities by providing writers' tools through the platform, which we consider to be an input of the platform. What we provide as features includes providing age-appropriate and immediate feedback to students and scaffolding mentor text that is helpful to students in developing their own writing.

We can ask students, do you want to use the thesaurus? Not only are you letting them have a choice, but you're reminding them of writers' tools. I think what you want your writers to do is to be able to improve their writing, by improving their workflow choices, their structure, sentence skills, organization, and so paragraphing.

Teachers are intrigued by the potential of generative AI (LLM) to offer customized mentor texts, catering specifically to students' needs, including varying vocabulary levels and phonics.

It still is very time consuming to make sure that you're finding good and mentor texts that fit with what you're teaching to make sure you're hitting direct discrete skills that are real. Using mentor text is really a lot of teachers' work. That is one of the ways that we can use AI – to help us develop appropriate levels of writing models for each student in different states.

Our findings indicated teachers identified that the integration of gamification and culturally relevant imagery generation via Text-to-Image AI art generator (TTL) serve as effective strategies to augment engagement and motivation within writing activities.

And it seems like it's absolutely a game that you can play with the AI. like, you write the description of the alien? you give the prompt, AI gives you the image.

Teachers highlighted the potential of utilizing AI art generators (TTL) to produce culturally relevant images, thereby amplifying student engagement during story creation.

Who's your character? Are they white? Are they black? You know, so that now they can have something that represents them? Right? If they're a black kid, they want to see a story about a black kid, they don't want to see a white kid with blond hair all the time.

Consequently, we identified the following logic models: *Context/Users* (struggling writers and low-level readers), *Inputs/Activities* (playful literacy activities, text chats, culturally responsive imagery selection, writing models, feedback, translanguaging, and writing strategies), *Outputs* (increased amount of time spent in writing, word count, genre-specific writing strategies), and *Outcomes* as fluent in spelling, typing, word processing, lexical diversity, syntax, and verbal reasoning, ultimately increase engagement and motivation in writing (see Table 1). These can be facilitated via the development of AI-powered story authoring platforms through conversational design, chatbot systems, and customizable editing stations.

**Table 1**  
*A Logic Model for AI-powered Story-authoring Platforms for Struggling Writers*

Context & Users	Inputs & Activities	Outputs	Outcomes
Make writing more accessible for those struggling writers and low-level readers	Playful literacy activities Text Chats Selecting Culturally Responsive Images Writing Models	Time spent in writing Word count Genre-specific writing strategies	Fluent in spelling, typing, and word processing Lexical Diversity Synthax, verbal reasoning
ELL Racial Minorities Low socioeconomic status households (SES)	Feedback Translanguaging Writing Strategies		Engagement (increased amount of time spent writing), Motivation in writing

## Discussion

The study represents an initial step in navigating the emergence of AI tools that can be integrated into learning experiences by constructing a logic model for the design and development of AI-powered story-authoring platforms tailored for young learners. In order for AI-driven educational technology to be effective, it needs to resonate with educators' needs, concerns, and aspirations. We developed our logic model as a result of our collaborative efforts with K-12 educators and synthesized academic literature. Our approach underscores the importance of involving educators from the beginning, fostering informed perspective on the design of learning technologies.

Our findings emphasize that while Generative AI offers exciting new possibilities, its adaptation in K-12 educational settings should be approached thoughtfully. Key aspects include ensuring cultural relevance, amplifying student engagement, and nurturing learners' motivation and self-efficacy in writing. The teachers' perspectives illuminated the value of integrating gamified elements and culturally-responsive imagery, which not only elevates the writing experience but also validates and respects students' identities.

Furthermore, the discussion around mentor texts, especially the potential for AI to offer tailored writing examples, highlights the transformative potential of AI. This could alter the teacher's role from that of a content provider to a facilitator, offering students more autonomy in their learning journey and equipping them with tools that cater to their specific needs. However, as with any innovative technology, challenges remain. The concerns about authenticity, agency, and the potential for bias or misinformation underscore the need for rigorous testing, regular updates, and perhaps most importantly, ongoing dialogue with educators and stakeholders.

## Conclusion

In conclusion, our study's contribution lies in not only providing a pathway for the design of AI-powered educational tools but also in emphasizing the value of a collaborative approach. By bridging the expertise of AI specialists with the lived experiences of educators, it is possible to ensure the development of tools that are both cutting-edge and grounded in real-world educational needs. Our proposed logic model should be implemented in diverse educational settings and the impact it has on student outcomes should be explored in future research.

## References

- Azungah, T. (2018). Qualitative research: deductive and inductive approaches to data analysis. *Qualitative research journal*, 18(4), 383-400.
- Brusilovsky, P. (2023). AI in education, Learner Control, and human-ai collaboration. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-023-00356-z>
- Digital Promise. (2023). Learner Variability Navigator: Learner Variability Project. Learner Variability Navigator | Learner Variability Project. <https://lvp.digitalpromiseglobal.org/>
- Graham, S., Bollinger, A., Olson, C. B., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). Teaching Elementary School Students to Be Effective Writers: A Practice Guide. NCEE 2012-4058.
- Han, A., & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. *Interaction Design and Children* <https://doi.org/10.1145/3585088.3593867>.
- Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Peppler, K. 2024. Teachers, Parents, and Students' Perspectives on Integrating Generative AI into Elementary Literacy Education. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '24*, May 11–16, 2024, Honolulu, HI, USA: Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642438>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). Chatgpt for good? on opportunities and challenges of large language models for Education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Long, D., Blunt, T., & Magerko, B. (2021). Co-designing ai literacy exhibits for informal learning spaces. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–35. <https://doi.org/10.1145/3476034>
- OpenAI. (2023, March 27). *GPT-4 technical report*. arXiv.org. <https://arxiv.org/abs/2303.08774>
- Stegemann, K. C., & Jaciw, A. P. (2018). Making It Logical: Implementation of Inclusive Education Using a Logic Model Framework. *Learning Disabilities: A Contemporary Journal*, 16(1), 3-18.
- Wiggins, G. P., & McTighe, J. (2008). *Understanding by design*. Association for Supervision and Curriculum Development.

## Source Memory and Collaborative Learning: The Role of Group Composition and Conflicting Information

Oktay Ülker, University of Duisburg-Essen, oktay.uelker@uni-due.de  
Daniel Bodemer, University of Duisburg-Essen, bodemer@uni-due.de

**Abstract:** Source memory, i.e., remembering the origin of information, is important for applying retrospective social learning strategies, such as academic help-seeking. We examined effects of group composition (heterogeneous: learning partners with differing knowledge levels vs. homogeneous: same knowledge levels) and conflicting information (with conflict vs. without conflict) on source memory and learning in a pseudo-collaborative learning scenario ( $N = 128$ ). Multinomial models allowed estimations of source memory unconfounded by guessing. Group composition and conflicting information did not influence learning but source memory: participants better remembered which partner shared certain information in heterogeneous groups with differing knowledge levels, especially the learning partners with the highest expertise. Contrary to our expectations, source memory was worse in a context with conflicting information. Considering source memory, and thus students' help-seeking processes, can help to reveal long-term effects of group awareness tools that present learners' knowledge levels, ultimately helping to derive holistic implications for collaborative learning.

### Introduction

Being aware of learning partners' cognitive attributes, such as their knowledge levels regarding learning topics (i.e., group awareness, Bodemer & Dehler, 2011), enables effective social strategies: questions can be directed to highly knowledgeable peers and assistance can be offered to those with lesser knowledge. Group awareness tools support such strategies by presenting, for example, knowledge related information of partners (Bodemer et al., 2018). However, effective strategies sometimes need to be applied in retrospect, after collaboration or after learners received certain information from their peers. For instance, teachers might form groups in classrooms with distributed expertise and information for one-time collaboration. After collaboration, a learner might remember a piece of information and then try to remember which learning partner shared it. This process helps to efficiently seek additional information from the right person or select the appropriate individual as a learning partner. Thus, for efficient academic help-seeking processes, remembering the source of information (i.e., source memory, Johnson et al., 1993), becomes crucial. According to Makara and Karabenick (2013), deciding whom to ask is essential for (successful) help-seeking. Some group awareness tools might support such decision-making (e.g., Ollesch et al., 2022). However, sometimes group awareness information may not be present after using the tool or after collaboration, underlining the potential relevance of source memory. In (computer-supported) collaborative learning (CSCL), effects of rather person-related aspects (like *group composition*) or rather content-related factors (like *conflicting information*) on learning and collaborative processes have been examined. However, only few studies have examined combined effects of group composition and conflicting information on learning. Additionally, it remains unclear how these factors affect source memory, and thus retrospective social strategies such as academic help-seeking.

Regarding *group composition*, heterogeneous groups, in which learners' knowledge levels or expertise differ, sometimes outperform homogeneous groups with the same (medium) knowledge levels (e.g., Zhang et al., 2016). Groups that differ in this regard profit from useful interactions between learners, like asking questions to more knowledgeable partners or providing explanations for less knowledgeable partners (Webb, 1989). Moreover, sources of information seem to be better remembered when they differ regarding their expertise or trustworthiness than when sources do not differ in those regards (Thomm & Bromme, 2016). In heterogeneous groups, source memory for the individual partners might also differ: for example, Nadarevic and Erdfelder (2013) have shown that source memory for sources with high credibility and low credibility is better than for sources with uncertain credibility. Source memory might be enhanced when remembering the source is beneficial in the given context.

When students seek information on a learning topic, they may encounter *conflicting information* (information which is inconsistent with prior beliefs or knowledge). The presence of conflicting points of view in a social context can lead to socio-cognitive conflicts, which can be considered as beneficial for learning (Mugny & Doise, 1978). According to the idea of constructive controversy (Johnson & Johnson, 2009), (socio-)cognitive conflicts, for example, motivate the search for more information when students deal with controversial ideas. Regarding source memory, studies show that in individual learning, the source of information is better remembered in situations where presented information is conflicting with prior beliefs (e.g., Maier & Richter,

2013). Readers pay more attention to sources to understand the reasons for the discrepancies, resulting in better mental representations of source-content-links (for a review, see Bråten & Braasch, 2018). We test whether these effects, which were found in individual learning scenarios, can be transferred to a collaborative learning scenario.

## Methods

### Participants, design, and hypotheses

The sample consisted of  $N = 128$  participants (90 female, 35 male, 3 diverse) with their age ranging between 17 and 35 years ( $M = 21.32$ ,  $SD = 3.11$ ). Most of them (126) were students. We applied a  $2 \times 2$  design with the between-subject factors *group composition* regarding the knowledge level of the learning partners (heterogeneous group vs. homogeneous group) and *conflicting information* regarding the to be learned topic (with conflict vs. without conflict). Participants were randomly assigned into the four experimental groups (each  $n = 32$ ).

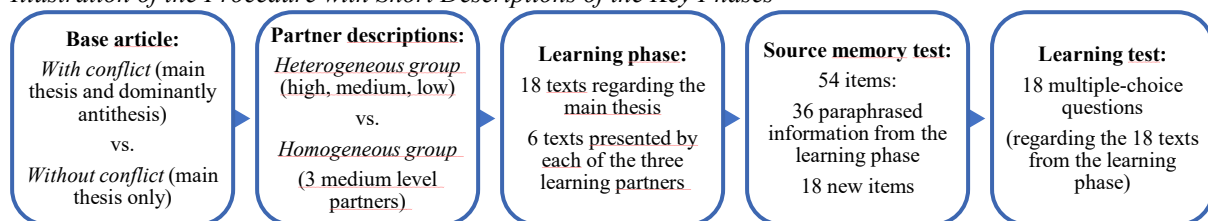
In our pseudo-collaborative study, participants received information from three learning partners and conducted a source memory test and knowledge test afterwards. We expect learning outcome to be higher in heterogeneous groups than in homogeneous groups ( $H-L-1$ ) and higher in contexts with conflicting information than without ( $H-L-2$ ). Regarding source memory, we assume better source memory in heterogeneous groups than homogeneous groups ( $H-SM-1$ ). Moreover, we expect source memory to be better for the high knowledge partner ( $H-SM-2a$ ) and the low knowledge partner ( $H-SM-2b$ ) than the medium knowledge partner, as remembering these sources helps to indicate the reliability of information and whether certain information might be right or wrong. Finally, source memory should be better in contexts where learners receive information which is conflicting with prior received information ( $H-SM-3$ ).

### Material and procedure

Phases of the experiment are depicted in Figure 1. The learning subject covered theories of the mass extinction of dinosaurs: the meteorite hypothesis (henceforth *main thesis*) and the volcanism hypothesis (henceforth *antithesis*). Participants were initially given base articles. Here, in the “with conflict” groups, participants were informed about two controversially discussed hypotheses regarding the extinction of the dinosaurs. Importantly, the text was heavily in favor of the antithesis. Participants in the “without conflict” groups were only informed about the main thesis. Next, participants were informed that the following information regarding the main thesis (allegedly) originated from the collaboration of three students, whose knowledge levels were measured with a validated test: in the groups “heterogeneous group”, the three partners had differing knowledge levels (high, medium, low), while in the groups “homogeneous group”, the three partners all had medium knowledge levels. Participants were instructed to imagine to learn collaboratively with the three partners, who were represented by randomly assigned names and avatars. Stars in avatars depicted their knowledge levels (3 stars = high, 2 stars = medium, 1 star = low). Participants were informed about the upcoming learning test, however, not about the source memory test.

**Figure 1**

*Illustration of the Procedure with Short Descriptions of the Key Phases*



In the learning phase, participants received information on the main thesis through 18 texts with similar length ( $M_{\text{Words}} = 50.94$ ). Each partner was randomly assigned to the texts and presented 6 texts on single pages. Participants could spend between 30 and 60 seconds per text. A source memory test then followed: participants were presented with pieces of information which were either paraphrased information from the texts in the learning phase or new information, which were not presented before. Participants indicated whether a piece of information was new or which of the three partners presented it. The test consisted of 54 trials, with 2 paraphrased information from each of the 18 texts and 18 new information. Learning outcome was assessed with a learning test consisting of 18 multiple-choice questions, each based on one of the texts from the learning phase. Finally, participants provided self-assessed prior knowledge (−3, “very low” to +3, “very high”), which was rather low ( $M = -1.12$ ,  $SD = 1.57$ ) and did not differ between experimental groups.

## Results

### Learning outcome

A  $2 \times 2$  ANOVA regarding learning outcome (see Table 1 for descriptives) revealed no significant main effects of group composition,  $F(1, 124) = 2.10, p = .150, \eta_p^2 = .02$ , or conflicting information,  $F(1, 124) = 0.23, p = .630, \eta_p^2 < .01$ , and no significant interaction,  $F(1, 124) = 0.50, p = .479, \eta_p^2 < .01$ , contradicting *H-L-1* and *H-L-2*.

**Table 1**

*Means (and standard errors) of Learning Outcome (number of correctly solved questions) and Source Memory Parameter Estimates*

	Heterogeneous		Homogeneous	
	With conflict	Without conflict	With conflict	Without conflict
Learning outcome	11.66 (0.32)	11.53 (0.26)	12.06 (0.24)	12.72 (0.27)
Source memory				
High knowledge partner	.33 (0.06)	.37 (0.06)		
Medium knowledge partner(s)	.03 (0.09)	.13 (0.08)	.02 (0.03)	.17 (0.03)
Low knowledge partner	.24 (0.06)	.31 (0.07)		

Source memory parameter estimates were analyzed with multinomial processing tree models and reflect the probability of remembering the learning partner as the source of information.

### Source memory

Source memory (see Table 1 for descriptives) was analyzed with the multinomial processing tree (MPT) model of source monitoring (Bayen et al., 1996), adapted for three sources (e.g., Nadarevic & Erdfelder, 2013). Such models allow to estimate probabilities of (source) memory processes unconfounded by often found guessing biases. We found a base model which fit the data well,  $G^2(27) = 30.09, p = .310$ . In the base model, the source memory parameters were restricted within the groups with homogeneous group compositions, meaning that source memory does not differ when partners all have the same (medium) knowledge level. In accordance with *H-SM-1*, sources were better remembered in heterogeneous groups than in homogeneous groups,  $\Delta G^2(2) = 28.55, p < .001, w = .06$ . High-knowledge sources are better remembered than medium-knowledge sources,  $\Delta G^2(2) = 12.22, p = .002, w = .06$  (supporting *H-SM-2a*). However, source memory for low-knowledge partners was (barely) not significantly better than for medium-knowledge partners,  $\Delta G^2(2) = 5.86, p = .053, w = .04$  (contradicting *H-SM-2b*). Additionally, contrary to *H-SM-3*, source memory was better in contexts *without* conflicting information than in contexts with conflicting information,  $\Delta G^2(2) = 11.32, p = .003, w = .04$ .

## Discussion

Group composition and conflicting information did not influence learning, contradicting *H-L-1* and *H-L-2*. Heterogeneous groups profit from interactions between more and less knowledgeable learners, like asking questions and providing explanations (Webb, 1989). Also, beyond cognitive conflicts, discussions of controversies and consensus-finding may be decisive (Johnson & Johnson, 2009). However, our study was pseudo-collaborative and thus lacked (useful) interactions between learners, potentially explaining why there were no effects on learning. We conducted a pseudo-collaborative study because our aim was to systematically examine the effects of group composition and conflicting information on source memory. Therefore, we prioritized high internal validity, for example, by controlling for linguistic expressions and standardizing the information, as well as standardizing the number of contributions from learning partners).

Source memory, however, was affected by group composition and better in heterogeneous groups. This supports *H-SM-1* and highlights the benefits of heterogeneous group compositions for learners, not only in achieving better learning outcomes (in real-collaboration), but also regarding better source memory, which can support help-seeking. While high-knowledge partners were better remembered than medium-knowledge partners (supporting *H-SM-2a*), low-knowledge partners were (barely) not significantly better remembered than medium-knowledge partners (contradicting *H-SM-2b*). In collaborative learning, learners seem to focus on which information might be right and less on which information might be wrong. Also, conflicting information impaired source memory, contradicting *H-SM-3*. Future studies could determine if socio-cognitive conflicts increase cognitive load, which can affect memory for the truth and falsity value of information, but not memory for the information itself (Nadarevic & Erdfelder, 2019), potentially explaining why there was no effect on learning.

Source memory was overall poor, especially for medium-level partners. However, medium-level partners need to be remembered as sources, so that they can be targeted questions in help-seeking processes. Instructions could direct learners' attention to the sources, possibly by integrating them into group awareness tools. Yet, future



studies need to test such instructions, as undesired effects might arise: in some contexts, better source memory is associated with worse memory for the information itself (e.g., Bell et al., 2022).

We demonstrated the use of MPT models in CSCL research, considering guessing biases. Such models can also be relevant for various CSCL scenarios and social memory processes, such as partner modeling (for an application of MPT models to analyze partner modeling, see Ülker & Bodemer, 2023). Some group awareness tools present knowledge related information of peers and aid learners in help-seeking processes in the moment of tool usage (e.g., Ollesch et al., 2022). Our study demonstrated that such tools could also aid learners in their long-term behavior, supporting source memory to know whom to ask for more information or to correctly assess the credibility of information, even when tools are not present anymore. Considering and adequately assessing source memory in future CSCL research may ultimately support learners in acquiring more (reliable) knowledge.

## References

- Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 197–215.
- Bell, R., Mieth, L., & Buchner, A. (2022). Coping with high advertising exposure: a source-monitoring perspective. *Cognitive Research: Principles and Implications*, 7, 82.
- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, 27(3), 1043–1045.
- Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). Routledge.
- Bråten, I., & Braasch, J. L. G. (2018). The role of conflict in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 184–201). Routledge.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28.
- Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher*, 38(1), 37–51.
- Maier, J., & Richter, T. (2013). Text belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, 31(2), 151–175.
- Makara, K. A., & Karabenick, S. A. (2013). Characterizing sources of academic help in the age of expanding educational technology: A new conceptual framework. In S. A. Karabenick & M. Puustinen (Eds.), *Advances in help-seeking research and applications: The role of emerging technologies* (pp. 37–72). Information Age Publishing.
- Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology*, 8(2), 181–192.
- Nadarevic, L., & Erdfelder, E. (2013). Spinoza’s error: Memory for truth and falsity. *Memory & Cognition*, 41(2), 176–186.
- Nadarevic, L., & Erdfelder, E. (2019). More evidence against the Spinozan model: Cognitive load diminishes memory for “true” feedback. *Memory & Cognition*, 47(7), 1386–1400.
- Ollesch, L., Kohlmann, J., Steiner, M., Tasdelen, O., & Bodemer, D. (2022). “uniMatchUp!”: An application for promoting academic help-seeking and group development. *Eleed*, 14(2).
- Schlusche C, Schnaubert L. & Bodemer, D. (2021) Perceived social resources affect help-seeking and academic outcomes in the initial phase of undergraduate studies. *Frontiers in Education*, 6, 732587.
- Ülker, O. & Bodemer, D. (2023). Remembering the knowledge of experts and novices in computer-supported collaborative learning: A multinomial processing tree approach. In JL. Shih, A. Kashihara, W. Chen & H. Ogata. (Eds.), *Proceedings of the 31st International Conference on Computers in Education*, Vol. 1 (pp. 286–291). Asia-Pacific Society for Computers in Education.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13(1), 21–39.
- Thomm, E., & Bromme, R. (2016). How source information shapes lay interpretations of science conflicts: Interplay between sourcing, conflict explanation, source evaluation, and claim evaluation. *Reading and Writing*, 29(8), 1629–1652.
- Zhang, L., Kalyuga, S., Lee, C.H., Lei, C., Jiao, J. (2015). Effectiveness of collaborative learning with complex tasks under different learning group formations: A cognitive load perspective. In: S. K. S. Cheung, L.F. Kwok, H. Yang, J. Fong, R. Kwan (Eds.), *Hybrid Learning: Innovation in Educational Practices* (pp. 149–159). Springer.

## Detecting Patterns of Idea Novelty and Complexity in Student Knowledge Building Discourses

Guangji Yuan, Chew Lee Teo, Zheng Chu  
guangji.yuan@nie.edu.sg, chewlee.teo@nie.edu.sg, zheng.chu@nie.edu.sg  
Nanyang Technological University

Jianwei Zhang, University at Albany, SUNY, jzhang1@albany.edu

Mei-Hwa F. Chen, University at Albany, SUNY, mhc@cs.albany.edu

Tianlong Zhong, Nanyang Technological University, tianlong001@e.ntu.edu.sg

Chun Yen Chang-Sundin, University at Albany, SUNY, cchang5@albany.edu

Alwyn Vwen Yen Lee, Nanyang Technological University, alwyn.lee@nie.edu.sg

Ong Kian Keong Aloysius, Nanyang Technological University, aloysius.ong@nie.edu.sg

**Abstract:** This paper explores the application of a framework for idea novelty in students' discourse for knowledge building. Knowledge building promotes collaborative discourse among students and supports them in expanding collective community knowledge. However, students often go beyond the sharing of information, and they contribute novel ideas that are vital to deepening community knowledge and expanding collective inquiry. Novel ideas not only reveal the character and quality of the discourse but also show how the conversation may extend to deepen the understanding of a challenging topic. This study attempts to illuminate novel ideas from students as they engage in knowledge building using the analytical lens of novelty. Data analysis from exploratory analysis and multiple correspondence analysis revealed patterns of how students contribute novel ideas to sustain their conversation. Utilizing advanced Machine Learning techniques, this study effectively identified and quantified patterns of idea novelty and complexity in student discourses, enhancing the understanding of collaborative knowledge construction.

### Introduction

Novelty is a pivotal force driving the expansion of human capabilities to address escalating challenges and push the advancement of collective knowledge (Kauffman, 1995; Tria et al., 2014; Yuan et al., 2023). In Computer-Supported Collaborative Learning (CSCL), the interplay of novelty and creativity gains increasing research attention. Novelty, as defined in the Merriam-Webster dictionary, is "new and not resembling something formerly known or used" (2002). Novelty acts as a catalyst in tasks that are divergent, potentially restraining creativity in tasks that are convergent-related (Gillebaart et al., 2013). Knowledge Building (KB) is a pedagogy focused on the community member's contribution to advancing the community's knowledge (Scardamalia & Bereiter, 1991), highlighting the knowledge and information carried out via knowledge building discourses. More specifically, knowledge building discourse is conceptualized by Scardamalia as a discourse that transcends mere idea sharing to foster a deeper understanding and continually identifies more profound problems for investigation. Other researchers further investigated the patterns of knowledge building discourse, where they found how members in the community progressively generate deeper questions, the creation of explanations and theories, the examination of ideas and hypotheses through evidence, constructive use of sources, mutual idea build-on, and shared reflection (Zhang et al., 2007; Teo et al., 2022). This revealed the importance of collaborative contributions in building upon each other's ideas, leading to more sophisticated understandings to achieve higher learning goals. A critical aspect of sustaining this discourse is the ability of participants to make novel and quality contributions that can extend existing knowledge boundaries. However, such aspects still remain unknown. This study aims to uncover the patterns of idea novelty and complexity in real-world learning settings and evaluating the quality of their knowledge building discourse using sophisticated machine learning techniques.

### Idea novelty and detection techniques

The genesis of novelty is rooted in several theoretical frameworks. It is the harbinger of uncharted possibilities, enabling us to "expand the adjacent possible" (Kauffman, 1995). This theory, from biological evolution to innovation evolution, posits that new ideas emerge incrementally, closely related to existing concepts. This process involves a blend of modification and recombination, leading to the formation of meaningful associations. Building on the existing literature, scholars further raised other definitions. For instance, Tria and colleagues assert that novelty is not an isolated phenomenon but emerges within a "cloud of other potentially new ideas or

experiences" that are thematically adjacent and can be triggered by it (Tria et al., 2014, p. 1). Thus, this shared "entourage" of ideas is subject to the dynamics of expanding into the adjacent possible. The evolution of data mining techniques in computer science has revolutionized novelty detection, defining it as the recognition of test data that significantly deviates from training data. The practical significance and complexity of this task have spurred a plethora of methodologies. These include, but are not limited to, frequentist and Bayesian approaches, information theory, support vector methods, extreme value statistics, and neural networks. These methods, developed over recent decades, primarily focus on constructing models from training sets devoid of or containing minimal instances of novel classes. Based on these models, novelty scores are then assigned to the data (Pimentel et al., 2014, p. 216). Towards a dynamic understanding of knowledge and innovation, current analytical paradigms often emphasize a static state of knowledge. However, advanced techniques are needed to increase the detection of idea novelty and complexity patterns to comprehend the ever-evolving complexity and fast accumulation of vast amounts of data. Based on the aforementioned literature backgrounds and challenges, this paper seeks to fill the research gap and understand students' idea contribution patterns through collaborative knowledge building discourse by posing two questions: RQ1: How are the six aspects of idea novelty related to idea complexity? RQ2: How to promote and automate the thematic coding and detection process??

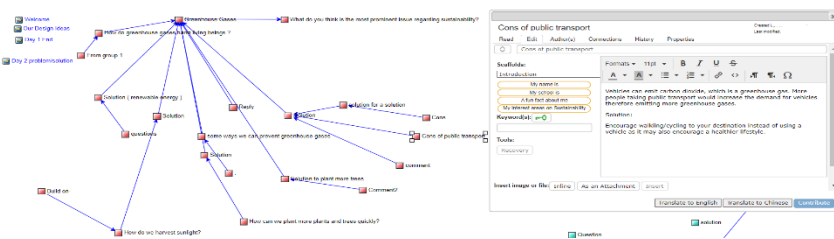
## Methodology

### Context and Background of Participants

Data was obtained from a summer program called Student Knowledge Building Design Studio (sKBDS, Teo, 2022; Yuan, 2023). sKBDS is a university-hosted initiative that brings together students from local schools to collaborate and tackle sustainability problems in a newly designed learning environment based on knowledge building principles. The data for this study were obtained from sKBDS in June 2021, November 2021, and October 2023. The three iterations involved 32, 32, and 35 local students, respectively. Data from three iterations was collected from an online Knowledge Forum(KF) platform.

**Figure 1**

*A Knowledge Forum Theme-based View with Students' Knowledge Building Discussions*



*A squared node illustrates each note, and connected lines between nodes indicate a build-on.*

### Data sources and analysis

Students' notes from KF served as the main data source to inform this study. The data included 155 notes, with an average of 16 words per note. These notes were selected due to the nature of the long extension of the build-on note structures (Figure 1). Students' novelty contributions were analyzed using an original framework developed by Yuan and colleagues (2023). The framework consisted of six dimensions, namely, New Concept, New Connection, New Context, New Rise Above, and New Question (details refer to Yuan, 2023). For measuring students' idea novelty, exploratory data analysis was applied to understand the potential associations among these six dimensions and complexity (1-unelaborated fact, 2-elaborated fact, 3-unelaborated explanation, 4-elaborated explanation; refer to Zhang et al., 2007). A summary of the data used in this study and the coded results is illustrated in Table 1. The interconnected relationship among each theme, indicates an overall correlation among these two main schemes. R package FactoMineR is applied for Multiple Correspondence Analysis (MCA) to investigate correlations among all dimensions, including six sub-dimensions in novelty and four in idea complexity. The multi-label function achieved by the Python BCE with Logits Loss package.

**Table 1**

*Coding Schemes and Results of Idea Complexity and Idea Novelty of the Current Study*

Scheme	Sub-Scheme	sKBDS	sKBDS	sKBDS
		Jun2021	Nov2021	Oct23
Idea Complexity	Unelaborated Facts	14	45	6
	Elaborated Facts	7	5	10

	Unelaborated Explanations	2	23	14
	Elaborated Explanations	0	3	26
Idea Novelty	New Concept	14	52	33
	New Connection	4	28	20
	New Rise-Above	0	2	4
	New Question	1	22	13
	New Source	0	5	3
	New Context	0	8	5

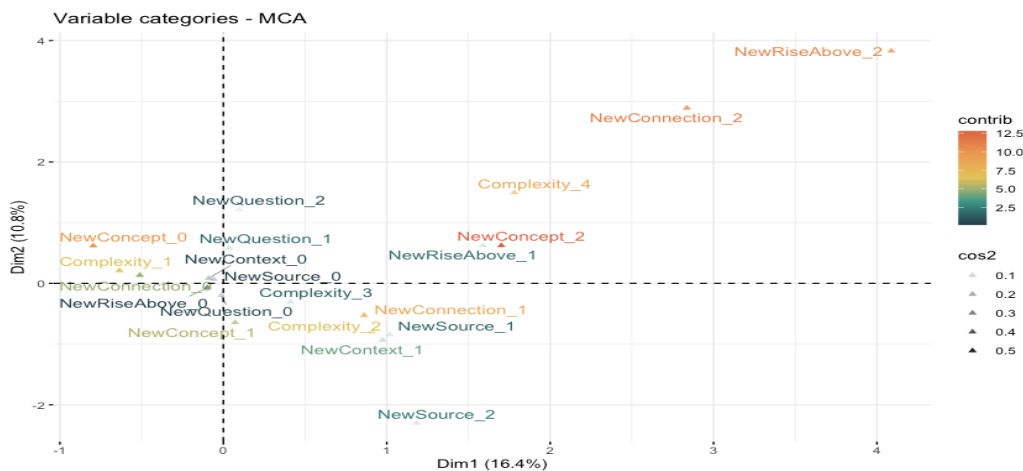
## Results

### RQ1: How are the six aspects of idea novelty related to idea complexity

To investigate the holistic relationship between novelty and idea complexity, we conducted a correlation study among these two themes, the Spearman coefficient is selected for this calculation. Besides the moderate correlation (0.48) between novelty and complexity, the six dimensions under the novelty framework also show patterns of correlation with complexity. The results show that a new concept and a new connection are substantially correlated (0.51). Moderate correlations are discovered between complexity and a new concept (0.49) and complexity and a new connection (0.42). There are slight positive associations between the following pairs: new source and new context (0.25), new rise above and new connection (0.25), complexity and new rise above (0.22). Negative correlations are negligible. Researchers further applied multiple correlation analysis (MCA), In multivariate and categorical analysis, multiple correlation analysis (MCA) extracts eigenvalue features corresponding to the amount of information retained by principal dimensions after dimensionality reduction (Figure 2). The color scales *cos2* and *contrib* refer to the confidence level of representation and the contribution level to the definition of dimensions 1 and 2. In quadrant 4 of Figure 2, it is shown that complexity 2, new context 1, new source 1, and new connection 1 are loosely clustered, indicating small associations. In quadrant 1, a substantial new connection and a substantial new rise above are also somewhat correlated.

**Figure 2**

*Multiple Correspondence Analysis (MCA) of Six Sub-indicators of Novelty and Idea Complexity*



### RQ2: How to promote and automate the thematic coding and detection process?

To further accelerate the note coding and detecting process, researchers manually coded all 155 with long build-on structures and tested three methods, namely BERT, LightSIDE, and GPT4.0, as state-of-the-art Large Language Models in Natural Language Processing. The LightSIDE application (Mayfield & Rose, 2013) for statistical machine learning algorithms tested with models such as Support Vector Machine (SVM), Naive Bayes, etc. It is worth to mention that a model with high accuracy does not necessarily indicate high reliability because, with an unbalanced dataset, if a model takes a random guess based on the statistical significance of one class, the kappa score will drop, and this phenomenon also means that the text features are not identified. The results show in Table 2 indicate that each detection method with the highest score in a specific category.

**Table 2**

*Comparative Performance of Machine Learning Models in Idea Complexity and Novelty*

		BERT	LightSIDE	Gpt4.0
Code Category	Subcode	Accuracy /Kappa Score	Accuracy/Kappa Score	Accuracy/Kappa Score
Idea Complexity	Unelaborated Facts	70.97%; 0.4583	70.32%; 0.4875	55%; 0.45
	Elaborated Facts			
	Unelaborated Explanation			
	Elaborated Explanations			
Idea Novelty	New Concept	48.39%; 0.0100	Naive Bayes; 62.58%; 0.3641	58%; 0.229
	New Connection	74.19%; 0.2832	SVM;75.48%; 0.4374	77%; 0.50
	New Rise-Above	93.55%; 0.0	Naive Bayes; 92.9%; 0.1243	66%; 0.10
	New Question	100%; 1.0	Logistic Regression 88.39%; 0.6481	86%; 0.64
	New Source	100%; NA	Naive Bayes; 89.68%; 0.0642	85%; 0.24
	New Context	96.77%; 0.0	SVM; 91.61%; 0.1135	79%; 0.14

## Discussion

Understanding how collaborative conversation expands over time is vital to creative discourse analysis. More specifically, detecting idea novelty and complexity is crucial to understanding the mechanisms of advancing a community's idea improvement. Members' new contributions play a core role in expanding the conversation through debates and build-ons. As new techniques are applied to detecting novelty, it is essential to understand how each dimension functions in active, real-life knowledge building communities. Using the idea novelty and complexity coding framework helped us understand the nature of the conversation and the behavior patterns in students' active intention to advance their knowledge. The findings of RQ1 shed light on the understanding the relationship and dynamics of knowledge building discourse in a collaborative real-world environment, which also highlights the importance of recognizing the potential novel contributions from students. RQ2 further explored the new machine learning techniques to accelerate the labeling process, and the accuracy score reveals a promising implication for the next step. Despite the limitations of the study, such as a limited amount of data for training, these results still show a promising way of adopting the new detection approach in further ML labeling, which will also shed light on larger dataset analysis for the next step.

## References

- Dictionary, M. W. (2002). Merriam-webster. *On-line at <http://www.mw.com/home.htm>*, 8(2).
- Kauffman, S. A. (1996). Investigations: The nature of autonomous agents and the worlds they mutually create. SFI working paper # 96-08-072. Santa Fe Institute.
- Mayfield, E., & Rosé, C. P. (2013). LightSIDE: Open source machine learning for text. In *Handbook of automated essay evaluation* (pp. 124-135). Routledge.
- Pimentel, M. A., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal processing*, 99, 215-249.
- Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the learning sciences*, 1(1), 37-68.
- Teo, C. L., Lee, A. V. Y., Ong, A., & Yuan, G. (2022). A three-year design-based study: Cultivating students' collaboration using principle-based approach in student Knowledge Building Design Studio. AERA.
- Tria, F., Loreto, V., Servedio, V. D. P., & Strogatz, S. H. (2014). The dynamics of correlated novelties. *Scientific reports*, 4(1), 5890.
- Yuan, G., Zhang, J., & Park, H. (2023). Testing the Novelty Analytic Framework for Online Collaborative Knowledge Building Discourse. In *International Conference of Learning Sciences (ICLS)*. Montreal, Canada.
- Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9-and 10-year-olds. *Educational Technology Research and Development*, 55, 117-145.

## Adaptive Instructional Support in Computer-Supported Collaborative Learning: A Meta-analysis

Lixiang Gao, Ludwig-Maximilian-University (LMU), Munich, Lixiang.Gao@campus.lmu.de  
Frank Fischer, Ludwig-Maximilian-University (LMU), Munich, frank.fischer@psy.lmu.de  
Matthias Stadler, University Hospital, LMU, Munich, Matthias.Stadler@med.uni-muenchen.de  
Olga Chernikova, Ludwig-Maximilian-University (LMU), Munich, o.chernikova@psy.lmu.de

**Abstract:** This meta-analysis investigates the effectiveness of adaptive instructional support across different learning outcomes in CSCL and the role of different learning activities used to adapt. The results indicate that adaptive instructional support in CSCL had a large positive effect on individual cognitive learning outcomes, medium effects on group cognitive and non-cognitive learning outcomes, as well as individual non-cognitive learning outcomes. When adaptive instructional support targeted individual or/and collaborative learning activities in CSCL, it benefited cognitive learning outcomes at both individual and group level, but the results were mixed for individual non-cognitive outcomes. We conclude that (1) adaptive instructional support enhances a broad spectrum of learning outcomes in CSCL; (2) the effects of adaptive instructional support on individual learners' non-cognitive learning outcomes may not be as substantial as its impact on individual cognitive learning outcomes; (3) the design of adaptive learning activities should carefully consider learners' non-cognitive outcomes.

### Problem statement

Computer-Supported Collaborative Learning (CSCL) is understood as a setting, where two or more people attempt to learn something together with the support of computers (Dillenbourg, 1999; Kirschner & Erkens, 2013). CSCL is considered a beneficial way of learning to elicit and co-construct new knowledge (Dillenbourg, 1999), improve problem solving skills (Dillenbourg, 1999), promote communication and collaboration skills (Zheng et al., 2023) and foster learning in its social nature (Jeong et al., 2019). The quality of CSCL considerably relies on pedagogical and social support by means of tools, scripts and scaffolding (Kirschner & Erkens, 2013). Adaptive instructional support, characterized by its capacity to modify the learning environment to meet the specific needs of learners, has long been of great interest and widely believed to improve the effectiveness of learning settings (Plass & Pawar, 2020). Providing adaptive instructional support to CSCL may improve the learning outcomes by balancing instructional support and addressing needs of students and possible negative aspects caused by overly structured peer interaction (Walker et al., 2009).

Although quite strong theoretical arguments assume benefits of implementing adaptive instructional support within CSCL, the practical application in collaborative learning yields mixed results. These include inconsistencies in overall learning outcomes, and specific effects on various aspects of learning and learners' experience (e.g., Chen et al., 2018; Fontaine et al., 2019; Radkowsch et al., 2021). A systematic approach is needed to identify key factors that contribute to effective learning with adaptive instructional support in CSCL environments (Rummel et al., 2016). This meta-analysis offers insights into the effects of adaptive instructional support in CSCL and examines the moderating role of different CSCL environments design features.

### Theoretical background

Scholars in teacher education, learning sciences, and educational technology are increasingly focused on designing digital tools that offer adaptive instruction tailored to the needs of individual learners or groups (Plass & Pawar, 2020). According to Plass & Pawar's (2020), adaptive instructional support means pedagogical support provided in a learning environment which could be adjusted to effectively meet learners' needs, and thus enhance learning. In the context of CSCL, this means providing the pedagogical support that can diagnose a range of learner variables, and accommodate a learner's or a group of learners' specific needs to improve learning outcomes. In contrast, fixed or static instructional support is characterized by scaffolding decisions that are predetermined and established during the design phase of the intervention (Beland et al., 2017).

The existing empirical research on the effects of adaptive instructional support in CSCL yields mixed findings. A previous meta-analysis investigated the effect of adaptive instructional support in CSCL found that it positively impacts learners' knowledge achievement and skill acquisition (Chen et al., 2018). However, results from another meta-analysis of computer-based scaffolding in STEM education shows that the effects adaptive instructional support including performance-adapted scaffolding don't perform better than fixed scaffolding and non-scaffolding (Beland et al., 2017). Another meta-analysis study conducted in health professional learning field

also revealed that the adaptive e-learning didn't outperform non-adaptive e-learning, classroom-based instruction and paper-based instruction (Fontaine et al., 2019). Furthermore, adaptive instructional support is often criticized for hindering learners' agency and thus undermining learners' motivation, as it might provide students with limited authority by forcing them to obey the decision of the system (Le & Jia, 2022). On the other hand, empirical research shows that adaptive instructional support in CSCL (e.g., adaptive collaboration scripts) may not have an adverse impact on students' intrinsic motivation (Radkowsch et al., 2021). To sum up, the findings of empirical research on the effects of adaptive instructional support are mixed and have limited generalizability. Furthermore, the rapid advancement of technology in recent years suggesting significant opportunities to personalized support, a new research synthesis would be beneficial.

The effects of CSCL can be examined across individual and group dimensions, encompassing cognitive and non-cognitive outcomes (Chen et al., 2018; Zheng et al., 2023). Therefore, this study aims at exploring these dimensions. Individual cognitive learning outcomes refer to the increase of the amount and accessibility of knowledge or accuracy of recall speed of new knowledge (Gagne, 1984). Individual non-cognitive learning outcomes refer to the learning outcomes at the individual level, excluding cognitive aspects, including coping strategies for affective/motivational fluctuations (O'Donnell et al., 1987), skills or performance of interacting with others, meta-cognitive knowledge and meta-cognitive strategies. Group cognitive learning outcomes include group performances refer to the overall performance of the group (i.e., solutions found to solve problems, product created and project designed jointly) or the quality of collaborative knowledge-building processes during collaboration (i.e., the quality of transmission of ideas) (Zheng et al., 2023). Group non-cognitive learning outcomes refer to the learning outcomes at the group level, excluding cognitive aspects, including the quality of group social activities and social relationships (Zheng et al., 2023), and the process in which group members collaboratively regulate collective activities (Zydney & Warner, 2016). Another important feature of the adaptive instructional support is what are the activities targeted by it. For example, individual learning activities (and needs of individual learner) or collaborative learning activities (e.g., sharing resources, negotiating) or both can be facilitated (Radkowsch et al., 2020).

To sum up, the current study aims to identify the effects of adaptive instructional support on learning outcomes in CSCL context and uncover the moderating role of learning outcome types and learning activities supported by adaptive instructional support.

## Research questions

This study aims to address the following three research questions.

RQ1: What is the overall effect of adaptive instructional support in CSCL on learning outcomes compared to fixed and non-scaffolding?

RQ2: What are the effects of adaptive instructional support in CSCL on different students' learning outcomes, including (1) individual cognitive learning outcomes, (2) individual non-cognitive learning outcomes, (3) group cognitive learning outcomes, (4) group non-cognitive learning outcomes?

RQ3: To what extent do the effects of adaptive instructional support in CSCL on learning outcomes differ if they prompt (a) individual learning activities, (b) collaborative learning activities, (c) both individual learning activities and collaborative learning activities?

## Methods

To address the research questions of this meta-analysis following databases were searched: Academic Search Complete, ERIC, PsychInfo, OpenDissertations. The search terms were (adapt\* OR computer\* OR cognitive tutor OR intelligent tutoring system OR personalized) AND (scaffold\* OR instruct\*) AND (skill OR competenc\* OR performance OR knowledge OR group learning) AND (CSCL OR computer supported collaborative learning OR collaborative learning OR cooperative learning) with no restriction on where the terms occur (title, abstract, descriptor, or full text). The search results were obtained on June 8, 2023. After deleting the duplicates, the search resulted in 4553 articles. To be included in the analysis, studies had to be experimental, provide sufficient statistical data, include adaptive intervention, which targeted collaborative learning (final  $N = 74$ ). Each eligible study was coded by two coders and all discrepancies were resolved through discussion. This meta-analysis used a random-effects model and Hedges'  $g$  estimation of effect sizes (Borenstein et al., 2009).

The coding scheme for this paper included learning outcomes: (1) individual cognitive learning outcomes (coded if study reported measuring knowledge on concepts, knowledge procedures, and knowledge application in individual level); (2) individual non-cognitive learning outcomes (including affective/motivational, social and meta-cognitive outcomes in individual level); (3) group cognitive learning outcomes (e.g., group task performance or group task-related interaction); (4) group non-cognitive learning outcomes (e.g., group social interaction quality or group co-regulation). In regard to the learning activities facilitated by adaptive instructional support, these

included both individual learner's activities within the collaborative learning process and group activities performed by groups of learners (coded as 'I' for individual, 'C' for collaborative, or 'Both'). An example for supporting individual activities would be providing learners with the prompts “please don’t be anxious” when the system detected the anxious emoji sent by that learner. An example for adaptive instruction supporting collaborative activities in CSCL was to prompt learners to relate their answer to the statement by other learners (Tegos & Demetriadis, 2017).

## Results

In regard to RQ1, the analysis indicates that adaptive instructional support had a significantly large positive effect on collaborative learning in CSCL. The average effect of adaptive instructional support (RQ1) on CSCL was found to be  $g = .73$ ,  $SE = .09$ ,  $p < .001$ .

Different types of learning outcomes were found to be a significant moderator (RQ2) (1) individual cognitive learning outcomes ( $g = .99$ ,  $SE = .16$ ,  $p < .001$ ,  $N = 55$ ); (2) individual non-cognitive learning outcomes ( $g = .53$ ,  $SE = .13$ ,  $p < .001$ ,  $N = 24$ ); (3) group cognitive learning outcomes ( $g = .50$ ,  $SE = .10$ ,  $p < .001$ ,  $N = 30$ ); (4) group non-cognitive learning outcomes ( $g = .53$ ,  $SE = .11$ ,  $p < .001$ ,  $N = 17$ ). In regard to RQ2, the analysis indicates that adaptive instructional support significantly enhanced individual cognitive learning outcomes with a large effect size, whereas it only moderately affected individual non-cognitive learning outcomes. Regarding group learning outcomes, both cognitive and non-cognitive aspects were affected to a medium degree, showing comparable impacts.

In regard to RQ3, while both adaptive instructional support which targeted collaborative learning activities and the one targeting individual learning activities might enhance learners' cognitive learning outcomes whether it is group or individual level. However, the effects on non-cognitive learning outcomes were mixed. When adaptive instructional support targeted individual learning activities in CSCL (RQ3), there were positive effects on both individual cognitive and non-cognitive learning outcomes, but the effect on individual cognitive was descriptively higher ( $g = .81$ ,  $SE = .15$ ) than on individual non-cognitive learning outcomes ( $g = .51$ ,  $SE = .21$ ). When adaptive instructional support targeted collaborative learning activities, the results showed a significant high positive effect on individual cognitive level learning outcomes ( $g = .98$ ,  $SE = .21$ ). However, it produced moderate effects on group cognitive learning outcomes ( $g = .52$ ,  $SE = .12$ ), group non-cognitive learning outcomes ( $g = .48$ ,  $SE = .16$ ), and individual non-cognitive learning outcomes ( $g = .59$ ,  $SE = .21$ ). When adaptive instructional support targeted both individual and collaborative learning activities, there were similar middle to high positive effects on individual and group cognitive learning outcomes, group non-cognitive learning outcomes. Yet, it had only a minimal impact on individual non-cognitive learning outcomes ( $g = .25$ ,  $SE = .17$ ).

## Discussion

This meta-analysis focused on the effects of adaptive instructional support on learning outcomes in CSCL context, taking different aspects of learning outcomes into account. This analysis reveals that adaptive instructional support in CSCL might have a large positive effect on the overall learning outcomes. Specifically, it has a large positive effect on individual cognitive learning outcomes, medium effects on group cognitive learning outcomes, group non-cognitive learning outcomes and individual non-cognitive learning outcomes. This is consistent with the findings of other related meta-analysis study (Chen et al., 2018), and confirms the claim that providing students with adaptive instructional support during their CSCL is beneficial to collaborative learning. The divergent findings between this study and the meta-analysis study in STEM education (Belland et al., 2017) may suggest that the effectiveness of adaptive instructional support may differ across learning domains and pedagogy methods. Interestingly, there are more advantages on cognitive learning outcomes than non-cognitive learning outcomes when adaptive instructional support facilitate collaborative learning activities or facilitate both individual and collaborative activities. One of the possible explanations for this phenomenon would be that the intervention periods in these empirical studies were relatively short (usually around 1 hour), while the improvement of non-cognitive factors is more likely to occur in long-term learning scenarios (e.g., Schoor et al., 2014).

A few limitations of this study should be mentioned: (1) scarce descriptions of the interventions in the primary studies, leading to insufficient data for some planned analyses. (2) the coding scheme used was rather coarse-grained to discriminate between some fine nuances in adaptive learning environments design, there was significant amount of heterogeneity, which could not be explained.

In conclusion, integrating adaptive instructional support into CSCL might bring considerable positive effects on learning outcomes. It may be advantageous to incorporate considerations of how adaptive instructional support influences students' non-cognitive learning outcomes when designing adaptive instructional support which targets individual or/and collaborative learning activities. Implementing adaptive instructional support in CSCL environments will become much easier with recent developments in AI. Given the big effect sizes of



adaptive CSCL support, educational institutions can apply the knowledge generated in this meta-analysis to substantially improve the effectiveness of their learners' learning.

## References

- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis. *Rev Educ Res*, *87*(2), 309-344.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The Role of Collaboration, Computer Use, Learning Environments, and Supporting Strategies in CSCL: A Meta-Analysis. *Review of Educational Research*, *88*(6), 799-843. <https://doi.org/10.3102/0034654318791584>
- Dillenbourg P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed), *Collaborative-learning: Cognitive and computational approaches* (pp. 1-19). Oxford: Elsevier.
- Fontaine, G., Cossette, S., Maheu-Cadotte, M. A., Mailhot, T., Deschenes, M. F., Mathieu-Dupuis, G., Cote, J., Gagnon, M. P., & Dube, V. (2019). Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. *BMJ Open*, *9*(8), e025252.
- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review*, *28*.
- Kirschner, P., & Erkens, G. (2013). Toward a Framework for CSCL Research. *Educational Psychologist*, *48*(1), 1-8. <https://doi.org/10.1080/00461520.2012.750227>
- Le, H., & Jia, J. (2022). Design and implementation of an intelligent tutoring system in the view of learner autonomy. *Interactive Technology and Smart Education*, *19*(4), 510-525.
- O'Donnell, A. M., Dansereau, D. F., Hall, R. H., & Rocklin, T. R. (1987). Cognitive, social/affective, and metacognitive outcomes of scripted cooperative learning. *Journal of educational psychology*, *79*(4), 431.
- Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275-300. <https://doi.org/10.1080/15391523.2020.1719943>
- Radkowsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International journal of computer-supported collaborative learning*, *15*(1), 5-47. <https://doi.org/10.1007/s11412-020-09316-4>
- Radkowsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, *75*. <https://doi.org/10.1016/j.learninstruc.2021.101487>
- Rummel, N., Walker, E., & Aleven, V. (2016). Different Futures of Adaptive Collaborative Learning Support. *International Journal of Artificial Intelligence in Education*, *26*(2), 784-795.
- Schoor, C., Kownatzki, S., Narciss, S., & Körndle, H. (2014). Effects of Feeding Back the Motivation of a Collaboratively Learning Group. *Electronic Journal of Research in Educational Psychology*, *12*(1), 191-210. <http://dx.doi.org/10.14204/ejrep.32.13077>
- Tegos, S., & Demetriadis, S. (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Journal of Educational Technology & Society*, *20*(1), 99-111.
- Walker, E., Rummel, N., & Koedinger, K. R. (2009). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction*, *19*(5), 387-431.
- Zheng, L., Niu, J., Long, M., & Fan, Y. (2023). An Automatic Knowledge Graph Construction Approach to Promoting Collaborative Knowledge Building, Group Performance, Social Interaction and Socially Shared Regulation in CSCL. *British Journal of Educational Technology*, *54*(3), 686-711.
- Zydney, J. M., & Warner, Z. (2016). Mobile apps for science learning: Review of research. *Computers & Education*, *94*, 1-17. <https://doi.org/10.1016/j.compedu.2015.11.001>

## Acknowledgments

The research was performed with the financial support by DFG FOR2385; FI792/12 and China Scholarship Council. We really appreciate Dr. Anika Radkowsch's suggestions and feedback during the process of revising the coding scheme of this study. Many thanks to Chenyuqi Xu, Johanna Tóth, and Ji-Young Min for their support in literature screening and coding.

# Navigating Academic Pathways with Collaborative Data: An Analysis of Enrollment Decision-Making

Nathaniel W. Cradit, Meghan Krawczyk  
ncradit@umich.edu, megkraw@umich.edu  
University of Michigan

**Abstract:** College students make course choices based on a variety of information sources within complex postsecondary systems. To support informed academic decision-making, a university developed a data platform, “Atlas,” to offer access to past course, evaluation, and faculty information. In this qualitative study using a think-aloud protocol, students built a course schedule in Atlas while explaining how they made meaning of the data in their academic, personal, social, and goal contexts. Findings included seeking trustworthy and authentic data, using course workload indicators to balance academics with holistic life responsibilities, and being empowered with faculty data to identify instructors who would best support their learning needs. We demonstrate Atlas as a collaborative system whereby students use course evaluations to support one another in navigating the academic community.

## Objectives

College and university students make decisions about course enrollments using intersecting data sources (e.g., word-of-mouth advice and anecdotes from peers, institutional data systems, subject or field guidance, career expectations) (Chen et al., 2022; Marshall, 2016; Scott & Savage, 2022). The array of potential paths through a degree program can require support to navigate, such as from academic advisors who help students make meaning of these diffuse information sources to select courses best aligned with their goals.

To support students in identifying best-fit course and faculty options, the University of Michigan (U-M) developed an online tool, Atlas, to provide transparent access to academic data. The platform is a collaborative tool to support students and others in the academic community in navigating this vast set of academic data and the information asymmetries common in postsecondary institutions. Atlas launched in an early form in 2014 and after several iterations in the decade since, is now used by more than 97% of U-M undergraduates. Atlas data include course schedule information, historical major and course profile details (e.g., typical course sequences, past grade distributions), faculty profiles, and aggregated elements of past student evaluations of teaching, including perceptions of course workload, and the respect, preparedness, and clarity of faculty. Students commonly use Atlas as a schedule-building tool prior to officially enrolling in courses, whereby they can test various schedule options and view detailed institutional data and student evaluation information about potential courses. In fact, in preparation for the fall 2023 semester, 34,299 students built 60,030 hypothetical schedules within Atlas. We view Atlas as a tool to mediate collaboration among students by aggregating elements of prior students’ evaluations and academic performance data to inform current students’ decision-making.

Perhaps a result of the uniqueness of the system in the postsecondary data landscape and the shifting informational power dynamics it fosters, faculty have raised questions about how students make sense of Atlas data. In a prior internal analysis of faculty Atlas use, professors questioned whether the platform contributed to a consumerization of higher education, and if it has unintentionally become a malign tool for students to shirk responsibility by seeking the easiest path forward. We therefore investigated how students used Atlas as a tool in their course decision-making via a think-aloud protocol to capture their cognitive processes in real time. The research question guiding this study was: how do students use an online system for academic data transparency to inform their course enrollment decisions?

## Theoretical framework

This research is rooted in expectancy-value theory (EVT), which examines achievement-related choices through the cultural, social, and psychological influences on choice and persistence (Wigfield et al., 2021). EVT suggests an individuals’ choice, persistence, and performance can be explained by their belief in how well they will perform a task and how much value they assign to completing that task (Loh, 2019; Wigfield & Eccles, 2000). EVT has previously guided other studies of course choice as it illustrates the various factors students consider when making enrollment decisions (Bong, 2001).

For this study, we foreground the expectations students have of their success in a course and the value they assign to enrolling in a course. Expectancies for success are defined as the belief an individual has about how

well they will perform a task. Subjective task value is defined as the personal value an individual assigns to a task, including four components: attainment value, intrinsic value, utility value, and cost (Wigfield et al., 2021; Wigfield & Cambria, 2010). Attainment value explains the relationship between one's performance of a task and their personal and social identities (e.g., a student with a 4.0 GPA enrolling in a course with a high median grade). Intrinsic value looks at the anticipated enjoyment the individual will get from the task. Utility value is the value of a task in relation to the individual's current goals (e.g., a student learns a skill required of them in their future career), while cost considers what one will negatively experience by performing the task (e.g., a large amount of time and effort associated with a course) (Wigfield et al., 2021). These different values illustrate the complexity of what students consider when choosing courses as they must weigh their decisions in relation to their individual situation. EVT, therefore, helped us capture the nuances students named when selecting courses with Atlas.

## Methodology

Given our research question seeking to understand how students make sense of the data within Atlas and use them to inform their course selection decision-making, we used a think-aloud protocol for this study. A type of cognitive interviewing, think-aloud research gives a defined task to a study participant and asks them to narrate their thinking, reactions, and emotions during the process of completing that task (Koro-Ljungberg et al., 2013; Wolcott & Lobczowski, 2021). Participants were interviewed via web conference and were asked to share their screen as they built a schedule for the fall 2023 semester via Atlas. Participants verbalized their thoughts as they clicked on various interface elements, their reactions to the data they found on Atlas, and the ways these data contributed to their course selections. Interview videos including captured screen shares were recorded, transcribed, and analyzed using a flexible coding approach for in-depth interview data (Deterding & Waters, 2021). This systematic analysis method is a contemporary process, inclusive of modern software tools, to make inductive and deductive passes through data and expedite reanalysis toward development of rigorous findings. Through this approach, we first applied index codes to begin building the analytic database, then subsequently applied analytic codes to capture themes relevant to this paper.

## Participants

Participants were recruited via a two-phase strategy in February and March 2023. First, given the near-universal use of Atlas among students at the institution and the enthusiasm they show for the platform, we posted a message seeking participants on the university's Reddit forum, or "subreddit." This allowed us to informally tap into ongoing student-to-student discussions of course registration and planning for the upcoming fall semester and reach participants who were primed to discuss their experiences. This strategy yielded nine participants. To expand our participant pool, we next contacted a random sample of students who had logged into Atlas since the start of the academic year, stratified by college. This let us capture the course selection experiences of students across an institution with decentralized academic policies and enrollment procedures determined by individual colleges. Twenty-five students participated in this study, including 21 undergraduate and four graduate students.

## Findings

We identified the following themes in response to our research question and EVT-informed analysis: seeking authentic and trustworthy evaluative data, strategizing course choice to balance life responsibilities, and using academic data as a tool for individual empowerment in faculty selection. In the following sections, we illustrate these findings with interview data from study participants.

### Data trustworthiness

Predominantly across the interviews students described their level of trust in the data Atlas presented. Many spoke about trusting the data, but not feeling that it was entirely authentic or other reasons their trust was nuanced (n=14). For example, students were gathering as much information about courses and professors as they could from various sources in the context of EVT cost; one being Atlas, but additionally they turned to Rate My Professor for student comments, which Atlas does not show. One described their trust in the professor evaluation data:

I think Atlas is not giving me the authentic information that I'm looking for. You're not gonna get that raw information on Atlas that I would like from other sources like Rate My Professor. People are brutal there, but it's the truth. Some measures of the professors [in Atlas] were respect and preparedness. That's not really giving me what I need in order to pick a course. I mean not to say the metrics they have on there are bad, but what I'm looking for is a little bit more nitty-gritty, a little bit more brutal, more honest.

Others described trust within their contexts. Course workload was the most referenced Atlas data point, depicted as a percentage of students on past course evaluations that said the course was more work than others. Students made meaning of this data point, an indicator of EVT cost, saying:

I trust what is reported here. And then I think a workload of 63% for an upper-level class is not the same as a workload of 90% for a 100-level class just because of who is responding.

I think it's pretty accurate, because sometimes, in the middle of the class, like when the class is in progress, I'll go back to Atlas to see what the instructor's ratings were, and normally I agree with the ratings, especially like the workload.

## Balancing life's workload

This nuanced interpretation of course workload data helped students consider their comprehensive life's workload for a prospective semester. While nearly all participants identified workload as a core decision factor, it was often related to their total course workloads, their lives outside of the classroom, and the EVT utility value they ascribed (n=12). Participants shared:

My extracurricular takes up a lot of my time. I consider that as a course in a sense because it's kind of what I've chosen to maximize out of my university education. Some people might do that in terms of classes, but I'm maximizing my university education in terms of the experiences I'm getting and that's just more valuable to me as a person.

I value my own personal time, and I also think that sometimes it's just not worth the workload. I can learn a lot of information in the class without having to do a lot of work and busy work. I also paired that with just how much I get out of the class. I'd say workload isn't the reason why I take a class, but it can be a reason why I won't take a class.

I have classes with heavy workloads. This semester I have two classes that are really time consuming, but I used Atlas to try and get my other classes that I was taking to be a really light workload or really good professors, so I didn't have to worry about them when I was doing these two really hard classes. I like to look at the workload because it will usually tell me how much homework I'll have outside of class. You use it to balance. You're not always looking for the lowest possible workload, but rather seeing how all of those fit together.

## Empowered faculty Selection

Students also used Atlas data to gain a sense of empowerment in their faculty selections (n=10). One student described having professors who also teach graduate level courses as important, while others discussed feeling like the professor can determine the experience of a course, considerations spanning EVT's intrinsic and utility value dimensions.

I do think that it's cool when Atlas shows they're teaching more upperclassmen or graduate students. That's favorable for me because I am planning on going to grad school at one point, and I think professors that have taught upperclassmen usually have a longer lasting relationship with me just because there's less turnover.

I definitely look at these ratings, the clarity, respect, and preparedness ones, because a professor can make or break your experience and really determine the trend of the class. The thing I look at most though is clarity because I feel like that's the most important one out of those three. Also respect in general is the next most important. Respect is important because I know a lot of professors in my department are blatantly sexist and racist. I want to make sure the respect percentage is high because I know there are teachers that I do want to avoid at the school.

I am very reliant on accommodations for a disability, and you can look at my transcript and find the semesters where I had professors who said, "we're not going to accommodate this request."

I've seen lots of friends who struggled in courses where professors weren't flexible, and those courses will have a higher proportion of lower grades [on Atlas], or of people dropping out or failing. If those are going to be high...that's saying to me that the professor is not good at writing exams or inflexible in the expectations of students or is not taking feedback. If you're just listening to rumors, you're not gonna get the full picture. I'd rather have [Atlas] than allow the rumors to exist.

## Discussion and implications

We investigated students' use of a collaborative online tool for academic data transparency in their course enrollment decisions, identifying data trustworthiness, balancing holistic life demands, and empowered faculty selection as core themes. We note the ways students leveraged Atlas data toward personal empowerment, sensemaking faculty ratings and course data toward inclusion and equity for their individual learning needs.

While not explicitly a part of our study, participants commonly spoke of Atlas influencing their own course and teaching evaluations while discussing the preceding themes. Students regularly described completing evaluations in context of how they understood their answers contributing to Atlas data; they subsequently made meaning of Atlas data in how they understood past students' perspectives and experiences. We therefore view Atlas as a tool enabling asynchronous collaboration across generations of students, in that they both complete course evaluations with an eye toward messages they hope to send future Atlas users and seek Atlas data in part to understand the messages past students may have been trying to give them. We anticipate future research to examine this phenomenon of student-to-student collaboration across time more comprehensively.

We demonstrate that sharing transparent academic data, including portions of teaching evaluations, allows students to benefit from each other's past experiences. As they seek to maximize goal attainment, students use Atlas data to frame expectations and craft a schedule that balances their needs. A unique tool for student empowerment through data transparency, we suggest future researchers explore how students leverage other information sources toward similar ends, including external, online faculty reviews. Given the flawed nature of student evaluations of teaching, we also see an opportunity for students to better understand the implications and utility of their evaluations as collaborative tools to support one another in navigating the academic community.

## References

- Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology*, 26(4), 553-570.
- Chen, Y., Fu, A., Lee, J. J.-L., Tomasik, I. W., & Kizilcec, R. F. (2022). Pathways: Exploring academic interests with historical course enrollment records. *Proceedings of the Ninth ACM Conference on Learning @ Scale*, 222-233. <https://doi.org/10.1145/3491140.3528270>
- Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research*, 50(2), 708-739.
- Koro-Ljungberg, M., Douglas, E. P., Therriault, D., Malcolm, Z., & McNeill, N. (2013). Reconceptualizing and decentering think-aloud methodology in qualitative research. *Qualitative Research*, 13(6), 735-753. <https://doi.org/10.1177/1468794112455040>
- Loh, E. K. Y. (2019). What we know about expectancy-value theory, and how it helps to design a sustained motivating learning environment. *System* 86, 102-119. <https://doi.org/10.1016/j.system.2019.102119>
- Marshall, J. (2016). Online course selection: Using course dashboards to inform student enrollment decisions. *Open Learning*, 31(3), 245-259. <https://doi.org/10.1080/02680513.2016.1227699>
- Scott, M., & Savage, D. A. (2022). Lemons in the university: Asymmetric information, academic shopping and subject selection. *Higher Education Research and Development*, 41(4), 1247-1261. <https://doi.org/10.1080/07294360.2021.1887094>
- Wigfield, A., and Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, 30, 1-35. <https://doi.org/10.1016/j.dr.2009.12.001>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81. <https://doi.org/10.1006/ceps.1999.1015>
- Wigfield, A., Muenks, K., & Eccles, J. S. (2021). Achievement motivation: What we know and where we are going. *Annual Review of Developmental Psychology*, 3, 87-111. <https://doi.org/10.1146/annurev-devpsych-050720-103500>
- Wolcott, M. D., & Lobczowski, N. G. (2021). Using cognitive interviews and think-aloud protocols to understand thought processes. *Currents in Pharmacy Teaching and Learning*, 13(2), 181-188. <https://doi.org/10.1016/j.cptl.2020.09.005>

# An Investigation of Dashboard in Collaborative Inquiry: The Dynamic Interplay between Technology and Pedagogy in Classroom Orchestration

Chen Feng, Indiana University, carrfeng@iu.edu  
Haesol Bae, The State University of New York, hbae4@albany.edu  
Daeun Hong, Indiana University, dh37@iu.edu  
Krista Glazewski, North Carolina State University, kdglazew@ncsu.edu  
Cindy E. Hmelo-Silver, Indiana University, chmelosi@indiana.edu  
Xiaotian Zou, Indiana University, xz107@iu.edu  
Tianshu Wang, North Carolina State University, twang43@ncsu.edu  
Thomas Brush, Indiana University, tbrush@iu.edu

Seung Lee, Bradford Mott, James Lester  
sylee@ncsu.edu, bwmott@ncsu.edu, lester@ncsu.edu  
North Carolina State University

**Abstract:** This study investigates the uses of a teacher dashboard in a technology-rich, problem-based learning (PBL) environment incorporating collaborative game-based learning for middle school science inquiry. The dashboard, designed through participatory approaches with teachers, provides real-time data visualization of students' learning activities, aiming to reduce teachers' orchestration load and enhance PBL facilitation. Findings indicate the teacher actively orchestrated classroom activities using the dashboard, balancing between digital insights and personal observation. The study extends current models of dashboard interaction by integrating classroom observations and teachers' prior knowledge into the interpretation phase. It emphasizes the need for dashboards to offer strategic-level support, enabling teachers to tailor instruction effectively. This approach not only reduces the orchestration load but also promotes meaningful student engagement and collaborative inquiry in a PBL setting.

## Introduction

Classroom orchestration describes the real-time management activities in multiple social planes (Dillenbourg et al., 2018) such as technology-rich problem-based learning (PBL) classrooms, in which teachers must manage multiple complex activities. Moreover, when the technological tools were designed and developed for collaborative learning, they might fall short in offering support for teacher orchestration with a rich set of activities (Dillenbourg, 2013). Orchestration load refers to the efforts required for teachers to conduct learning activities in the classrooms, including both physical components such as moving around the classroom, and cognitive components, such as providing facilitation to support students' learning. Introducing new technological tools into the classroom can increase teachers' orchestration load, bringing complexity and uncertainty for collaborative learning (Dillenbourg, 2013). As a result, teachers under a high orchestration load may benefit from support to better enable the facilitation of complex learning activities. To meet the need for supportive resources, we designed a teacher dashboard to 1) reduce teachers' orchestration load, and 2) promote PBL facilitation to support students' learning. The teacher dashboard in this study refers to a real-time tool for teachers, providing data visualization of students' learning activities and presenting instant recommendations through learning analytics. It was designed and developed to help teachers access learning and collaborative processes as well as make informed decisions, with the underlying goal to reduce the orchestration load in classrooms and empower teachers to provide meaningful facilitations for students (Bae et al., 2023).

Research has been exploring the design stage of teacher dashboards (Holstein et al., 2019), aiming to foreground stakeholder opinions and teachers' needs, while some also examined their practical application in classrooms. The theoretical models (e.g. van Leeuwen et al., 2021) on how teachers translate information on the dashboard into actions describe the process of awareness (knowing the information on the dashboard is available for them), interpretation (sense-making based on the provided information), and enactment (making pedagogical actions). Experimental studies have revealed that the type of support provided in the dashboard and the time pressure will influence teachers' interpretation of the situations (van Leeuwen & Rummel, 2022). However, few studies have explored how teachers used a dashboard to support their classroom orchestration and inform their pedagogical decision-making in real collaborative inquiry classrooms. This study, as a case study, aims to investigate the following question: In a technology-rich PBL classroom, how does the teacher apply the dashboard

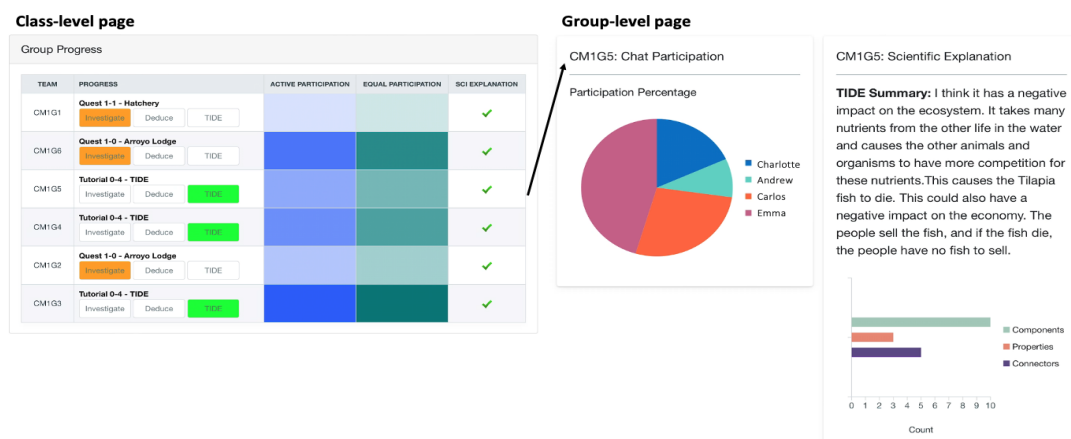
to support classroom orchestration? The finding will provide a further understanding of teacher orchestration in problem-based learning, and design implications for teacher-facing orchestration technology.

## Teacher dashboard to support collaborative game-based learning

Following the PBL approach, [GAME] was designed for middle school students to engage in collaborative scientific inquiry and promote social knowledge construction. In this setting, students work in teams of three to four to tackle a problem of ecosystems by collecting information, sharing ideas, and negotiating with peers through chat. The game consists of a tutorial and three major quests, each following the same structure of exploration and negotiation. Students need to submit their group summary after each quest.

While students are engaging in the game, their group interaction within the learning environment is captured, analyzed, and displayed for teachers through the dashboard. The dashboard consists of both class-level and group-level information (see Figure 1). On the class-level page, each row represents each group while the columns include group number, their current progress in the game, active and equal participation, and “scientific explanation” indicating the quality of the group summary. The participation was captured and calculated based on the number of in-game chats, following the rationale that the more intense the color saturation is, the less active or unequal the participation is, therefore the more attention the group might need from the teacher. To investigate a certain group, the teacher can tap the group row and the group-level information will be provided. On the group-level page, teachers can view the relative distribution of chat participation, in the form of a pie chart, and their submitted group summary. The dashboard displays the number of components (science concepts), properties (description of the phenomenon), and connectors (the logic of the argument) words used in their group summary.

**Figure 1**  
*Teacher Dashboard*



## Methods

### Participants and data collection

The case study was conducted in a science classroom middle school in the midwestern, U.S. Across five 70-minute class sessions, 22 six-grade students in six groups were participating in PBL using the [GAME], while one teacher Shelly (pseudonym), with more than 20 years of teaching experience, orchestrating the classroom with provided concurrent dashboard. The dashboard was displayed on a tablet so the teacher could carry it around the classroom and interact with it whenever needed. Shelly was provided a brief instruction on the dashboard along with the job aid explaining features. In our session, we emphasized the goal for actionable dashboard analytics to enhance, but not override, teacher decision-making and insights into individual and small group activity. We provided different displays and reasoned about ways in which low activity for one profile of student would be reasonable, but for a different profile, this would necessitate a check-in. Four types of data were collected during the implementation: 1) 350 minutes of videos on classroom orchestration captured by a camera following the teacher during the class; 2) screen recording of the tablet when the dashboard was in use; 3) five 10-15 minutes interviews recorded right after each class session; and 4) fieldnotes for each class session.

### Data analysis

To investigate how the teacher used the dashboard in the classroom, we first aligned the classroom videos with the dashboard screen recordings in each session to capture what the teacher was looking at when interacting with

the dashboard. Informed by the aligned video data, we then coded teacher actions into four categories with two levels (See Table 1). For example, when the teacher tapped into group 5 (see Figure 1), the action was coded as “Observe” and “Group”, and when the teacher made an announcement to the whole class, the action was coded as “Intervene” and “Class”. Two researchers in the research team coded and reviewed the codes collaboratively. Coding issues were resolved by discussion with changes to categories as needed.

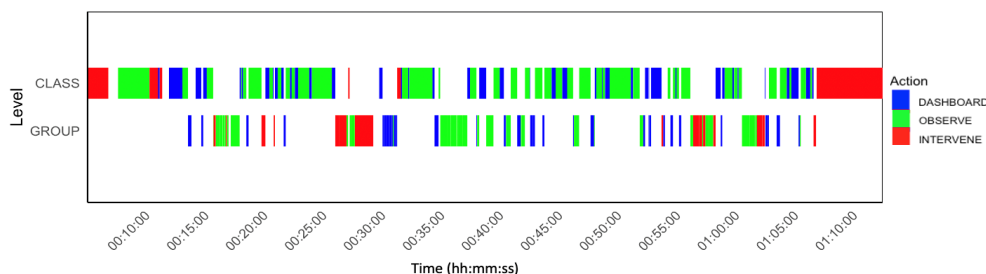
**Table 1**  
*Coding Scheme of Teacher Actions in Classrooms*

Category	Code	Description
Action	Dashboard	Interacting or browsing information on the dashboard
	Observe	Making observations on certain groups without intervening
	Intervene	Talking to individual students or groups, or making class announcements
	Other	Other activities, such as talking to researchers or leaving the room
Level	Group	Groups or individuals in groups
	Class	Class-level interaction or observation without a specific target
	Other	Usually co-occurring with Other in Action

## Findings

Our preliminary analysis showed the orchestration the teacher was engaged in with the dashboard and continuing observation in the physical environment. Due to limited space, in this paper, we will present one 70-minute class session from day 3 out of the five sessions to unpack how the teacher orchestrated the class across levels. On day 3, the teacher was familiar with the dashboard information and the students were on track in the game. In this session, the teacher interacted with students across all three social planes for 14.2 minutes for 21.1% of the whole session time, including the briefing at the start and the debriefing at the end. Shelly’s active orchestration was demonstrated by her close monitoring across the session with the dashboard (16.3 mins, 24.2%), and with observation of the classroom (34.2 mins, 50.7%). Her attention shifted across levels (see Figure 2) during the class while cross-checking the dashboard information with her observation of the classroom.

**Figure 2**  
*Action Distribution Over Time in a PBL Session*



## Enactment informed by dashboard, observation, and prior knowledge of students

In our analysis, we sought to understand teacher activity patterns across the span of the lesson by coding each action at units of one second. As displayed in Figure 2, almost every instance of dashboard access (blue) is followed by teacher observation (green). There are almost no instances in which teacher intervention (red) follows dashboard access (blue), indicating the teacher did not intervene immediately following dashboard access. For example, Shelly accessed the class-level dashboard identifying Group 3 needed attention for participation (see Figure 1), and then she walked around Group 3 to observe the situation. After confirming the dashboard information with her observation, she decided to intervene by suggesting “Hey, talk to your group more.” It suggests that the teacher did not intervene solely based on dashboard information, but instead, triangulated the dashboard information with her own observation of the classroom. This pattern aligns with Shelly’s reflection in the post-session interview: “Because you are a teacher, you are a great eavesdropper, and you can laser focus on conversations at each of the tables. I prefer that to (in-game) chat because I also like to watch their facial expressions. You can’t put that in chat, right?” While the dashboard provides behavioral and cognitive status during the collaborative inquiry, the observation in the classroom enriches Shelley’s understanding of the situation by how she observes students’ affective status. When the teacher identified a discrepancy between dashboard information and her observation for participation, she tended to interpret the situation with her prior knowledge of students. For example, Shelly described a situation in Group 1 when one student was helping another in face-



to-face conversation and it was not captured as participation in the dashboard: “I’ve got in my mind to watch when, like for example, Kale, he’s not going to say a word, he just isn’t. Colin is his guide dog. But as I’m watching where the active participation things are (not equal), I’ve paid attention to that group and I’m finding it. Yeah, it’s equal.” Thus, while the displayed participation information was designed to provide actionable information to teachers, the teacher did solely not rely on this. Shelly brought rich prior knowledge of her students into the classroom and therefore had expectations on participation patterns for different groups. Though the visualization for participation gradients could guide her attention on certain groups, she quickly made the decision not to intervene since she was “not surprised at all.”

### The group summary: “Delightfully robust”

Shelly reflected that among all the information on the dashboard, she found the part of group-level scientific explanation most meaningful to her: “I need to know where they are, but this [scientific explanation] tells me their sense of making. And their thoughts and their problem-solving. And their ability to agree.” She was able to identify the “delightfully robust” group summary, through not just the quantity or length of explanation but also the quality and accuracy as “Science is not about speed. It’s about accuracy.”

## Discussion

This study demonstrates an active orchestration practice with a teacher dashboard in a collaborative PBL game-based learning context. Based on this investigation, the teacher’s attention shifted regularly between group and class levels to monitor what was happening in the classroom. The design of the dashboard centered the teacher’s role by having teachers take pedagogical actions supported by the analytics—the teacher’s decision-making was informed not only by the dashboard but also by her observation of the classroom as well as prior knowledge of students. Meanwhile, the teacher focused on the quality and accuracy of students’ scientific explanations on the dashboard, emphasizing pedagogical attention on reasoning over mere game progression.

The findings align with the current understanding of dashboard interaction following awareness, interpretation, and enactment (van Leeuwen et al., 2021). However, we have expanded the phase of interpretation by integrating other sources of information such as classroom observation and teachers’ prior knowledge of students. This has implications for dashboard designs, particularly for any advising functions within the dashboard, providing strategy-level support such as questions to ask and concepts to address for teachers so that they can better tailor their instruction to meet the diverse needs of their students. This functionality would allow teachers to quickly identify key areas of focus, enabling them to offer more personalized and effective pedagogical support. By leveraging data-driven insights on the dashboard, teachers could provide just-in-time facilitation, fostering both collaborative inquiry and scientific discussion in groups.

## References

- Bae, H., Feng, C., Glazewski, K., Hmelo-Silver, C. E., Chen, Y., Mott, B. W., Lee, S. Y., & Lester, J. C. (2023). Co-designing a Classroom Orchestration Assistant for Game-based PBL Environments. *TechTrends*, 67(6), 918–930. <https://doi.org/10.1007/s11528-023-00903-4>
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485–492.
- Dillenbourg, P., Prieto, L. P., & Olsen, J. K. (2018). Classroom Orchestration. In *International Handbook of the Learning Sciences* (pp. 180–190): Routledge.
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher–AI Complementarity. *Journal of Learning Analytics*, 6(2), 27–52. <https://doi.org/10.18608/jla.2019.62.3>
- Molenaar, I., & Knoop-van Campen, C. A. N. (2019). How Teachers Make Dashboard Information Actionable. *IEEE Transactions on Learning Technologies*, 12(3), 347–355. <https://doi.org/10.1109/TLT.2018.2851585>
- van Leeuwen, A., Knoop-van Campen, C. A. N., Molenaar, I., & Rummel, N. (2021). How Teacher Characteristics Relate to How Teachers Use Dashboards: Results From Two Case Studies in K-12. *Journal of Learning Analytics*, 8(2), 6–21. <https://doi.org/10.18608/jla.2021.7325>
- van Leeuwen, A., & Rummel, N. (2022). The function of teacher dashboards depends on the amount of time pressure in the classroom situation: Results from teacher interviews and an experimental study. *Unterrichtswissenschaft*, 50(4), 561–588. <https://doi.org/10.1007/s42010-022-00156-9>

## Acknowledgments

This study was supported by US National Science Foundation DRL-1561486, DRL-1561655, and DRL-2112635.

# Proposing an Evaluation Framework on Educational Technology that Supports Synchronous Online Collaborative Learning

Yu Xia, Emporia State University, [yxia1@emporia.edu](mailto:yxia1@emporia.edu)  
Shulong Yan, University of California, Davis, [shlyan@ucdavis.edu](mailto:shlyan@ucdavis.edu)

**Abstract:** With online collaboration becoming increasingly common after the recent COVID-19, the design and development of online collaborative learning environments has also become essential in supporting such learning. A critical part of the development process is evaluating various technologies used to support collaborative learning. This study takes a social material lens towards synchronous online collaborative learning, builds on the definition and elements of collaborative learning, and proposes a preliminary framework to evaluate technology in terms of how it supports various aspects of the collaboration process.

## Introduction

Human activities transform the material world, including tools and human beings themselves, which consists of material and collaborative processes (Stetsenko, 2005). Research on computer-supported collaborative learning (CSCL) has long provided evidence that collaborative learning has positive effects on learning outcomes, including domain knowledge gain and collaborative skills (Stahl, 2017). With the emergency remote teaching necessitated in the recent COVID-19, synchronous online teaching and learning has increased drastically, and both instructors and students throughout K-20 struggled with engagement and interaction issues. When lecturing on video conferencing platforms such as Zoom, it became harder to facilitate active learning or collaborative learning. This study is situated in the context of synchronous online collaborative learning and aims to provide a preliminary framework to evaluate the technology tools used to support such learning contexts. This framework aims to design a socio-technological learning context, select technologies, predict potential challenges using the tech tools to support certain collaborative learning activities, and provide possible solutions and alternatives to those challenges. The following section briefly defines and categorizes features of collaborative learning in this study, describes the social material lens we adopt, and the evaluation framework for educational technology that supports online synchronous collaborative learning.

## Online synchronous collaborative learning

A most widely used definition of collaboration describes it as a construction of shared understanding through interaction with each other, in which the participants are committed to or engaged in shared goals and problem-solving process, as in “a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (Teasley & Roschelle, 1995, p. 70). As such, collaborative learning is a nested phenomenon of individual and social processes where both play critical roles (Borge & Xia, 2023). In this study, we focus on how groups interact and learn together rather than how collaboration helps individuals learn. In collaborative learning, the group, composed of individual members, co-constructs shared meanings, during which processes they negotiate sense-making as a group. Collaborative learning happens when the small group jointly problem-solves and overcomes the constraints they face in interconnected socio-technological learning systems.

Over the past decades, while computer systems have been used to support online groups and communities, most are in the forms of asynchronous communication, such as blogs, forums, wikis, etc. (Xia & Borge, 2019) or through text-based synchronous communication (Stahl, 2017). While telecommunications for synchronous, virtual face-to-face communication such as Skype and Zoom are widely adopted, they are mostly used for personal and business purposes. For research on synchronous collaborative learning process where sociocultural theories of learning were adopted and video recordings were collected (e.g., Bause et al., 2018), the analysis was focusing on utterance, without examining the interplay between artifacts and learners and the non-verbal modes of interaction, which is critical in establishing the togetherness in virtual spaces.

## Taking a socio-material lens on CSCL

Luria (1928) argued for the bi-directional effect between humans and tools, revealing the intricate relationship between social and material. Vygotsky (1978) explored this relationship from the mediation angle and argued that with the mediation of tools, humans could broaden new capacities provided by the tools through use and control. Further, the tools not only extend the physical capacities, but ultimately modify humans’ mental processes and re-construct the system. Though Vygotsky’s conceptual understanding sheds light on social and material

entanglement, scholars argue that separating social and material is futile (Johri, 2011). Instead, Sørensen (2009) argued that “we must describe in detail which entities enter into which relations that form which patterns in order to be able to characterize precisely the ability of a form of learning to connect to other entities” (p.181). This argument aligns with the definition of affordances by Norman (2013), i.e., “an affordance is a relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used” (p. 11). Both views highlight the importance of identifying active relationships that might 1) establish between multiple users, 2) between multiple users and tools, 3) form among different materials, 4) and further the interconnected network of technology, users, learning behaviors, and learning goals.

Taking such a socio-material view on CSCL, we argue the need to view technology as essential in such online learning environments, which provides the socio-technological system support and might also change how learners collaborate. Those changes brought by technologies could enable and constrain certain aspects of collaboration. For example, in Mehto et al.’s (2020) study, they found that different design tools used in the collaborative design process enabled learners to externalize ideas through visualization and diverse means of contribution but can also constrain simultaneous participation and create separate tasks for individuals. We believe different tools form an inseparable socio-technological ecological system that allows researchers to characterize different patterns.

In the recent *International Handbook of Computer-Supported Collaborative Learning* edited by Cress et al. (2021), a section devoted to an integrated theory of CSCL outlines its seven elements (pp. 36-39). Informed by this integrated theory and in response to the need to evaluate the socio-technological systems to support synchronous online collaborative learning, we propose a technology evaluation framework. Emphasizing synchronicity and sociomateriality, we specifically focus on the following elements: discourse and interaction, interactional mediation by CSCL environments, knowledge/design artifacts, shared understanding, and agency. In addition, the direct implications for technology in such a framework (see Stahl & Hakkarainen, 2021, Figure 2., p. 36) includes “design for emerging ecologies of socio-digital technologies,” developing CSCL practices, “multi-level international, artifact-related, and material mediation.”

## An evaluation framework

Synthesizing the definition and elements of collaborative learning from a socio-material lens, we propose the following framework to integrate social and material together as a unit of analysis to evaluate how they can support each aspect of collaborative learning. This framework includes four key units and criteria for evaluating technology tools - *Social-Artifacts*, *Technology-Togetherness*, *Technology-Synchronicity*, and *Multilevel Participation*. More details are presented in Table 1.

*Social-Artifacts* is a social product, as it is co-constructed by a group of learners through negotiations. It should consist of both individual and group artifacts, including their multiple versions during the process of interactive design or knowledge building, and the design elements available on the platform or created by group members. This unit corresponds to the element of knowledge artifacts including student designs. *Technology-Togetherness* in a socio-technological space means that each member is able to see others in a shared virtual space where all have access to their group artifacts and to others’ movements and actions. To achieve the sense of togetherness in a digital space, the conversation and interaction in such space need to be intuitive and natural. The shared space for group interaction also highlights the interactional mediation by the environment. Group members bring in their knowledge, expertise, and experiences; at the same time, much information is readily available online. With the activities that learners are tasked to do, they together form an interconnected learning web, reflecting how the CSCL environment mediates interactions.

Related to the sense of being together, a third important dimension of synchronous online collaborative learning is *Technology-Synchronicity*. Common communication channels should allow synchronicity in collaboration including text-based chat, audio chat, and video chat, each having its limitations. For example, when learners are busy building artifacts, they might not want to type to articulate their ideas, while video chat might require high bandwidth for smooth and stable communication. Also, the gadgets used to support video or audio chat can be unavailable on some occasions, too. These considerations are around technology but also around social interactions. The last dimension is the multiple levels in collaborative learning, including levels of individual, small group, and class community participation. To allow *multilevel participation*, the socio-technological ecology needs to allow conversational and interactional flow within and between groups.

**Table 1**  
*An Evaluation Framework on Tools that Supports Synchronous Online Collaborative Learning*

Unit	Criteria	Description	Theoretical Elements in Stahl & Hakkarainen, (2021)
<i>Social - Artifacts</i>	One single group artifact	All members will work on the same group artifact which can go through multiple iterations.	Knowledge/ design artifacts
	Various elements of the artifact	All members have access to the design elements to contribute to the group artifact.	
<i>Technology - Togetherness</i>	Visual presence of each other	All members can see other members in their group, whether in the form of video or avatar in virtual worlds.	Discourse and interaction; interactional mediation by CSCL environments; knowledge/design artifacts
	Shared space to show actions	All members are in the same shared space where all can see what others are doing.	
	Shared view on digital artifacts	All members have easy access to the same group artifact and are able to see all changes made to the artifact.	
<i>Technology - Synchronicity</i>	Text based chat	Members have access to chats open to the whole community, to other group members, or to an individual.	Discourse and interaction; interactional mediation by CSCL environments
	Audio chat	Members have access to audio chat open to the group.	
	Video chat	Members have access to video chat open to the group.	
<i>Multilevel Participation</i>	Individual participation	Individual members have easy and equal participation in the group and community learning process.	Agency; emerging ecologies; multi-level interactional, artifact-related, and material mediation
	Group interaction	All members can easily engage in group interaction.	
	Cross group interaction	All groups can easily interact with other groups.	

We would like to emphasize that, though we present the framework in a table with the goal to develop a rubric with a clear unit, we take an integrated view of socio-technological learning environments that represent holistic, synergetic ecologies. Also, it is important to identify the players in such learning environments and the roles they can potentially play in the interconnected socio-technological ecology. An evaluation effort on educational technology in synchronous online learning needs to address the orchestration of multiple tools.

## Conclusion and next steps

This paper reports a preliminary evaluation framework for examining technology tools that support online synchronous collaborative learning, focusing on socially shared artifacts, technological features that support togetherness, technological features that support synchronicity, and the sociotechnical system that support smooth multilevel participation. This work bears crucial implications in the long term as we as a society are moving into a future where synchronous online interaction is going to be more common. From this work, we see two important next steps from this preliminary framework.

First, building on the framework, we propose to develop a rubric and establish the reliability and validity of the rubric. This end product will fill the gap as no such evaluation tool is currently available to evaluate socio-technological environments for synchronous online collaborative learning. It will further help with the design and development of such learning environments, which have the potential to benefit more diverse groups of learners.

Second, as indicated in the integrated theory of CSCL (Cress et al., 2021), the other essential dimension is practice, i.e., facilitating constructive and agentic teaching and learning practices around CSCL. These practices would include specific strategies of 1) using technology to facilitate interaction, 2) documenting and developing artifacts, 3) facilitating communication and establishing shared understanding across levels, and 4)

promoting equitable and agentic participation. Such a comprehensive framework on technology evaluation would greatly encourage the intentional design of CSCL environments and ongoing reflection and examination of CSCL practices.

## References

- Bause, I. M., Brich, I. R., Wesslein, A. K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning*, 13(1), 7-33.
- Borge, M., & Xia, Y. (2023). Beyond the individual: The regulation and negotiation of socioemotional practices across a learning ecosystem. *Journal of the Learning Sciences*, 32(3), 325-375.
- Cress, U., Rosé, C., Wise, A.F., & Oshima, J. (Eds.). (2021). *International Handbook of Computer-Supported Collaborative Learning*. Springer Cham.
- Johri, A. (2011). The socio-materiality of learning practices and implications for the field of learning technology. *Research in Learning Technology*, 19(3), 207-217.
- Luria, A. R. (1928). The problem of the cultural behavior of the child. *The pedagogical seminary and journal of genetic psychology*, 35(4), 493-506.
- Mehto, V., Riikonen, S., Kangas, K., & Seitamaa-Hakkarainen, P. (2020). Sociomateriality of collaboration within a small team in secondary school maker-centered learning project. *International Journal of Child-Computer Interaction*, 26, 100209.
- Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.
- Sørensen, E. (2009). *The materiality of learning: Technology and knowledge in educational practice*. Cambridge University Press.
- Stahl, G. (2017). Group practices: A new way of viewing CSCL. *International Journal of Computer-Supported Collaborative Learning*, 12(1), 113-126.
- Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In Cress, U., Rosé, C., Wise, A.F., & Oshima, J. (Eds.), *International handbook of computer-supported collaborative learning*, 23-43.
- Stetsenko, A. (2005). Activity as object-related: Resolving the dichotomy of individual and collective planes of activity. *Mind, Culture, and Activity*, 12(1), 70-88.
- Teasley, S. D., & Roschelle, J. (1995). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie, & S. J. Derry (Eds.). *Computers as cognitive tools* (pp. 229-257). Hillsdale, NJ: Lawrence Erlbaum.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Xia, Y., & Borge, M. (2019). A systematic review of the quantification of qualitative data in proceedings of international conferences on CSCL from 2005 to 2017. In *Proceedings of the 13th International Conference on Computer-Supported Collaborative Learning* (pp. 620-623), Lyon. International Society of the Learning Sciences, Inc. [ISLS].

# Exploring the Dynamics and Trends of Knowledge Exchange: A Structured Topic Modeling Approach of the CSCL Conference Proceedings

Rogers Kaliisa, Department of Education, University of Oslo, Norway. rogers.kaliisa@iped.uio.no  
Sonsoles López-Pernas, School of Computing, University of Eastern Finland, Finland. sonsoles.lopez@uef.fi  
Kamila Misiejuk, Center for Learning & Technology, University of Bergen, Norway. Kamila.Misiejuk@uib.no  
Mohammed Saqr, School of Computing, University of Eastern Finland, Finland. mohammed.saqr@uef.fi  
Crina Damsa, Department of Education, University of Oslo, Norway. crina.damsa@iped.uio.no

**Abstract:** This paper presents the topical trends of Computer-Supported Collaborative Learning (CSCL) through a structured topic modelling of CSCL conference proceedings. The study highlights the multidisciplinary nature of CSCL, revealing theoretical, methodological, and epistemological diversity. Noteworthy findings include a decline in interest in scripting and concept maps, reflecting an evolving emphasis on learner autonomy and the study of collaboration based on various artifacts. The impact of technological advances, particularly the focus on multimodal collaboration analytics, indicates a dynamic interplay between technology and CSCL discourse. As the field stands on the precipice of the artificial intelligence (AI) era, there is anticipation that AI will significantly influence CSCL methodologies, offering opportunities for enhanced collaboration analytics and adaptive learning environments.

## Introduction and background

Computer-Supported Collaborative Learning (CSCL) is a research field at the intersection of pedagogy and digital innovation (Jeong et al., 2014). Since the early 1990s, CSCL has been established as an interdisciplinary field (Ludvigsen & Mørch, 2010), building on related scientific fields such as the learning sciences, communication studies and computer science. CSCL studies range from two individuals/learners solving a well-defined problem in a formal learning setting to large communities that emerge spontaneously and comprise thousands of users sharing their informal knowledge over longer periods of time supported by technology (Dillenbourg et al., 2009). In 1991, a significant milestone was achieved with the CSCL workshop held in Carbondale, IL, USA. This gathering marked a precursor to the annual CSCL conferences and laid the foundation for a community of scholars passionate about the transformative power of technology in collaborative learning. The inaugural CSCL conference, as we recognize it today, took place in 1995 in Bloomington, USA (Shelley et al., 2015), and since then it has served as the main platform for CSCL researchers and practitioners, to converge, exchange ideas, and showcase their contributions.

The CSCL conferences have, over the years, accumulated a wealth of knowledge and insights encapsulated in the form of peer-reviewed research papers, presentations, and discussions. These proceedings are a testament to the intellectual capital of the CSCL community. They offer an invaluable resource for scholars seeking to map the trajectory of the field, identify key trends, and comprehend the vital questions that continue to drive CSCL research forward. In this context, bibliometric analysis emerges as a powerful tool to systematically explore and map the intellectual landscape of the CSCL conferences. This study aims to provide a comprehensive picture of the CSCL ecosystem by scrutinizing the vast repository of research outputs published at CSCL to understand the vital questions that continue to drive the CSCL community and better understand its development, areas that demand attention, and potential avenues for future research and innovation.

## Related studies and justification for the present study

Previous bibliometric studies focusing on CSCL have made valuable contributions but exhibit limitations, justifying the need for our current study. Håkleiv et al, cited in Vogel et al (2019), delved into the CSCL research community, employing bibliometric analysis on ISLS and CSCL proceedings to reveal co-authorship patterns, authors' research fields and geographical distribution. However, their study focused solely on CSCL proceedings from the 2015 and 2017 editions of the conference, which is a very limited snapshot that does not reflect the field's breadth or temporal evolution. Additionally, while some extensive syntheses of CSCL research exist (e.g., Jeong et al., 2014), they are limited by thematic breadth. In contrast, our study extends existing efforts by employing latest advances in bibliometric analysis coupled with Natural Language Processing (NLP) to comprehensively analyze the content and identify the key trends and topics that have shaped the discourse in the CSCL community—a dimension less emphasized in prior studies. Through this data-driven approach, we aim to provide

a more holistic understanding of the evolving landscape in CSCL research to guide researchers on the topics to focus on and the areas that require further investigation. Our goal is to unveil the key topics and trends that have shaped the history of the CSCL community's discourse until the present day. We seek to answer the following research question: *What are the topical areas and trends that have shaped the CSCL conference over the years?*

## Methodology

The first step in our analysis was to search for all the articles published in the CSCL conference proceedings. We chose the Scopus database due to its quality metadata and coverage (Norris & Oppenheim, 2007). The search resulted in 2,697 articles. We downloaded the metadata from Scopus and used the *UDpipe* library and bibliometrix R library to process the bibliometric metadata (López-Pernas et al., 2023). Specifically, we extracted the relevant *parts of speech* (adjectives, nouns, and verbs) and lemmatised words into their most simple (i.e., dictionary) forms (e.g., reading was converted to read). To identify the *topics* of research—the aim of our study—we used structural topic modeling (STM) on the articles' textual metadata, namely titles, abstracts, and author keywords. STM is a well-known unsupervised machine learning method that allows the detection of hidden patterns (grouping of similar coherent themes together under the same cluster) (Roberts et al., 2016). STM allows researchers to identify terms that commonly occur together and form coherent themes instead of simple keyword analysis-, which is limited in depth or diversity. We used the STM implementation of the R library *stm* (Roberts et al., 2019). Since the number of topics in the data was unknown beforehand, we estimated several models ranging from 5 to 50 topics. In order to identify the optimal number of topics, we relied on a combination of fit indices (i.e., semantic coherence and exclusivity) and human judgment, by which we sought to maximize the relevance of the topic terms and minimize the overlap between topics as well as the possible ambiguity of the relevant terms. A consensus was reached by four researchers on 37 topical areas. To facilitate the analysis, we further grouped the topics into seven categories and listed their most representative terms. We also analyzed the temporal evolution of the topics, which allowed us to identify thematic trends in CSCL proceedings research. To investigate the relationship between the different topics, we constructed a network of topic co-occurrence to identify topic communities using Louvain modularity (Ghosh et al., 2018).

## Findings

### Topics of research

The analysis of topics showed that the CSCL conference has witnessed developments across various topics over the years (Table 1). First, the topics capturing the *Actors*—or groups of interest—and how CSCL studies are designed suggest that the CSCL community has maintained a strong focus on group-level analysis and the exploration of emergent, socially constructed properties of interactions (Dillenbourg et al., 2009). This focus has propelled the development of methodologies for analysing and interpreting group interactions, redirecting focus from individual cognition to understanding group-level processes. The topics within the *Design & implementation* category focus on scripted activities and interventions within the CSCL community to guide interactions through predefined structures (Weinberger & Fischer, 2006). This is despite the mixed reactions to the effectiveness of scripts, with some researchers arguing against their use due to undermining learners' agency (Fischer et al., 2013). The topics related to *Methods* show that advancements in technology have also impacted the methodological approaches used by CSCL researchers, with broad topics such as data visualisation and visual representation, appearing alongside terms such as *analytics, data, multimodal, visual, representation, and simulation*, which are rooted in the advancements in big data analytics. The topics capturing the *Theoretical frameworks* suggest that a broad spectrum of theoretical perspectives is employed by CSCL researchers, represented by terms such as regulation, epistemic, social, cognitive and knowledge building. This diversity could explain the theoretical richness of the community, with some researchers interested in the social and interaction aspects while others focusing on epistemic elements such as discourse, knowledge building, and how learners regulate their learning. Furthermore, the topics focusing on the *Discipline* or area of application show that CSCL research spans various educational domains such as mathematics, science education, and computing. This interdisciplinary approach demonstrates the applicability of collaborative learning across diverse academic disciplines and highlights the adaptability of CSCL to different disciplinary contexts. Lastly, the topics representing *Tools & technology* in CSCL showed that advancements in technology play a pivotal role in shaping the landscape of CSCL. In particular, the analysis showed that CSCL research has moved beyond the use of standard computers as the main form of technology to include technologies such as *games, mobile devices, wikis, simulations, and virtual reality* in different collaborative learning environments.

### Topic trends

The analysis showed increased attention to the *discipline of application*, such as *science* or *mathematics*, indicating a move from developing cross-disciplinary to more discipline-specific applications (see Table 1). The change in the popularity of *visual representation* before 2010, which included virtual and augmented reality, was exceeded around 2020 by *data visualization around 2020* with multimodal data as a prominent part of this topic. The centering of *actors*, such as *students* and *teachers*, spiked twice in the early years of CSCL research and around 2020. The dominance of *teachers*, especially in the second spike, could be explained by the need to support the instructors' online teaching during the COVID-19 pandemic. The product of learning studied in CSCL research has primarily revolved around *knowledge building* since 2005. This long-term emphasis underscores the enduring importance of collaborative learning in knowledge construction. Similarly, the interest in analyzing *interactions* through discourse, social interactions or argumentation remained stable over the years. Interest in *scripting* and *concept maps* seems to have declined over the years, while interest in *learning resources* and *design* has become more popular recently. While interest in *EdTech* had several spikes over the years, the interest in other *tools & technology* (e.g., *discussion tools*, *wikis*) has remained relatively stable over the years.

**Table 1**  
*Topics Identified Using STM along with the Five Most Representative Terms of Each Topic*

Category	Topic	Terms
Actors	Communities	community, network, participation, practice, youth
	Students	student, change, study, conceptual, understanding
	Teachers	teacher, professional, development, classroom, teaching
Design & implementation	Concept map	concept, construction, map, collaborative, knowledge
	Learning resources	learn, resource, learning, sciences, learner
	Peer assessment	peer, assessment, feedback, student, write,
	Pedagogical design	design, pedagogical, classroom, support, orchestration
	Design	design, base, idea, student, isl
	Collaboration task	collaborative, task, solve, individual, learn
Discipline	Scripts	collaboration, script, support, knowledge, learn
	Mathematics	computational, mathematics, mathematical, math, school
	Science education	science, student, inquiry, scientific, classroom, support
Interaction	Information exchange	share, information, object, personal, space
	Discourse	discourse, base, assess, measure, coding
	Social interactions	social, interaction, learn, network, virtual,
	Interaction analysis	analysis, interaction, network, discourse, analyze
	Argumentation	student, argumentation, argument, discussion, study
	Virtual interactions	child, virtual, museum, interaction, collaborative
Methods	Research methods	cscl, research, method, future, question
	Visual representation	representation, learn, visual, physical, study
	Data visualization	data, visualization, analytic, collect, video
Theoretical frameworks	Culture	culture, education, skill, learn, narrative
	Strategies	strategy, collaborative, regulation, learn, process
	Complex systems	system, learn, complex, learning, collaborative
	Models	model, activity, base, learn, collaborative
	Knowledge building	knowledge, building, student, discourse, collective
Tools and technology	Wikis	contribution, multi, creative, wiki, generation
	Games	game, learn, space, play, environment
	Multimedia	digital, engineering, video, media, skill
	Discussion tools	tool, discussion, awareness, communication, synchronous
	Forums	discussion, forum, student, collaborative, tutor
	Mobile learning	mobile, learn, classroom, device, student
	Online learning	online, learn, collaborative, study, collaboration
	Edtech	learn, technology, learning, educational, collaborative

## Discussion and conclusion

The purpose of the current paper was to explore the trends of knowledge sharing within the CSCL community by identifying the key topics, trends, and topic co-occurrences based on the work published at the CSCL conference. One notable observation is the diversity in topics and methods within CSCL, which signifies a rich and



multidisciplinary field. The presence of concepts such as knowledge-building, social interaction, and argumentation, and methods such as interaction analysis, discourse, and analytics implies that CSCL researchers have different methodological and epistemological stances on what counts as learning, which also influences the nature of methods and units of analysis applied by the different researchers. While some researchers see the epistemological and methodological differences within CSCL as a challenge and have called for a unified framework to reconcile these differences (Wise & Schwarz, 2017), others see the diversity as a strength that promotes multivocality and productive tensions within the CSCL field. Another notable trend in CSCL methods is the increasing focus on data visualization and visual representation, a trend consistent with the CSCL's historical interest in automated analysis to support collaboration, evident in methods such as group awareness tools (Bodemer & Dehler, 2011), adaptive scripts (Vogel et al., 2019), and other intelligent support mechanisms for collaborative groups. At the same time, the focus on data visualization could be explained by the influence of methodological and technological advancements such as learning analytics (LA), dedicated to developing tools that support collaboration in formal learning environments, a focal point of CSCL (Kaliisa et al., 2022).

The analysis further showed a noteworthy decline in interest in scripting over the years which could be attributed to a critique that excessive reliance on scripts, particularly those featuring strict guidance, has the potential to undermine learners' self-determination and intrinsic motivation (Fischer et al., 2013) and with a small positive effect on domain learning and collaboration skills (Radkowsch et al., 2020). Moreover, there was a reduced interest in studying concept maps which could be explained by the changing focus in the community to study collaborative learning around a range of artifacts (*for example, the recent work in collaboration analytics using simulations to study collaborative problem-solving in professional settings such as nursing, engineering*) other than the traditional knowledge artefacts such as concept maps. In conclusion, this study has provided a bird's eye view of the evolving landscape of CSCL research. The multidisciplinary nature of CSCL is evident, showcasing theoretical, methodological, and epistemological diversity. Notably, the influence of technological advances, particularly the focus on multimodal collaboration analytics, underscores the ongoing interplay between technology and CSCL discourse. As we stand on the brink of the AI era, there is anticipation that AI will significantly shape the future of CSCL methodologies, presenting opportunities for enhanced collaboration analytics and adaptive learning environments. We acknowledge the limitations of our study, which focused exclusively on CSCL conference proceedings. The absence of insights from other dissemination channels, such as journals, necessitates caution in generalizing our findings to the entire spectrum of CSCL research.

## References

- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, 27(3), 1043-1045.
- Dillenbourg, P., Fischer, F., & Järvelä, S. (2009). The Evolution of Research on Computer-Supported Collaborative Learning. In *Technology-Enhanced Learning* (Issue June 2014, pp. 3–19).
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in CSCL. *Educational psychologist*, 48(1)
- Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *IJCSCL*, 9, 305-334.
- Kaliisa, R., Rienties, B., Mørch, A. I., & Kluge, A. (2022). Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies. *Computers and Education Open*, 3, 100073.
- Ludvigsen, S., & Mørch, A. (2010). Computer-supported collaborative learning: Basic concepts, multiple perspectives, and emerging trends. *The international encyclopedia of education*, 5, 290-296.
- López-Pernas, S., Saqr, M., & Apiola, M. (2023). Scientometrics: A Concise Introduction and a Detailed Methodology for Mapping the Scientific Field of Computing Education Research. In *Past, Present and Future of Computing Education Research* (pp. 79–99). Springer International Publishing.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. In *Journal of Informetrics* (Vol. 1, Issue 2, pp. 161–169).
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2).
- Vogel, F., Jeong, H., Yoon, S., Håklev, S., Guillain, L. V., Abassi, N. G., & Hmelo-Silver, C. (2019). Understanding CSCL through the lens of research syntheses.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *IJCSCL*, 12, 423-467.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in CSCL. *Computers & education*, 46(1), 71-95

# Scripting Undergraduates' Interdisciplinary Collaborative Learning to Enhance Their Interdisciplinary Competence

Guo Su, Wenli Chen, Gaoxia Zhu, Chencheng Le, Lishan Zheng  
nie22.sg4560@e.ntu.edu.sg, wenli.chen@nie.edu.sg, gaoxia.zhu@nie.edu.sg, ccle@ntu.edu.sg,  
nie23.zl0281@e.ntu.edu.sg

National Institute of Education, Nanyang Technological University, Singapore

**Abstract:** Interdisciplinary competence is critical to tackle the issues that cannot be addressed by a single discipline. Existing research indicates that meaningful interdisciplinary collaborative learning can potentially develop interdisciplinary competence. This research designed and implemented macro- and micro-scripts to scaffold university students' interdisciplinary collaboration and evaluated students' interdisciplinary competence. Results found that generally, students showed great competence in integrating disciplinary knowledge, a dimension of interdisciplinary competence.

## Introduction

Society is confronted with a growing number of complex or wicked concerns and addressing these challenges necessitates knowledge, skills and methods beyond a single discipline (Schmidt, 2008). Higher education institutions have launched various initiatives to develop students' interdisciplinary competence, "a capacity to integrate knowledge and modes of thinking in two or more disciplines or established areas of expertise to produce a cognitive advancement" (Boix Mansilla et al., 2000, p. 17). Interdisciplinary competence is a multi-dimensional concept that includes students' knowledge and skills regarding integrating and synthesizing disciplinary insights, and their perceived value and beliefs on interdisciplinary work and approaches (Lattuca et al., 2013). Tripp and Shortlidge (2019) proposed a conceptual framework consisting of five criteria to achieve interdisciplinary competence. They indicated that *Collaboration* criteria would likely help students understand their own disciplinary grounding and achieve integration across disciplines (Tripp & Shortlidge, 2019). Therefore, meaningful interdisciplinary collaborative learning has the potential to promote students' interdisciplinary competence. However, meaningful interdisciplinary collaborative learning does not happen by nature. According to the literature, collaborative learning can be supported through computer-supported collaborative learning (CSCL) script. Macro-scripts are pedagogical models that model a sequence of activities to be performed by groups, while micro-scripts are dialogue models that students are expected to utilize and progressively internalize (Dillenbourg & Hong, 2008). Jigsaw is a macro-script that allows students to share the sub-topics with peers in small groups and cultivates interdependence and equal contribution (Chang & Benson, 2022). Both macro- and micro-scripts are needed to enhance students' interdisciplinary competence since interdisciplinary teams may experience difficulties in communication, resulting in conflicts and collaboration failure in practice (Repko, 2008). Epistemic and social scripts are commonly used scripts. Epistemic scripts are task strategies that help the group structure the topics to be discussed, while social scripts specify and sequence learner interactions (Weinberger et al., 2005). This study attempts to design, implement, and evaluate the epistemic and social scripts as the micro-scripts which will be embedded in the macro-script, Jigsaw script, to improve novice learners' interdisciplinary competence. The study extends existing research on script design to the field of interdisciplinary learning in which students experience conflicts and difficulties in communication and knowledge integration, and therefore, may need epistemic and social support. The research question is: how did undergraduates perform in terms of interdisciplinary knowledge and skills?

## Methodology

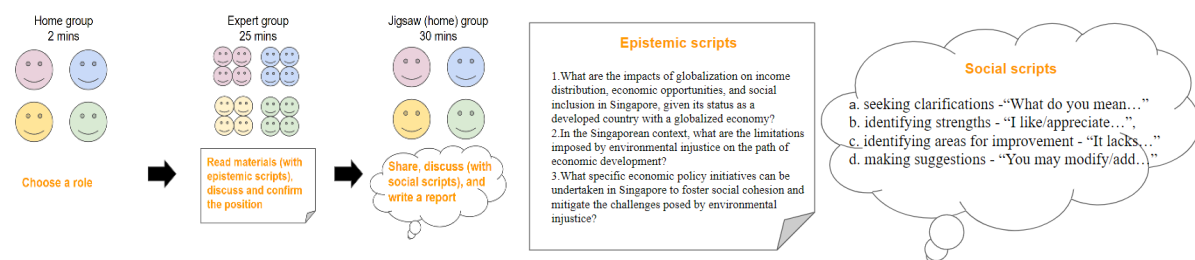
### Context and participants

A total of 27 undergraduate students (14 female) in a Singapore university, aged 18 to 26, participated in this study. Twenty-six students were second-year students, and one of them was a third-year undergraduate. The participants coming from various disciplines were randomly grouped into interdisciplinary groups of 5 or 6. This research was conducted in a course requiring students to systematically analyze sustainability challenges, especially in Singapore. Students brought their mobile devices to the tutorial lab.

### Research and pedagogical design

This study was the first iteration of a design-based research study. Design-based research is usually adopted to tackle authentic problems in real-world contexts (Collins, 1999). It is appropriate here since this study aims to address the challenges in interdisciplinary collaborative learning, such as communication barriers among team members coming from different backgrounds, knowledge integration problems, and unequal contribution as a common problem in collaborative learning (Repko, 2008). Researchers and the instructor co-developed the scripts and interdisciplinary collaborative learning activities. Figure 1 shows the pedagogical design, including Jigsaw, epistemic and social script. Co-designing with the instructor - a domain expert, the research team determined 6 roles for students to choose from anthropologist, economist, environmentalist, politician, sociologist, and social worker to promote equal contribution in interdisciplinary collaborative learning. First, each group member chose a role within 2 minutes. In the following 25 minutes, the same roles constituted the corresponding ‘expert group’ to read the assigned material accordingly and discuss the questions for developing disciplinary grounding. Prompt questions, designed as epistemic scripts, were shown at the bottom of reading materials to guide the knowledge integration. Figure 1 shows the example prompt questions for economists. Finally, the experts went back to their home groups to share and make a consensus in 30 minutes. Four categories of sentence openers for facilitating communication toward the shared goal were provided. After discussing, each group was required to write a group report focusing the two questions: “Should we focus on economic growth and development or prioritize environmental injustice when addressing social inclusion?” and “How do we reconcile the potential conflicts between economic progress and environmental injustice in pursuing social equity?” on Padlet (<https://padlet.com/>), an online platform.

**Figure 1**  
*Pedagogical Design*



## Data collection and analysis

Two types of data were collected to answer the two research questions: group artefacts and survey. First, to assess students’ interdisciplinary knowledge, 5 group artefacts written after “experts” went back to their home groups were collected and evaluated. A rubric for assessing the quality of group artefacts was adapted from Mansilla et al.’s (2009) and Tripp and Shortlidge’s (2020) rubrics (see Table 1). The score of interdisciplinary knowledge is the sum of the scores of the sub-dimensions. Two authors scored the five groups’ reports individually. The inter-rater reliability (Cronbach’s Alpha) of the four dimensions is .80, .83, .95, and .82, respectively. The disagreements were discussed. Second, participants completed a survey after the session to report their perceived interdisciplinary skills using the 5-point Likert scale developed by Lattuca et al. (2013). The internal consistency reliability (Cronbach’s  $\alpha$ ) was .92, .70 and .88 respectively in this research. Since the scale was developed to measure the interdisciplinary competence of engineering students, items were adapted to discipline-neutral items for this research. Mean values and standard deviations were calculated to describe students’ perceived interdisciplinary skills measured by the survey.

**Table 1**  
*The Rubric of Assessing Students’ Interdisciplinary Knowledge*

		1	2	3
Purposefulness	Frame the problem with an integrative approach	The paper does not contain an identifiable purpose, or the purpose is unclear.	The student clearly states a purpose that calls for an integrative approach.	The student clearly states a purpose that calls for an integrative approach and provides a clear rationale or justification for taking this interdisciplinary approach.
	Stakeholder consideration	Students considered only 1 or no stakeholder.	Students considered 2 or 3 stakeholders.	Students considered more than 3 stakeholders.

Disciplinary grounding	Range of disciplinary perspectives	Students considered only 1 or no discipline/ perspective.	Students considered 2 or 3 disciplines/ perspectives.	Students considered more than 3 disciplines/ perspectives.
	Disciplinary reasoning	Students do not explain the contributions of disciplines to the solution.	Students explain the contributions of some disciplines to the solution.	Students explain the contributions of all disciplines to the solution.
	Credibility of disciplinary knowledge	Does not include information from reliable sources (eg., peer-reviewed articles).	Includes reliable information that is irrelevant to the problem.	Includes reliable information that is relevant to the problem.
Integration	Number of disciplinary integration	No evidence of disciplinary integration.	Students raised a piece of evidence of disciplinary integration.	Students raised two or more pieces of evidence of disciplinary integration.
	Depth of disciplinary integration	Students do not explain how the disciplines selected support/ intertwine each other.	Students list disciplines/ experts' contributions without building off the knowledge/methods from each contributor.	Students leverage contributing disciplines and/or experts by building off knowledge/methods to effectively address the problem/task in a way that one contributor cannot.
Critical awareness	Social impact	Students do not include local or broader societal impacts nor what/who will be affected.	Students include only local or broader (not both) societal impacts and do not sufficiently explain what/who will be affected.	Students include local and broader societal impacts and sufficiently explain what/who will be affected.
	Limitations	Students do not include limitations or resolutions.	Students include potential limitations but do not explain resolutions to overcome these barriers.	Students include potential limitations of the plan and sufficiently explain resolutions to overcome these barriers.

## Findings

Table 2 shows the means and standard deviations of the five groups' interdisciplinary knowledge. Generally, students performed well in the dimensions of Purposefulness, Disciplinary Grounding and Advancement through Integration. Among the four dimensions, only Critical Awareness has a mean value below half of the full score, indicating that many students were not aware of the importance of being critical of their solutions.

**Table 2**  
*Students' Interdisciplinary Knowledge (N = 27)*

Purposefulness (Full mark = 6)		Disciplinary Grounding (Full mark = 9)		Advancement through Integration (Full mark = 6)		Critical Awareness (Full mark = 6)	
Mean	SD	Mean	SD	Mean	SD	Mean	SD
3.500	.612	4.800	1.037	3.500	1.000	2.400	.418

Twenty out of twenty-seven students responded to the survey. Table 3 shows that overall, students perceived that they had high-level interdisciplinary skills and had the capacity of "reflexivity", an ability to decide when to reconsider the direction of thinking and problem-solving approaches. However, the overall mean value is lower in Recognizing Disciplinary Perspective, indicating their perceived understanding of disciplinary knowledge and how to apply disciplinary knowledge in different situations.

**Table 3**  
*Students' Perceived Interdisciplinary Skills*

N = 20	Interdisciplinary Skills	Reflective Behavior	Recognizing Disciplinary Perspective
--------	--------------------------	---------------------	--------------------------------------

	Mean	SD	Mean	SD	Mean	SD
Overall	4.031	.893	4.150	.662	3.850	.899

## Discussion, limitations and conclusion

As the first iteration of design-based research, this study designed, implemented, and evaluated the integrated macro- and micro-script design to support undergraduate students' interdisciplinary collaborative learning. The study found that generally, student groups performed well in the dimensions of Purposefulness, Disciplinary Grounding and Advancement through Integration based on the evaluation of group artefacts. However, the awareness of reflection on the limitations and social impacts of their solutions needs to be improved. Furthermore, survey data shows us that students held more positive perceptions of their interdisciplinary skills and reflective behaviors than their perceptions of recognizing disciplinary perspectives. Informed by this iteration, explicit guidance should be provided in scripts to support students to reflect on the contributions and limitations of disciplines. Students' lower level of critical awareness indicates that they rarely reflect on the impacts and limitations of their interdisciplinary work critically. This finding is in keeping with earlier research indicating that it is not easy to engage in reflection and that it cannot be assumed that students have this competence unless they are trained in learning activities (Horn et al., 2022). This calls for training reflection through scaffolding in which learners receive guidance with a degree of structure that fits their competence levels and become gradually prepared for more self-directed learning (Horn et al., 2022). The study has some limitations. First, the duration of the iteration can be longer so that the script design can be applied to more topics, though the current design was implemented in a completed session. Second, more insightful results would be found if more process data, such as their dialogue and on-screen behaviors, could be collected. However, it was hard to require participants to record their discussions and behaviors since they participated voluntarily. In conclusion, this study identifies good practices and areas for improvement and provides some implications for the macro- and micro-script design of the next iteration of design-based research.

## References

- Boix Mansilla, V., Miller, W. C., & Gardner, H. (2000). On disciplinary lenses and interdisciplinary work. In S. Wineburg & P. Grossman (Eds.), *Interdisciplinary curriculum: Challenges of implementation* (pp. 17-38). New York: Teachers College Press.
- Chang, W. L., & Benson, V. (2022). Jigsaw teaching method for collaboration on cloud platforms. *Innovations in Education and Teaching International*, 59(1), 24-36.
- Collins, A. (1999). The changing infrastructure of education research. In E. Lagemann & L. Shulman (Eds.), *Issues in education research* (pp. 289-298). San Francisco: Jossey-Bass.
- Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, 3, 5-23.
- Horn, A., van der Meij, M. G., Willems, W. L., Kupper, F., & Zweekhorst, M. B. (2022). Developing interdisciplinary consciousness for sustainability: using playful frame reflection to challenge disciplinary bias. *Sustainability: Science, Practice and Policy*, 18(1), 515-530.
- Lattuca, L. R., Knight, D. B., & Bergom, I. (2013). Developing a measure of interdisciplinary competence. *International Journal of Engineering Education*, 29 (3), 726-73.
- Mansilla, V. B., Duraisingh, E. D., Wolfe, C. R., & Haynes, C. (2009). Targeted assessment rubric: An empirically grounded rubric for interdisciplinary writing. *The Journal of Higher Education*, 80(3), 334-353.
- Repko, A. F. (2008). *Interdisciplinary research*. Thousand Oaks, CA: SAGE.
- Schmidt, J. C. (2008). Towards a philosophy of interdisciplinarity: An attempt to provide a classification and clarification. *Poiesis & Praxis*, 5, 53-69.
- Tripp, B., & Shortlidge, E. E. (2019). A framework to guide undergraduate education in interdisciplinary science. *CBE—Life Sciences Education*, 18(2), es3.
- Tripp, B., & Shortlidge, E. E. (2020). From theory to practice: gathering evidence for the validity of data collected with the interdisciplinary science rubric (IDSR). *CBE—Life Sciences Education*, 19(3), ar33.
- Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33, 1-30.

## Acknowledgments

This study was funded by Nanyang Technological University (NTU), Singapore (Edex grant #023339-00001) and administered by NTU. Any opinions, conclusions or implications in this material are those of the author(s) and do not necessarily reflect the views of NTU. The study was approved by the NTU IRB (number: IRB-2023-048).

# The Influence of Interaction Patterns on Performance in Online Collaborative Problem Solving

Somi Park, Bookyung Shin, Seonyoung Keum, Youngsun Choo  
see06212@snu.ac.kr, shinbrooke@snu.ac.kr, keum0815@snu.ac.kr, daria0115@snu.ac.kr  
Seoul National University

Mi Song Kim, University of Western Ontario, mkim574@uwo.ca  
Young Hoan Cho, Seoul National University, yhcho95@snu.ac.kr

**Abstract:** Collaborative problem solving (CPS) is a key competency that plays an important role in addressing complex and ill-defined problems. To explore the essential process of CPS, this study analyzed high and low-level CPS performance groups' online interaction patterns by measuring the average time spent on each CPS action category and visualizing the connectivity between the categories through Epistemic Network Analysis (ENA). This study found that successful CPS depended on forming a shared mental model in the initial stage of the process and strategically transitioning from cognitive interactions into the composition of the final solution underpinned by content-related regulation. Socio-emotional interactions also contributed to enhancing group dynamics throughout the overall CPS process.

## Introduction

The advancement of society leads to an increase in the complexity of the problems we face. To solve ambiguous ill-structured problems that do not have definitive answers, it is necessary to critically examine solutions from multiple perspectives (Jonassen & Cho, 2011). This is the reason why collaborative problem solving (CPS) is considered a future key competency (OECD, 2017). For successful CPS, learners need to share ideas about a problem, integrate different solutions, synthesize conflicting viewpoints, and regulate the whole problem-solving process. Since not all groups navigate this process well, instructors need to provide adaptive support to improve the group performance based on knowledge of how learners interact and collaborate with each other, especially in online learning situations. However, there is a dearth of research focused on the process of CPS. Even studies that address CPS process are limited in scope, due to a confined interest in cognitive aspects, leading to a failure to deal with interactions with regulatory or socio-emotional aspects (Zhang et al., 2022). As such, the relationship between interaction patterns and CPS performance in online learning environments needs to be further explored. Our research questions are as follows: (1) How are high-level performance groups different from low-level ones in time on CPS actions? (2) How are high-level performance groups different from low-level ones in patterns of CPS actions?

## Literature review

The OECD (2017) defined CPS as the capacity to effectively share knowledge and skills with others to solve a problem. CPS not only plays an important role in learner-centered education including problem-based learning, project-based learning, and flipped learning, but also contributes to the enhancement of learner's well-being by engaging them in collaborative activities (Andrews-Tood & Forsyth, 2020). Given the significance of CPS, considerable research has been conducted on modelling generalized CPS competencies, exploring factors influencing the outcomes of CPS, and investigating effective instructional scaffolding for successful CPS.

Previous studies treated cognitive and social dimensions as key factors influencing CPS, especially emphasizing the importance of prior knowledge and socio-emotional characteristics of individual learners (C. Graesser et al., 2018). However, Barron (2003) found that the CPS outcomes of groups are not solely determined by abilities of individual members. The fact that it is difficult to expect a high-quality CPS outcome if learners fail to collaborate in an effective way highlights the importance of the process and patterns of collaboration (Zhang et al., 2022).

## Methods

Fifty undergraduate students (34 females, 16 males) participated in this study. Two learners were paired into teams; one received reading material advocating for the regulation of genetically modified (GM) food, while the other received material presenting a contrasting perspective. After five minutes of reading, they spent 25 minutes collaboratively writing an argument about the regulation of GM food with specific evidence and reasons in real-time online interaction via Zoom, using Google Documents.

We collected the written solutions of each team and recorded the whole CPS process. The scores of the group CPS solutions were evaluated by two raters (Cohen's kappa = .67) based on the holistic scoring rubric proposed by Jonassen and Cho (2011) which was modified to the context of the study. The revised rubric sets criteria to assign scores from 0 to 4 based on the presence of claims, reasons and evidence, counterarguments, and consideration of rebuttals. According to the results, performance levels were differentiated into two levels: high-level performance groups (17 teams) that received scores of 4 and 3, and low-level performance groups (8 teams) that received scores of 2 and 1 ( $M = 3.24$ ,  $SD = 1.07$ ). For the analysis of the recorded videos, two researchers segmented them into semantic units upon agreement and independently assigned codes (see Table 1) to each segment (Cohen's kappa = .92). The coding scheme was restructured by drawing inspiration from the studies conducted by Zhang et al. (2002), and Han and Cho (2019). To examine collaboration patterns, we calculated the average time high and low-level performance groups spent on each CPS action category (see Table 2). In addition, we conducted Epistemic Network Analysis (ENA). The ENA network model illustrates the connectivity between actions by creating links between nodes that correspond to each code when they manifest concurrently (Shaffer et al., 2016). We generated network models for both high and low-level performance groups, then subtracted them to compare the differences between the two (see Figure 1).

**Table 1**  
*Coding Scheme*

Categories	Codes	Explanations
Cognitive interactions	Problem definition (PD)	Specifying the meaning and scope of the task.
	Information sharing (IS)	Sharing the contents of the reading material or individual knowledge with each other.
	Negotiation (N)	Expressing one's own opinion or asking questions. Integrating opinions through agreement/disagreement.
	Solution writing (SW)	Summarizing information or writing solutions on a shared document.
Collaborative regulation	Content-related regulation (CR)	Setting directions or evaluating contents of the solution.
	Procedure-related regulation (PR)	Planning group work by setting up steps, assigning roles or managing time.
Socio-emotional interaction	Positive emotion (PE)	Complimenting or expressing positive feelings about contributions.
	Negative emotion (NE)	Expressing negative feelings about contributions.
	Empathy (E)	Socially reacting to others to form a sense of community.

## Results

### Average time for CPS actions

We first calculated how much time high-level and low-level performance groups spent on average per each CPS action category and conducted an independent sample t-test (see Table 2). In the realm of 'cognitive interactions', low-level groups allocated more time on every action category except for SW. Among them, a significant difference was observed between the two groups in the case of IS ( $p=.016$ ). Learners from high-level groups did not need much time to understand the given task, share relevant information, and propose or accept opinions about the problem. This implies that they successfully formed shared mental models, a common knowledge or comprehension shared by group members (Hsu et al., 2011), in a minimal amount of time and concentrated on the actual composition of the solutions instead. In terms of 'collaborative regulation', high-level groups engaged more in regulatory activities determining, monitoring, and revising the direction of contents rather than exploring and adjusting procedural elements such as setting up steps of the work, assigning roles, or managing time. Particularly a significant difference for CR was evident between the high and low-level groups ( $p=.041$ ). No significant differences were observed in the case of 'socio-emotional interaction' because of the infrequent occurrences of those activities throughout the whole CPS process. However, high-level groups tended to exhibit more PE and E, whereas low-level groups displayed a greater extent of NE.

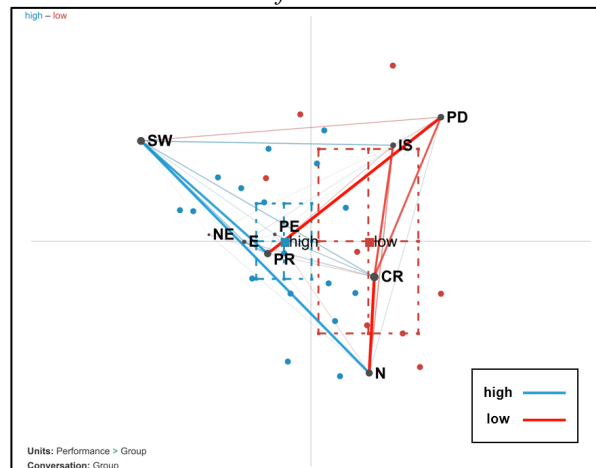
**Table 2**  
*Average Time(minutes) of CPS Action Categories*

Categories	Codes	High performance M (SD)	Low performance M (SD)	<i>p</i>
Cognitive interactions	Problem definition (PD)	0.54 (0.68)	1.10 (0.87)	.093
	Information sharing (IS)	3.33 (2.17)	5.86 (2.47)	.016*
	Negotiation (N)	3.91 (2.86)	5.23 (2.46)	.273
	Solution writing (SW)	13.94 (4.56)	9.96 (4.73)	.056
Collaborative regulation	Content-related regulation (CR)	12.30 (5.01)	8.08 (3.26)	.041*
	Procedure-related regulation (PR)	1.77 (0.86)	2.27 (0.68)	.159
Socio-emotional interaction	Positive emotion (PE)	0.11 (0.15)	0.09 (0.17)	.766
	Negative emotion (NE)	0.06 (0.13)	0.17 (0.45)	.498
	Empathy (E)	0.65 (1.44)	0.28 (0.20)	.477

### Connectivity between CPS actions

We measured the degree of association among CPS action categories through ENA and conducted a Mann-Whitney U test to identify the differences in pattern between the high and low-level groups (see Figure 1). The result reveals a statistically significant difference in pattern between the two networks along the X axis ( $U=19.00$ ,  $p<.01$ ,  $r=.72$ ). The fact that ‘PD’ is positioned on the far right whereas ‘SW’ is on the very left on the graph indicates that the X axis represents whether the focus of the activities during the CPS process is on exploration or comprehension of the given problem, or on execution for the production of an actual solution. The connection between SW and other categories is prominent in high-level groups, which means that high performing approached the task in a profoundly strategic manner. In addition, E is strongly linked with PR and SW, showing that learners from high-level groups encouraged each other by providing supportive emotional backing for the task execution steps or contents proposed by the other group member. Low-level groups demonstrated a pronounced connection between PD-PR compared to high-level groups. This can be attributed to an insufficient understanding of the requirements of the presented problem, leading to prolonged discussions on the issue and manifesting in regulatory activities such as time monitoring.

**Figure 1**  
*Subtracted ENA Model of Overall CPS Process*



For further analysis, we divided the entire process into early, middle, and late stages. At the early stage, high-level groups demonstrated stronger connections between IS-SW, and N-SW. This implies that they established a foundation for their work in the initial phase of the CPS process by sharing and summarizing the different information and opinions of each member. In the middle and late stages, the connection between CR-SW remained consistently strong for the high-level groups, whereas low-level groups demonstrated a relatively higher connectivity between N-CR. High-level groups were adept at setting the direction of the contents. Based on this groundwork they wrote or critically revised the solutions, and thus achieved high-quality outcomes. Low-level groups, on the other hand, exhibited concurrent N-CR activities due to difficulties in reaching a consensus



on the argument. Therefore, the collaborative debate exhibited during the early and middle stages did not lead to SW, which in turn tended to take place in a hurry at the very last stage of the CPS process.

## Discussion

In summary, successful CPS requires establishing a shared mental model of the problem early in collaboration. Our results showed that groups with high performance were particularly efficient in executing the CPS process, sharing information, and understanding of the problem mostly occurring during the initial phases of the task and quickly moving on to solution writing. This showed strategic regulation based on shared understanding of the task. Directly connecting content-related regulation to solution writing, thus meeting the established task requirements, seemed particularly effective, though such regulation was also essential throughout the CPS process. Finally, fostering team activities through appropriate socio-emotional interactions, such as expressing empathy, were also shown to be vital.

Our research findings can provide several educational implications. For instance, our results imply that before collaboratively solving a problem, individuals should be guided to sufficiently share their personal understanding and background knowledge about the problem. This can help students form a shared mental model. These results are in line with previous research that emphasized the importance of shared mental models and co-regulation (Van den Bossche et al., 2011). Furthermore, scaffolding is necessary to facilitate group regulation related to the task content. This is also supported by the previous literature, which emphasizes that the nature of the task is important to collaborative patterns (Van Boxtel et al., 2000) and that adjusting regulation strategies to the necessities of the current status of the group is necessary (Malmberg et al., 2015). However, our findings add that content-related regulation needs to be particularly emphasized in the case of a writing-based task. Last but not least, instructors should provide various scaffolding to foster empathetic interaction between group members. Socio-emotional interaction is a characteristic of effective collaboration, which was highlighted by both our results and previous findings (Kwon et al., 2014). We expect that the findings will provide implications for the development of instructional methods and software that help improving learners' CPS processes and outcomes.

## References

- Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior, 104*, 105759.
- Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307-359.
- C. Graesser, A., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. L. (2018). Challenges of assessing collaborative problem solving. In Care, E., Griffin, P., & Wilson, M. (Eds.), *Assessment and teaching of 21st century skills* (pp. 75-91). Springer.
- Han, Y. J., & Cho, Y. H. (2019). Exploring online collaborative learning process with a visualization tool for a design task. *Journal of Educational Technology, 35*(3), 787-816.
- Hsu, J. S., Chang, J. Y., Klein, G., & Jiang, J. J. (2011). Exploring the impact of team mental models on information utilization and project performance in system development. *International Journal of Project Management, 29*(1), 1-12.
- Jonassen, D. H., & Cho, Y. H. (2011). Fostering argumentation while solving engineering ethics problems. *Journal of Engineering Education, 100*(4), 680-702.
- Kwon, K., Liu, Y. H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education, 78*, 185-200.
- Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high-and low-performing groups. *Computers in Human Behavior, 52*, 562-572.
- OECD (2017), *PISA 2015 results (Volume V): Collaborative problem solving*, OECD Publishing.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9-45.
- Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and instruction, 10*(4), 311-330.
- Van den Bossche, P., Gijssels, W., Segers, M., Woltjer, G., & Kirschner, P. (2011). Team learning: Building shared mental models. *Instructional science, 39*, 283-301.
- Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education, 183*, 104485.

## “We did it we did it!": Exploring Middle Schoolers' Collaborative Learning in an Online Machine Learning Curriculum

Changzhao Wang, University of Southern California, wangchan@usc.edu  
Jennifer Kahn, University of Miami, jkahnthorne@miami.edu

**Abstract:** To address the pressing need of AI and Machine Learning (ML) education at K–12 level, we designed and implemented an online curriculum that engaged middle school students in ML projects. Ten students from Grade 6–8 participated in a 16-hour online weekend program in Spring 2023. With ML computer tools, students collected data, trained and tested their own ML models, and shared and discussed ML products with each other. This study investigates how the designed constructionist, CSCL environment supported students' learning in ML practices. We used interaction analysis methods to analyze video recording episodes of students' collaborative learning in k-means clustering projects. Our findings show that students developed understandings of the clustering mechanism and labeling practices through iterations of collaborative observing, hypothesis making, investigating, and problem-solving. The study sheds light on the design of CSCL environment for ML education.

### Introduction

As Artificial Intelligence (AI) is being increasingly utilized everywhere, it is necessary to prepare youth for the future AI-driven society. Young students need opportunities to develop a deep understanding of AI and especially Machine Learning (ML)—computational algorithms that function as the “brain” of AI to support its intelligence (El Naqa & Murphy, 2015). Curricula development and research on learning environments for AI/ML are needed for K–12 populations. In particular, there is a need to advance curricula that facilitate and leverage learner communities and collaboration for learning with and about AI/ML. To fill in this gap, the first author designed and implemented a synchronous online AI/ML curriculum that engaged middle school students in ML projects, from collecting data to training and testing their own ML models. The design drew on both constructionist and CSCL principles to facilitate social, collaborative learning. In this study, we focus on how our designed learning environment facilitated students' learning of ML knowledge and practices. We ask, how does a constructionist, CSCL environment support middle school students to perform k-means clustering and understand the results?

### CSCL environments and constructionism in AI/ML education

The learning design for our CSCL environment is informed by constructionism, which suggests that students learn through constructing sharable and personally meaningful artifacts (Kafai, 2006; Papert, 1980), and the perspective that students learn AI/ML through participation in socio-technical practices (Lave, 2011). Among the limited number of published research on K–12 AI/ML education (e.g., Gresse von Wangenheim et al., 2021; Touretzky et al., 2019; Zhang et al., 2022), there have been some CSCL environments studied, in the form of group projects to create ML models with computer-supported AI/ML tools. Mariescu-Istodor and Jormanainen (2019) implemented a ML lesson for high schoolers to work in groups on image recognition tasks with designed tool. Wan et al. (2020) reported their use of the SmileyCluster tool for k-means clustering with high schoolers in pairs.

Learned from these successful explorations, we focused on image classification and k-means clustering as major ML contents and designed the ML projects for students to create ML products to solve real-world problems with AI/ML computer tools. The CSCL environment is also designed to encourage students to share and discuss their ML products with peers and teachers. In the process of working on the designed ML projects, students' social interactions with AI/ML computer tools, physical tools, problem contexts, and peers and instructors, provide them the opportunity to connect to existing knowledge and build up new knowledge.

### Methods

The current study comes from the second iteration of a larger design-based research project called Artificial Intelligence for Middle School Students (AIMS). The first iteration was a 5-day pilot online program in Summer 2022, and the second iteration was a 4-week online weekend program in Spring 2023.

### AIMS curriculum design

The 16-hour curriculum is designed for middle school students to learn about AI and particularly how ML works. It consists of three modules: the first module on supervised learning (6 hours), the second on unsupervised learning

(6 hours), and the third on reinforcement learning and AI ethics (4 hours). Each module includes short introductory lectures, projects, and discussions. In all project sessions, students first work individually to build their own ML projects, then share and discuss their ML products with peers and instructors.

Projects and discussions are key components to form the constructionist learning environment. The projects used Teachable Machine (TM; <https://teachablemachine.withgoogle.com/>) for supervised learning and Orange data mining tool (<https://orangedatamining.com/>) for unsupervised learning. With Orange, students made k-means clustering models by themselves, using embedded datasets or importing their collected data to Orange. Due to page limit, this paper specifically focuses on students' collaborative learning when doing k-means clustering with Orange. More information about the curriculum can be found in Wang (2023).

## Implementation and participants

Ten students (Grade 6–8) participated in the 2023 spring program, with eight attending at least 12 hours out of the total 16 hours' duration. Students were diverse in terms of their gender (three female, seven male), race, and ethnicity (two Latino/Hispanic, one Indian, seven Chinese), and physical location (five different US states and New Zealand). This free online program was completed over three (Saturday cohort) or four (Sunday cohort) weekends, from 10:00 am to 2:30 pm one day each week. Each cohort included 5 students. For Module 2, some students on Saturday did not install Orange properly, so the current study focuses on the Sunday cohort. The first author was the instructor throughout this program, occasionally assisted by a graduate student in STEM Education.

## Data collection and analysis

Data sources include Zoom recordings of all class sessions throughout the program, pre and posttests and surveys, and other sources such as researchers' observation notes and students' learning journals. This analysis focuses on the Zoom recordings of class sessions to examine how students learned with each other. We used interaction analysis methods (Jordan & Henderson, 1995) to micro-analyze episodes of students' collaborative learning (Barron et al., 2013). Our analysis focused on the following: During the class sessions of hands-on ML projects, (1) how students interact with each other, as well as the ML tool/platform and physical tools; (2) the evidence of learning ML knowledge and practices; and (3) how those interactions may have supported their learning. The analysis went through the following major phases: (1) Content logs; (2) Transcription; (3) Group view and discussion. The observations focused on students' social interactions in the designed learning environment. (4) Detailed interaction analysis tables. We added notes of physical action (e.g., gesture, facial expression), action or reaction on the computer, and analytical notes on interactions between different elements in the designed learning environment. More details in the data source and analysis can be found in Wang (2023).

## Findings

We briefly present the interaction analysis of two scenarios in which students had collaborative learning experiences through iterations of observing, making hypothesis for their observation, and investigating, and finally reaching a conclusion. Both scenarios occurred when students were sharing k-means clustering results: Scenario 1 was for clustering a traffic signs dataset, and Scenario 2 was for clustering a fruit dataset.

## Collaborative exploration of clustering mechanism

When one student shared his/her screen, all peer students carefully observed and interpreted the clustering result, then identified problems from the observation and investigated the problem together. We see evidence of students' collaborative observations and interpretations for the distance between dots on the scatter plot of clustering result in Scenario 1. The teacher first demonstrated an Orange project of using its embedded dataset Traffic Signs to do k-means clustering ( $k=4$ , i.e., resulting in four clusters). Then students individually built the same project. They needed to change the value of  $k$ , which would determine how many clusters are in the result, to explore clustering data into different number of groups. Each student shared their own project with the class.

Darsh (Male; Indian American; Texas; note that all student names are pseudonyms) shared his clustering result of three groups, and the teacher guided students to observe and elaborate on the clustering result. Three out of the four students in attendance verbally shared their observations and interpretations (Carlos [Male; Hispanic; Florida] was in a situation not convenient to speak). Lily (Female; Chinese American; Florida) observed in Darsh's scatter plot that there was a dot distant from any of the cluster centers, which led to students' exploration of the phenomenon together as Darsh shared his screen. A selection of the discussion is provided in Table 1. This episode involved rich interactions between the learner and the class (peers and teacher), as well as interactions between the learner and Orange (ML tool). Lily's interesting observation led to in-depth discussions among students and collaborative explorations of why certain dots were clustered together or not. As a result, Darsh could correctly and confidently predict and interpret which traffic signs the distant dots represent and vice versa.

**Table 1**

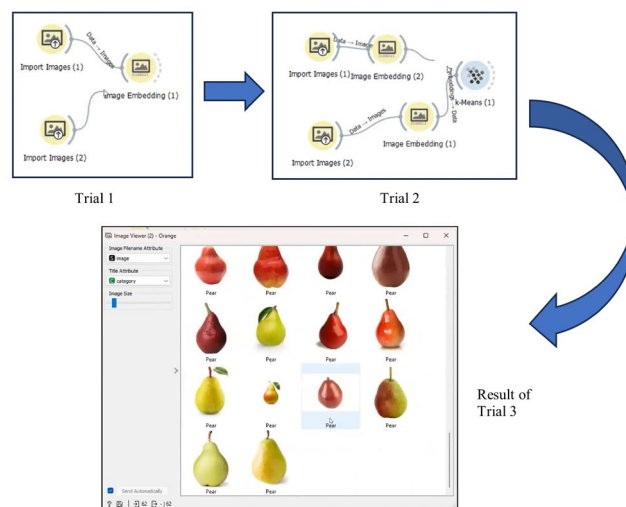
*Exploration of Distant Dots as A Result of Clustering Traffic Signs*

Turn	(Action/Reaction on Computer) Talk
20	Lily: Wait! There, there's a tiny blue dot! Like, in the middle of nowhere.
21	(Darsh found that dot and selected it.) Darsh: Oh, right, here!
22	(Darsh selected the entire blue group.) Teacher: Oh, I think. so, what do you think that tiny blue dot means?
23	Lily: It's something random.
24	Darsh: It's separated from the rest [interrupted]
25	Lucas: It's like, just like, something like, apart, more distancing out.
26	Lily: An outlier.
28	(Darsh selected the blue group, and showed the signs in that group. The teaching assistant also circled out the red sign so that everyone can see.) Darsh: It's actually in the blue cluster, but it's a red sign.
29	Teacher: so. Let's let's. We can make sure whether this is the the dot out there. So why not you only selecting that dot? You only select this dot.
35	(Darsh switched to the image viewer. The red warning sign showed up. It was the sign which we guessed to be the outlier. While the teacher was summarizing, Lily marked [via the annotation feature of Zoom] on Darsh's screen to point out an error.) Teacher: Yeah, and then use image viewer. Let's see. Oh, yeah, yeah! that's a good catch! So that's kind of that's kind of an exception among that group, so all of the other dots are in the blue color, and this one is red. Good, so that also matches what we have just interpreted: sometimes, so they may figure out some patterns or rules that doesn't make that much sense to human.
41	(Darsh selected the dot, and it did show up as the stop sign.) Darsh: The stop sign. So they're like, one exception for all of them. So, it's like, the ones that are similar, they are close to each other? But the ones that are different, like, if I click on that one, that's probably the stop sign. Yeah! The one that's far away from them is the stop sign. It's like: The ones that are similar, they are close to each other; the ones that has some trait, but it's not exactly similar, it's farther.

**Collaborative problem solving to label image data**

After observing each other's clustering results, students may recognize shared problems, then solve the problem collaboratively. We see evidence of this in Scenario 2. The teacher had students import dataset of apple and pear images into Orange and use that dataset to do k-means clustering. While sharing to class, three students, including Jack (M; Latino; Massachusetts), Darsh, and Lily, all observed similar phenomenon in their results: There were always two pear images mixed in a cluster of apple images. Jack commented that the result was "inaccurate," and brought up his "hypothesis": "What if we rename the apples to a different thing, so that it would be easier for the AI to determine which is a pear which is an apple." Then, Jack led the class to explore how to integrate fruit names to help the computer improve the clustering and had three collaborative trials (Trial 1–3 in Figure 1).

**Figure 1**  
*Screenshots of Jack's Explorations*



Through the three collaborative trials, students figured out how to add category labels to image data on Orange (although later they found out the labeling did not change their clustering results). All students were highly engaged in the exploration and provided in-time feedback to Jack. For example, students showed excitement when they agreed with Jack's plan; when Trial 1 and 2 failed, they actively provided suggestions. Finally in Trial 3, Jack followed the plan suggested by Carlos to add only one Import Images widget but put apple and pear images into separate subfolders under the image folder. He observed the Apple and Pear label showing up and shared this success with the class "Oh! We did it, we did it! Look!" All students were excited and celebrated for this achievement by smiling, waving body, and applauding, etc.

## Discussion and conclusion

In the collaborative learning scenarios, students carefully observed each other's ML products, identified problems, made hypotheses of the problems, and investigated their hypotheses together. Through iterations of the exploring process, they could finally reach a conclusion to the identified problems. Our Findings section present such examples of (1) how students understood that the distance between dots in clustering results is determined by the similarity between images, and (2) how students figured out the way to add category labels to images in Orange.

In conclusion, the designed curriculum supports students' learning in ML practices through collaborative observation, interpretation, exploration, and problem solving. Through iterations of hypothesizing and investigating, students can deepen their understanding of ML mechanism behind the problem and enhance their ML practices. Students are closely connected to support each other as a learning community in this designed constructionist, CSCL environment. We believe that our study makes a timely and significant contribution to the emerging field of K – 12 AI education, particularly on the design for online collaborative learning.

## References

- Barron, B. J., Pea, R., & Engle, R. A. (2013). Advancing understanding of collaborative learning with data derived from video records. In C.E. Hmelo-Silver, C.A. Chinn, C.K.K. Chan, A. O'Donnell (Eds.) *The international handbook of collaborative learning* (pp. 203-219). Routledge: New York.
- El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? In I. El Naqa, R. Li, & M. J. Murphy (Eds.), *Machine Learning in Radiation Oncology: Theory and Applications* (pp. 3-11). Cham: Springer International Publishing.
- Gresse von Wangenheim, C., Hauck, J. C. R., Pacheco, F. S., & Bertoneceli Bueno, M. F. (2021). Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. *Education and Information Technologies*, 26(5), 5733-5778.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, 4(1), 39-103.
- Kafai, Y. B. (2006). Constructionism. In *The Cambridge Handbook of the Learning Sciences* (pp. 35-46). Cambridge University Press.
- Lave, J. (2011). *Apprenticeship in critical ethnographic practice*. University of Chicago Press.
- Mariescu-Istodor, R., & Jormanainen, I. (2019). *Machine learning for high school students* Proceedings of the 19th Koli Calling International Conference on Computing Education Research, Koli, Finland.
- Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. New Yor: Brighton, Sussex: Basic Books; Harvester Press.
- Touretzky, D., Martin, F., Seehorn, D., Breazeal, C., & Posner, T. (2019). Special Session: AI for K-12 Guidelines Initiative. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 492-493), Minneapolis, MN, USA.
- Wan, X., Zhou, X., Ye, Z., Mortensen, C. K., & Bai, Z. (2020). *SmileyCluster: supporting accessible machine learning in K-12 scientific discovery*. Paper presented at the Interaction Design and Children Conference, London, United Kingdom.
- Wang, C. (2023). *AI education for middle school students: Critically exploring socioscientific issues with machine learning*. [Doctoral Dissertation, University of Miami].
- Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education*.

## Acknowledgments

We would like to thank all enrolled students for participating in the research, Tianyu Ma for assisting in the AIMS 2023 Spring Program as teaching assistant, and Hua Ran and Tianyu Ma for participating in the video analysis.

# Co-Designing an AI Tool to Support Discourse Based Math Instruction in a High-Dosage Tutoring Context

Chelsea Brown, Jeffrey Bush  
chelsea.brown-1@colorado.edu, jeffrey.bush@colorado.edu  
University of Colorado Boulder

**Abstract:** TalkMoves is an AI assistive tool that provides automated feedback to educators to support their daily teaching practices. While originally designed for classroom math teachers, this tool can be useful in a broader context. The University of Colorado Boulder and Saga Education formed a co-design team tasked with re-contextualizing TalkMoves for coaches of novice math tutors to use in their ongoing professional development. To effectively adapt an existing technology to a new problem space, the co-design team iteratively exchanged ideas of what exactly TalkMoves could achieve, as well as the specific needs of the coaches. Facilitators used strategies such as communal orientation, expansive dreaming, backcasting, and revoicing to promote productive collaboration. Three main goals emerged: maximize opportunities for user agency, center design around goal setting, and integrate the tool into the existing workflow. Any adaptation of an AI tool would benefit from this approach.

## Introduction and background

AI assistive tools have the potential to make a dramatic impact in education, especially in areas that are chronically under-resourced. Many students have been negatively impacted due COVID-19 and issues of inequity have been compounded; one proven way to address learning loss is through incorporating frequent tutoring (Dorn et al., 2020). Saga Education (Saga) has developed a unique and scalable approach to this challenge by recruiting novice tutors and pairing them with experienced coaches who provide ongoing professional learning (PL).

Coach feedback is critical in ensuring that tutors learn high leverage and equitable pedagogy, however, finding time to observe multiple tutors and sessions is both challenging and tedious. TalkMoves is an AI feedback tool that has shown success helping classroom teachers in their daily practice (Jacobs et al., 2022), and could potentially assist coaches by augmenting their ability to provide high quality and evidence-based feedback to tutors about their discourse based tutoring strategies. The current study focuses on the unique integration of TalkMoves into the existing workflow of Saga coaches by exploring three research questions. *RQ1*: What themes arise that explain how the co-design team of practitioners and researchers aligned their goals and priorities to re-contextualize the design of TalkMoves for a high-dosage tutoring context? *RQ2*: What generalizable facilitation strategies were used to enhance productive collaboration in the design process? *RQ3*: What design themes and goals emerged within and across co-design sessions that created synergy between the existing TalkMoves application and the existing Saga workflow?

To address these questions, we drew on many theories. *Co-design*, an established method drawing on sociocultural theory, incorporates expertise from all involved parties and intentionally structures projects to address learning within the context of motivation, identity, belonging, intertwined with cognitive development (Penuel et al., 2007). Collaborative efforts enhance applicability, increase adoption rates, and foster agency through successful partnerships. To adapt an existing tool to growing needs, we borrow the concept of *solutions mapping* from the field of engineering (Lee et al., 2020). Starting with an existing technology (TalkMoves) and finding a novel application (remote tutoring), the technology and the problem continually inform each other resulting in novel adaptations. TalkMoves was first a browser-based tool that took recordings of classroom lessons and used computational modeling to process the dialogue for talk moves (utterance level categorizations of teacher and student speech surrounding math content) and discourse features (e.g. talk ratio: time teacher talks vs. students, wait time: how long until a response, and quantity of math terms used). Talk moves are based on *accountable talk theory* which states that deliberate and intentional forms of communication can lead to more equitable participation in academic conversations (Michaels et al, 2008). A pilot study with classroom math teachers found that highlighting when and where talk moves were present (or not) during a lesson was beneficial and informative for the teachers (Jacobs et al., 2022; Suresh, in press).

## Methodology

The co-design team, made up of CU and Saga researchers as well as coaches and coach supervisors, met eight times remotely over four months to discuss the design of the interface for coaches and supervisors, which produced

a set of wireframes (interactive depictions) that captured many of the discussion points. Additional sources include collaborative digital whiteboard spaces, video memos for reflection, and one-on-one interviews of Saga coaches and supervisors. We used multiple strategies to assess the data to ensure triangulation of the emergent features and concepts (Saldaña, 2016). Analytic memos were written for each meeting, capturing the main points and trends. Videos and whiteboards were reviewed to collect all comments that pertained to what participants wanted to incorporate into TalkMoves. Comments were then iteratively reviewed and open coded (Corbin & Strauss, 2015) for similarity and initial groupings. Once consistent categories emerged, comments were inductively coded (Saldaña, 2016). Comparing codes to initial trends in the analytic memos allowed for a comprehensive understanding of the overall themes being captured from a fine grained to broader analysis. A micro-vignette (Barter & Renold, 1999) was created to showcase multiple themes, as well as facilitator strategies. Reviewing these sources of data establishes how researchers and Saga members reached a mutual understanding of how TalkMoves could be applied in a new context and provides insight into the specific features and facilitator moves that brought the main suggestions out.

## Findings

*RQ1: Creating a shared vision to re-contextualize TalkMoves:* Through co-adaptation and reflection about TalkMoves, we see how CU and Saga align priorities and create a shared vision of how the new platform could be mutually beneficial. We know from solution mapping that integrating a solution into a new problem space is an iterative exchange. Our analysis indicates four themes where researchers and Saga converge to align their goals and priorities to re-contextualize the design of TalkMoves in a high dosage tutoring context. First, the co-design team had to identify the user for whom TalkMoves would assist the most. Next, the CU researchers had to comprehend the specific context in which TalkMoves would be used. Simultaneously, Saga coaches had to learn what TalkMoves provided for classroom teachers and reconcile how it would apply in a tutoring setting. Finally, by incorporating the new design of TalkMoves, Saga coaches needed to imagine how their practices would change.

Central to creating a shared vision for the platform was identifying who the platform was being built for and why. First and foremost, CU and Saga researchers had to identify that the Saga coaches were best situated to use the feedback from TalkMoves, instead of tutors as originally thought. Tutors had neither time for co-design or PL about interpreting results, but this tool could be extremely helpful to coaches. Therefore, the next stage focused on understanding both the normal workflow of the coaches as well as the challenges TalkMoves could assist with. Each new layer of understanding informed how the application could best be adapted. The full picture of the coaches workflow was refined across every co-design session. If the co-design team had only met once or twice researchers would not know valuable details such as the challenge coaches faced with taking notes, that coaches were often the only adult in the physical classroom, or that supervisors did not have a standardized way of tracking whether a goal was achieved or not. All these details emerged in later sessions, and all informed the new application. Through trying to understand coaches' process and understanding what AI can achieve, new areas for the app to support were uncovered. Concurrently, the team focused on the tension between different contexts; what is needed to support tutoring versus classroom instruction. Coaches and classroom teachers both need information about what is occurring within a lesson; however, this looks very different across users and settings. In the original form of TalkMoves, teachers would upload recorded sessions themselves and different graphs for teacher talk moves, student talk moves, and other summaries were generated. It was left to the teacher to interpret what was useful. In the re-contextualized app, sessions are automatically uploaded. Coaches can navigate across each of their tutors, set the data presented to be grouped by class or date, and look at trends across specific talk moves.

Saga also adapted their approach to PL based on lessons learned from the co-design process. In addition to exploring TalkMoves and all the information it could provide, Saga also had to reconcile how *accountable talk theory* that guided TalkMoves' development was in line with institutional commitments. This was a delicate balance of both learning as the sessions unfolded and imagining what would be most beneficial in their practices for the future. By incorporating TalkMoves into their daily practice, this also meant that the coaches' practices for supporting tutors could be positively disrupted by this collaboration. An example of this can be seen in how coaches originally had specific details that they look for in the tutor's sessions that pertained to relationship building, rigor and mathematical knowledge. The TalkMoves application provides a new frame of reference to look for these types of details. One of the coaches stated in their final one-on-one interview that they were still looking for all of the correlations between talk moves and Saga frameworks and that they've had to step back and think a lot about how they identify things. So, while the inclusion of the app makes identification faster, it also requires the coaches to learn a new perspective. Situating TalkMoves into a new context required Saga to

understand what TalkMoves could offer as well as its future capabilities, and for CU researchers to understand what the critical issues and challenges Saga coaches experienced in their workflow.

*RQ2: Facilitator strategies:* Facilitators guided productive collaboration by establishing communal orientation, promoting expansive dreaming and backcasting, and continually revoicing collaborator ideas. While co-design is a mutually enriching experience, getting multiple stakeholders involved is only one piece of the puzzle; well-directed meetings is another vital component (Penuel et al., 2007; Hoadley, 2002). Facilitators ensured productive meetings by providing unique perspectives and implementing critical scaffolds. Two facilitators were boundary spanners (Akkerman & Bakker, 2011); one researcher had been a math teacher, and former employee of a startup that helped design Saga's digital math environment. The Saga facilitator was originally a Saga tutor and coach. Their backgrounds afforded them perspectives from practitioners and developers, increasing the boundary permeability of the design work to create a tool that meets the needs of both. The facilitator strategies provided intentional structure that encouraged new ideas to grow and build on each other. Each session was initiated by being framed around a group goal, creating communal orientation, and ensuring that conversations addressed how to assist coaches in their workflow. This approach promoted the acceptance of TalkMoves; if the focus was limited instead to only 'this is the tool, how would you change your work to use it', this would have stunted the ideas produced and would not have been as well received.

Early sessions incorporated expansive dreaming, while later sessions attempted to backcast those goals onto the technical and learning environment constraints of the tool (Quist & Vergragt, 2006). For example, before viewing TalkMoves, participants watched a five-minute video of classroom instruction and were asked to reflect on the different observations or data that they might expect to use (e.g., what might they expect to see for talk distribution between teacher and students). It was only after everyone had their own expectation set that the data from TalkMoves was presented. Having everyone imagine their typical practice first ensured that they could clearly compare their expectations to that of the app and imagine new uses. Facilitators continually tried to elicit ideas independent from the current app function so that participants would not limit their ideas. Additionally, reflection time, free interaction, and visual aids were used both during and outside of the sessions. Coaches and supervisors created a short reflection recording, had built in time during sessions to explore the app, and visual aids were always used as a jumping off point. Initial TalkMoves' graphs progressed to wireframes that displayed the suggestions produced from the meetings. These detailed examples helped to solidify what and how the coaches and supervisors imagined using the application in their typical workflow. Finally, facilitators were very intentional about how they engaged with collaborators. They frequently revoiced the suggestions or comments, which highlighted the value of everyone's contributions. Revoicing is also a talk move that has been identified as an effective way to promote accountable talk (Michaels et al., 2008), and was a common occurrence which often solidified ideas or uncovered new topics. Multiple elements worked together to nudge everyone to generate refined ideas of what was needed and how exactly TalkMoves could best work.

*RQ3: Emergence of critical design themes and overarching goals:* By examining all of the comments across sessions, 15 themes and 3 main design goals emerged. Goal one: the interface should enable user agency, and is supported by themes such as Filter (ability to adjust presentation of data through selection, sorting or comparison), Exemplars (specific high-quality examples), Sharing (how information is shared between coach, supervisor, and tutor) and View (specifics of how data is presented). Goal two: the interface should center around goal setting, and is supported by themes such as Trends (displaying data trends), Reflection (ability for tutors to reflect on data, observations and/or feedback), and Goals (relates to goal setting). Goal three: the interface should be integrated into the existing workflow, and is supported by themes such as Perspective (Refers to the different needs across different Saga positions), Bookmarking and note taking (Ability to take notes and mark points within the video/transcript), Classify and Quantify (Additional area identified for data annotation or quantification), Recommender system (Automatic system recommendations), Integration (specific to integrating app into current Saga practice), Alignment (connecting different types of data), Dashboard alert (quick signal on users home page as to status on areas of interest), and Summary (summarization of data or session). These goals were distilled by reviewing the emergent design themes that arose within and across sessions. Design goals are not mutually exclusive, and multiple themes support more than one goal.

Frequency of theme occurrence did not indicate importance. Frequent themes captured important components, however, their continued presence indicated that these areas needed to be elaborated upon or clarified as collaborators' understandings developed. Consistently distributed themes were readily understood concepts that specifically needed refinement in the context of the new tool. Infrequent themes demonstrate how important themes can emerge at different stages. Some themes seen early on do not continue since there was immediate consensus, while others appear later in response to fine-grained decisions. It may be difficult to develop new ideas until a stable picture is established and can only emerge through iterative learning. Themes are interleaved; often a single comment encapsulates multiple themes. For example, one participant stated that they want the autonomy



to send different data (e.g. specific notes, summary, strategy evidence, etc.) to different people since everyone has their own unique way of doing things. This one comment contributes to many themes (Notes, Summary, Sharing). While almost half of the ideas were present from the very first meeting and persisted through subsequent sessions, these ideas solidified as the co-design team built their joint understanding of what was needed and what was possible. This was not a trivial task that was completed in a single session. Instead, differing degrees of learning from each other continued through all of the meetings. Through this exchange we see solution mapping and how the team mutually designed the app interface.

## Conclusion

This study provides rich insights into the development and integration of a new AI tool into an existing educational practice. Solution mapping provides an informative frame for how RPPs can co-adapt tools and solutions to new and old problems. Key to re-contextualizing was allowing for time and space to align understanding of both the problem and solution. This process was aided by beginning sessions focusing on dreaming and later sessions having researchers and coaches share complementary expertise to backcast Saga's goals for an ideal future onto the reality of the technology and their workflow. This approach is advantageous for any re-contextualization or implementation of an AI tool.

Additional developments and changes to the tool or practice may only be adoptable after more time has established trust in the current iteration of TalkMoves. TalkMoves may be especially good at automatically identifying many short but critical moments for coaches to observe across multiple sessions. However, coaches maintained that they wanted to be able to follow closely to their normal practice of watching a full twenty minutes of a singular session. Even if the new way of reviewing would save time and give a broader overview, coaches were not ready to solely rely on the app. This is understandable; the coaches need to first trust the app before accepting more dramatic changes to their practice. In a collaboration such as this, designing a new tool can be a slower process in order to establish comfort and trust such that greater change can be accepted later. Trust both in the design as well as in the results is essential, however, it is only through initial use that this latter form of trust can develop. There is a delicate balance between designing a useful tool and designing for change, as these goals are not always in alignment. RPPs must be committed to co-adapting and looking for new ways to improve.

## References

- Akkerman, S. & Bakker, A. (2011) Boundary crossing and boundary objects. *Review of ed. res.*, 81(2), 132-169
- Barter, C., & Renold, E., (1999) The Use of Vignettes in Qualitative Research, *Social Research Update*
- Corbin, J., & Strauss, A. (2015) Basics of qualitative research: Techniques and procedures for developing grounded theory. *SAGE*.
- Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020) COVID-19 and learning loss- disparities grow and students need help. *McKinsey & Company*
- Hoadley, C.P. (2002) Creating context: Design-based research in creating and understanding CSCL. *Proceedings of Computer Support for Cooperative Learning (CSCL)*
- Jacobs, J., Scornavacco, K., Harty, C., Suresh, A., Lai, V., & Sumner, T., (2022) Promoting rich discussions in mathematics classrooms: Using personalized, automated feedback to support reflection and instructional change. *Teaching and Teacher Ed.* (112). DOI:10.1016/j.tate.2022.103631.
- Lee, J.W., Daly, S.R, Huang-Saad, A., Rodriguez, G., & Seifert, C.M., (2020) Cognitive strategies in solution mapping: How engineering designers identify problems for technological solutions. *Design Studies* 71 100967. DOI:10.1016/j.destud.2020.100967.
- Michaels, S., O'Connor, C. & Resnick, L.B. (2008) Deliberative Discourse Idealized and Realized: Accountable Talk in the Classroom and in Civic Life. *Studies in Phil. and Ed.*, 27, 283–297. DOI:10.1007/s11217-007-9071-1
- Penuel, W.R., Roschelle, J., & Shechtman, N. (2007) Designing formative assessment software with teachers: an analysis of the co-design process. *Research and Practice in Tech. Enhanced Learning*. 2(1).51-74
- Quist, J., & Vergragt, P., (2006) Past and future of backcasting: The shift to stakeholder participation and a proposal for a methodological framework, *Futures*. 38. 1027–1045. DOI:10.1016/j.futures.2006.02.010.
- Saldaña, J. (2016) The coding manual for qualitative researchers (3rd ed.). Thousand Oaks, CA: Sage.
- Suresh, A., Penuel, W., Jacobs, J., Martin, M. & Sumner, T. (in press) Using AI tools to provide teachers with real-time feedback on their classroom discourse patterns. In X. Zhai & J. Krajcik (Eds), *Uses of Artificial Intelligence in STEM Education*. Oxford University Press.

# Exploring Families' Science Talk Using a Mobile Augmented Reality App: An Epistemic Network Analysis Method

Leqi Li, Susan M. Land, Heather Toomey Zimmerman, Yanping Xu  
Lzl5599@psu.edu, sml11@psu.edu, haz2@psu.edu, ymx5173@psu.edu  
Penn State University

**Abstract:** This research applies the Epistemic Network Analysis (ENA) method to analyze 11 families' science talk as they engaged with mobile augmented reality (AR) to learn about cave formation. Results show that the design features in the *Cave Explorers* mobile AR app triggered four types of families' science talk. Families with high pre-post gain knowledge scores of the app content engaged in more frequent and detailed describing and identification talk styles when encountering the science content and the place-based observation prompts. Children in these families read more science content aloud and used it to make sense of their observations in the cave exhibit by making explanations and inferences.

**Keywords:** Informal learning, science learning, mobile augmented reality, family learning, epistemic network analysis, discourse analysis

## Introduction

With mobile technology and immersive augmented reality (AR) emerging in informal settings (Zimmerman et al., 2015), this study investigated how a mobile augmented reality app for families influenced learning about limestone cave formation as families visited a children's garden. Learning in informal settings is often studied by examining the conversation elaboration that occurs as people talk in naturally occurring groups in situ (Leinhardt et al., 2003). Accumulating evidence also states the essential role of family discourse in supporting children's science learning and science engagement (Vedder-Weiss, 2022). Mobile AR, conceptualized as a cultural tool mediating family's informal science learning, enhances learning possibilities in real-world environments by overlaying a virtual layer of digital content, which can be accessed on devices in outdoor spaces (Ryokai & Agogino, 2013).

Mobile AR has been used to scaffold scientific explanation-building, collaborative scientific argumentation, and scientific observation. Prior research (e.g., Zimmerman et al., 2015) found that families using a mobile AR app at an arboretum engaged in high levels of perceptual talk (i.e., describing and identifying). However, few articles have studied what role the design features of a mobile AR app play in triggering family science talk. To further understand the connection between families' science talk and specific design features of a mobile AR app, an epistemic network analysis (ENA) was applied to answer the following research questions: (1) How does families' science talk connect with specific design features of a mobile AR app?; and (2) Are there differences in science talk between families with higher vs. lower gain scores on a knowledge assessment?

## Methods: Context, participants, and data

*Cave Explorers* is a mobile AR app designed for families with children aged 5-12 to foster learning about how limestone caves form in the U.S. Karst landscapes. It was designed for families to use while visiting a large-scale cave exhibit in a children's garden in central Pennsylvania, U.S.A. Eleven (11) families with 26 participants (11 adults and 15 youths) were chosen as the data set from the total of 34 families. Each family was provided an iPad with the *Cave Explorers* app for a 10-20-minute AR experience. The app includes 7 learning activities (Figure 1). Families started the experience with a family discussion prompt about how caves form in their region (Figure 1a). After entering the cave exhibit, the app prompted the families to find clues about how water and rock interact and take 5 photos with the iPad (Figure 1b). The app then introduced content about how stalactites and stalagmites formed in caves (Figure 1c) and included an AR animation of how stalagmites and stalactites can grow together to form a column. When the families held up the iPad to scan a stalagmite on the cave exhibit floor, an AR animation appeared to simulate the formation of a column within the cave exhibit. Another AR animation of water dripping was triggered when the families scanned a ceiling crack (Figure 1d). The families watched an animated video about how caves form (Figure 1e), and then completed an activity where they tagged their cave photos with scientific descriptions (Figure 1f). The experience ended with a video animation about helping protect our environment (Figure 1g). Screen recordings of the iPads were captured throughout the experience. One family member also wore a cap with an attached GoPro video camera to collect audio-visual data. The same pre- and

post-questions about cave formation were asked to families before and after the experience; the questions were later scored and interrater (Cohen's Kappa = 82.4%). All families were ranked in order of their gain scores, from highest to lowest. Six families, selected from the top scores, comprised the highest-performing group, and 5 families selected from the bottom scores, comprised the lowest-performing group for further analyses. Since each family experienced the same AR app, and families typically read the content aloud, this method of selection minimized the high similarity in their discourse content.

ENA is a learning analytics technique that models the weighted structure of connections in discourse data by identifying co-occurrences in stanzas of coded discourse and modeling the weighted structure of co-occurrences (Shaffer et al., 2016). Research indicates that ENA can effectively analyze science communication (e.g., Bressler et al., 2019), especially when dealing with dynamic and temporal discourse (Csanadi et al., 2018). In this study, families' science talk was coded based on Allen's (2003) framework: Perceptual, conceptual, connecting, and affective talk.

Five design features of the *Cave Explorers* app were coded as: family discussion prompts; place-based observation prompts; photo-taking prompts; AR animations; and science content (text/video/graphics). Two researchers discussed the codebook and independently coded 25% of the transcripts. Agreement of inter-reliability was reached (Cohen's Kappa = 89.3%). After discussing the disagreements and making changes in the codebook, the transcripts coded by the first author were analyzed using ENA. In total, 734 family utterances were coded.

**Figure 1**

*Screen Capture of the Cave Explore MAR App*



Figure 1a: Talk with your family

Figure 1b: Explore inside the cave

Figure 1c: See stalactites and stalagmites

Figure 1d: Observe water in the cave

Figure 1e: Watch how cave forms

Figure 1f: Tag your photos

Figure 1g: Help protect our environment

## Findings

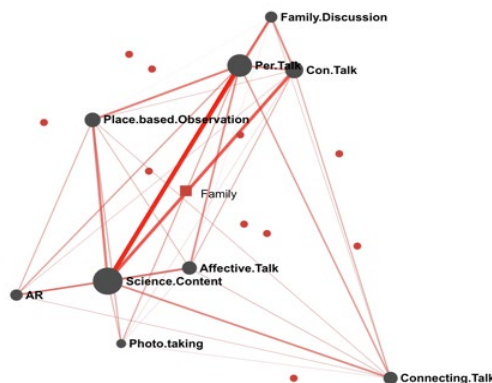
Nodes in the ENA network correspond to categories of science talk and/or design features of the AR app, and edges reflect the relative frequency of connection between two nodes. Thicker lines represent stronger connections, while thinner lines represent weaker connections. In Figure 2, the networks model the broad structure of connections among families' science talk and the *Cave Explorers*' design features. All five design features are connected to four categories of families' science talk. The most frequent talk, represented by the thickest edges in the network, happens between science content and perceptual talk, and then science content and conceptual talk. This means that during the experience, families read the science content in the app aloud, made observations in the cave exhibit, described what happens in the cave, and made sense of what they saw. The excerpt below shows an example of a family's science talk. Usually, parents held the iPad and read the content to the children. In this excerpt, when reading how a column forms, the mother identified a stalactite and stalagmite in the cave and inferred what adding four inches in a thousand years means to the six-year-old boy:

- Mom: "Calcium carbonate forms stalactite on the cave ceiling and stalagmite on the ground. When stalactite and stalagmite meet, they form a column. Caves' features form slowly, only adding four inches in a thousand years. *[reading science content]* Can you imagine?"
- Daniel: No.
- Mom: So, these things, you see them up there (stalactite)? And down below (stalagmite)? *[perceptual talk]* They take thousands of years to get four inches. That's like this big. *[conceptual talk]*
- Daniel: Oh, wow! *[affective talk]*

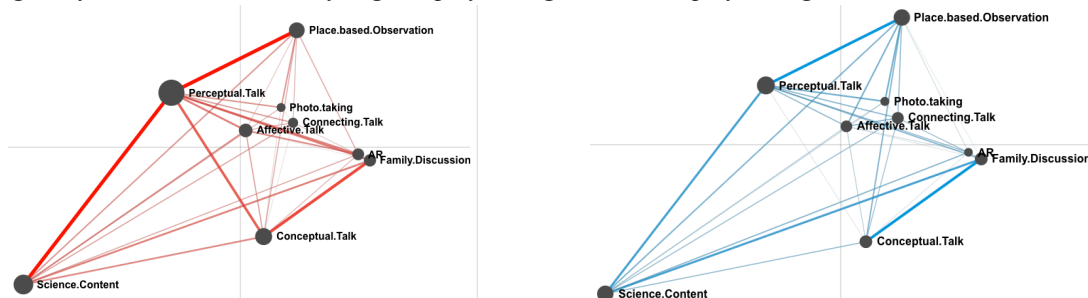
Figures 3 and 4 show the discourse network in the highest-performing family group (N=6) and the lowest-performing family group (N=5). Comparing the two networks in Figure 3, we can conclude that the overall structures of the highest-performing and lowest-performing family groups are similar, likely because families were provided with the AR app with the same content. However, the highest-performing families related perceptual talk with science content and the AR animations more frequently than the lowest-performing group. Additionally, they had more perceptual talk during the place-based observation activity, meaning more scientific observations were generated.

Figure 4 shows children's discourse networks in the highest-performing groups and lowest-performing groups. Both groups identified or described the science phenomenon in the place-based observation activity, but the children in the highest-performing group exhibited more diverse and spread structures, meaning they tended to lead the tour by frequently reading app content and directing the talk. There are connections between science content and affective talk, and science content and conceptual talk in the children's discourse in the highest-performing group, but these connections in the children's discourse in the lowest-performing group are not seen. Children who were cognitively and emotionally engaged with the science content had higher learning gains. Similarly, the connection between place-based observation and conceptual talk, and AR and conceptual talk are only shown in the highest-performing children's discourse, which indicates that these children not only described what they saw in the cave/AR simulation, they also explained what they saw and inferred what it meant to them. Children in the highest-performing families connected perceptual talk with conceptual talk. These co-occurrences suggest sense-making of their observations. In sum, children in the highest-performing group read more science content aloud, made sense of the science content by describing and comparing it to the cave exhibit, explained the science content in their own words, and emotionally engaged with the family discussion.

**Figure 2**  
*Figure of Mean ENA Network of Families (Per Talk: perceptual talk; Con Talk: conceptual talk)*



**Figure 3**  
*Figure of Mean ENA Network of Highest-performing and Lowest-performing Families*



*Figure 3a: ENA of highest-performance family group (N=6 families)*

*Figure 3b: ENA of lowest-performance family group (N=5 families)*

**Figure 4**  
*Figure of Children's Mean ENA Network of Highest-performing and Lowest-performing Families*

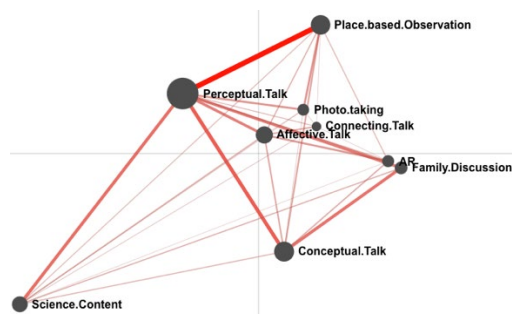


Figure 4a: Children's ENA of highest-performance family group (N=6 families)

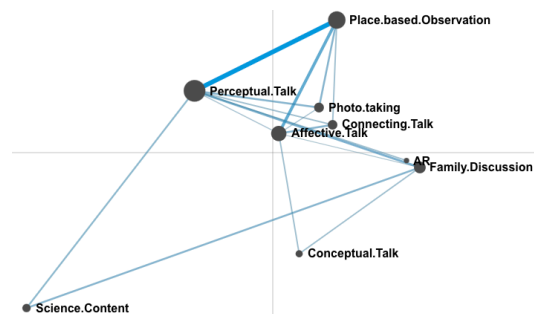


Figure 4b: Children's ENA of lowest-performance family group (N=5 families)

## Discussion and implications

Findings suggest that our mobile AR app successfully triggered four types of families' science talk during the learning experience. The networks' differences in the two-family groupings, especially children's talk, provide insights into how families can generate higher learning gains using a mobile AR app during informal science learning. More science prompts might be needed to trigger learners' perceptual talk and conceptual talk, for example, by asking more open-ended wh- questions to direct attention and elicit meaning-making (Eberbach & Crowley, 2017). Since almost all the families read the app content aloud, these wh- questions could be added to the AR and placed-based observation prompts, for example, "click NEXT to launch the AR and discuss why you think the column formed this way". Likewise, parents can encourage children to read science content aloud or have them lead the experience during informal science learning. Future research will conduct qualitative triangulation to better understand the nature of the talk and connections to the app features.

## References

- Allen, S. (2003). Looking for learning in visitor talk: A methodological exploration. In *Learning conversations in museums* (pp. 265-309). Routledge.
- Bressler, D. M., Bodzin, A. M., Eagan, B., & Tabatabai, S. (2019). Using Epistemic Network Analysis to Examine Discourse and Scientific Practice During a Collaborative Game. *Journal of Science Education and Technology*, 28(5), 553–566. <https://doi.org/10.1007/s10956-019-09786-8>
- Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 13, 419-438.
- Eberbach, C., & Crowley, K. (2017). From seeing to observing: How parents and children learn to see science in a botanical garden. *Journal of the Learning Sciences*, 26(4), 608-642.
- Leinhardt, G., Crowley, K., & Knutson, K. (Eds.). (2003). Looking for Learning in Visitor Talk: A Methodological Exploration. *Learning Conversations in Museums* (0 ed., pp. 265–309). Routledge. <https://doi.org/10.4324/9781410606624-14>
- Ryokai, K., & Agogino, A. (2013). Off the paved paths: Exploring nature with a mobile augmented reality learning tool. *International Journal of Mobile Human-Computer Interaction (IJMHCI)*, 5(2), 21-49.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9-45.
- Vedder-Weiss, D. (2022). Discourse genres and children agency in everyday family science engagement. *Learning, Culture and Social Interaction*, 37, 100649. <https://doi.org/10.1016/j.lcsi.2022.100649>
- Zimmerman, H. T., Land, S. M., McClain, L. R., Mohny, M. R., Choi, G. W., & Salman, F. H. (2015). Tree Investigators: Supporting families' scientific talk in an arboretum with mobile computers. *International Journal of Science Education, Part B*, 5(1), 44-67

## Acknowledgement

This project was made possible in part by the National Science Foundation, grant 1811424. The views, findings, conclusions, or recommendations expressed in this article do not necessarily represent those of the National Science Foundation. We thank colleagues Yu-Chen Chiu, YongJu Jung, Jeff Willians, who assisted with data collection and preparation. We express our appreciation to colleagues Nick Rossi, Mitch Lang, and Brad Kozlek who helped with technology design.

# Audio Analysis of Group Discussion Patterns in Noisy Classrooms Before, During, and After Teacher-Group Interactions

Chris Palaguachi, Cynthia D'Angelo, Elizabeth Dyer, Nessrine Machaka  
cwp5@illinois.edu, cdangelo@illinois.edu, edyer@illinois.edu, machaka2@illinois.edu  
University of Illinois Urbana-Champaign

**Abstract:** Teachers use a variety of teacher-group interactions to support, sustain, and increase small group work and group discussion. These strategies include conferring interactions like nudging and eliciting/probing that help students examine their previous or current conceptual understanding. Conversations found in these teacher-group interactions are complex and difficult to code for and interpret. In this study, researchers use qualitative methods to code teacher-group interactions, and computational methods to interpret the relationship between teacher-group interactions and group discussion before and after teacher-group interaction. The results of this study showed many of the teacher-group interactions increased in group discussion after teacher-group interaction occurred. When coding teacher-group interactions by conferring interactions, nudging had a higher percentage of interactions with increased group discussion after teacher-group interaction occurred compared to the other non-conferring interactions. In exploring teacher-group discussions, we observed how probing/eliciting and nudging change the types of conversations student have during and after teacher-group interactions.

## Introduction and background

Today, teachers are expected to be superhuman. In many schools, teachers are often expected to multitask and play many roles in the classroom. They observe, assess, evaluate, and facilitate over 27 students at any given time. Additionally, teachers are expected to be able to listen to every conversation in the classroom and to provide meaningful feedback to support students' conceptual understanding of domain material. However, this can be quite challenging, especially when it is not uncommon for students in small groups to have loud and noisy conversations with one another (Sherin & Star, 2011). Furthermore, the urgency to jump from one group to another to provide equitable feedback to all students while confronting these noisy challenges can make it difficult for teachers to evaluate the effectiveness of their own interventions. In this study, we investigate the feasibility of using students' non-lexical speech data in small groups to explore the relationship between group discussion patterns and teacher interventions. In this study, we aim to answer the following research questions:

- RQ1 – In what ways can non-lexical speech analytics like voice activity detection be used to measure the effectiveness of small group discussion in naturalistic classroom settings?
- RQ2 – In what ways do different types of teacher-group interactions influence small group discussion before, during and after teacher-group interaction?

## Theoretical framework and prior research

Educators are called on to craft interventions that help to support students' domain thinking and sensemaking through discursive dialogue. These types of interventions allow for educators to learn more about students' domain understanding and allows them to respond to their students' thinking. However, not all interventions are equally productive for promoting students' learning, building on students' ideas, or fostering students' agency through discussion (Munson, 2019). As previously mentioned, teachers are expected to respond to students by jumping from group to group. Although these responses can be seen as the teacher being responsive, a growing body of researchers are more interested in understanding teachers' responsiveness by further examining the quality of discussion they have with their students (Munson, 2019). Therefore, it is important for educators and researchers to have reliable tools like speech-based analytics that can help provide insights on the quality of teacher-group interactions on students' sense-making and group discussions.

Speech-based analytics, specifically lexical and non-lexical processing like Automatic Speech Recognition (ASR) and Voice Activity Detection (VAD) help in studying group dynamics and teacher-group interactions. Despite these affordances, challenges persist in using these tools in naturalistic settings. Currently, challenges include insufficient modeling of student speech and transcribing speech in noisy, multi-speaker environments. Between the two tools, the VAD within openSMILE offers a more adaptable solution. Aligned with the 12 core commitments of the new era of MMLA that focuses on intentional and privacy-conscious data

processing (Worsley et al., 2021), using non-lexical speech analytics like VAD can aid in evaluating small group work, collaborative problem solving, and the quality of teacher-group interventions.

## Methods

The study presented in this paper uses data from a larger observational dataset of secondary mathematics lessons in the United States (Dyer, 2016). Middle school and high school students were asked to work in small groups on math problems ranging from algebra to trigonometry. Each classroom had up to 27 students. Students were split into 5-7 groups depending on classroom size. Students were then audio-video recorded. Researchers used a pair of Sony HDR-MV1 or Zoom Q4 cam-recorders positioned at the front corners of the classroom capturing footage at 1080p to collect video. Zoom H1 microphones were stationed at the center of each group's table and used to collect group audio. In total, 106 classroom sessions were recorded from 10 instructors.

For this study, we selected one 100-minute lesson from the primary dataset. This classroom session had 27 students split into 7 small groups, and groups consisted of 3 or more students. Qualitative coding of the larger dataset determined that the session we selected had a high amount of group work and teacher-group interaction, and groupwork was a common participation structure used in this class (Hudson et al., 2021). Additionally, content log and timestamp data of when teacher-group interactions occurred were also coded from the previous study.

## Audio analysis and labeling data

For this study, we used openSMILE – an open-source audio processing program (Eyben et al., 2010) to process the audio. We used the Voice Activity Detection (VAD) feature within openSMILE to extract speech turn timestamps. In Palaguachi et al. (2022), researchers modified openSMILE's RMS threshold values to an auto-mode to account for the variability in students' pitch and loudness in group audio, and decreased the turn-segmenting threshold (nPre and nPost) values to detect shorter turns. Similar to Palaguachi et al. (2022), we merged the different group microphone data and categorized them by colors. A duration variable was created by subtracting the turns' end-time by its start time. The duration variable allowed us to note the time each turn of speech lasted whenever speech was detected by the VAD. Lastly, we filtered the dataset to get rid of any turns that were less than .5 seconds because those turns were unlikely to contain meaningful speech information.

To segment the audio before and after teacher group interactions, content log and turn data was merged. This included labelling turns with different group mic, teacher-group interaction, and interaction ID information. We used the content logs to create R scripts that allowed us to label the group discussion patterns detected 30 seconds before and after a teacher-group interactions occurred. We then labeled the type of teacher-group interaction based off our qualitative coding. Through our qualitative coding, we removed interaction 21 because the teacher did not have a verbal interaction with the group. After these changes, there were a total of 29 teacher-group interactions.

## Qualitative coding for teacher-group interaction

We qualitatively coded the different types of teacher-group interactions, drawing from Munson's (2019) conferring interaction and Ehrenfeld and Horn's (2020) entry-exit frameworks. Our coding scheme included codes for nudging, eliciting/probing, funneling, and other (i.e., interactions not about student thinking or non-conferring interactions) teacher-group interactions. Interaction 2 was dramatically longer than all the other interactions, and for this reason, interaction 2 was removed from this analysis. For the 28 teacher-group interactions, the inter-rater reliability between two coders yielded a Kappa of 0.884, signifying a high level of agreement. Both coders had a discussion on the three interactions that they did not agree on and came to a consensus.

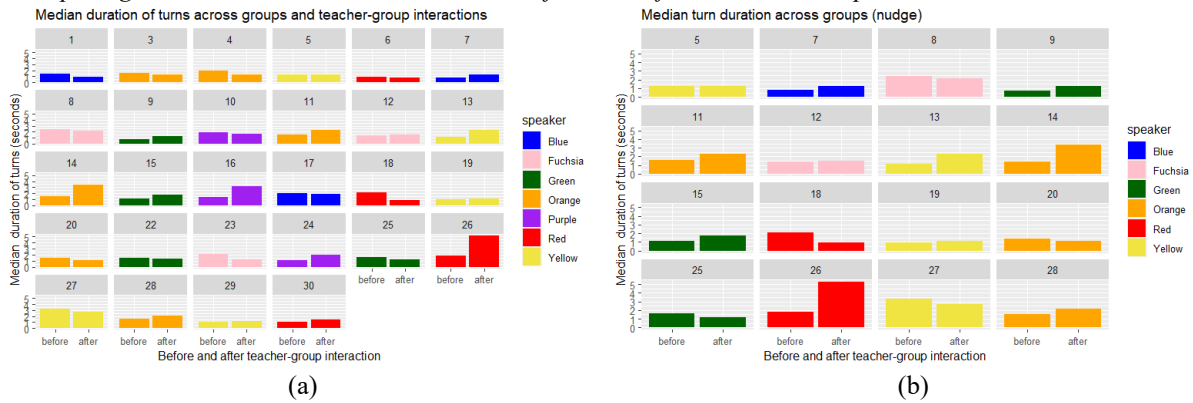
## Findings

When examining the group discussion patterns before and after teacher-group interaction, the main metrics we looked at were the median length of turns, the total duration of turns, and the number of turns for each interaction. When comparing the group discussion patterns before and after teacher-group interactions, we found that for the median duration of turns there were 15 out of 28 (54%) interactions that had increased median duration of turns after the teacher-group interaction compared to the group discussion before (see Figure 1a). In further investigating the median duration of turns before and after teacher-group interaction, we found that for some interactions there were instances of a dramatically larger increase in median duration of turns after teacher-group interaction, however, in cases where there were decreases in median duration of turns after teacher-group interactions the decrease tended to be small or a minor decline. Additionally, while exploring

three different metrics to analyze group discussion patterns (median turn duration, total duration of turns, and number of turns) we found, in all three metrics, there were greater percentages of interactions that had increased group discussion after teacher-group interaction.

**Figure 1**

*Comparing Median Turn Duration 30 Seconds Before and After Teacher-Group Interactions*



This is important to keep in mind as we further investigate the differences in conferring interactions and group discussion patterns before and after teacher-group interactions.

### Group discussion patterns before and after conferring teacher-group interactions

When examining the group discussion patterns before and after conferring and non-conferring interactions, we used median duration of turns as our key metric. Since nudging involved teacher uptake and student sense-making, we were curious if these interactions were related to increased group discussion after teacher-group interaction occurred. In this analysis we used the median duration of turns metric because it accounted for the skew in the turn duration distribution. When comparing the differences in median duration of turns before and after the conferring interactions, we found that 11 out of 16 (69%) of the nudging interactions had a greater median turn duration after the teacher-group interaction compared to the group discussion before (see Figure 1b). In contrast, 1 out of 5 (20%) of the eliciting/probing interactions and 3 out of 7 (43%) of other (non-conferring) interactions had a greater median turn duration after the teacher-group interaction. Since there were a greater number of nudging interactions detected, and since nudging had a greater percentage of increased median duration of turns after teacher-group interaction compared to all the other conferring interactions, we focused on the patterns across nudging interactions.

In further investigating the median duration of turns before and after nudging interactions, we found that the same pattern we observed in the previous analysis of teacher-group interaction was consistent with what we saw when observing the nudging interactions. In the nudging interactions we observed a majority of increased median duration of turns after teacher-group interaction occurred, and when median turn duration decreased after nudging, these decreases were minor. We also observed that the teacher did not use any nudging with the purple group, and only one nudging interaction with the blue group. In contrast, the orange group and yellow group both received four nudges. These findings suggest that when comparing group discussion patterns before and after teacher-group interactions, conferring interactions like nudging have a greater percentage of interactions with increased median turn duration after the teacher leaves the group.

### Qualitative findings of conferring teacher-group interactions and group discussion

For eliciting/probing teacher interactions, we found examples of when the teacher-group interaction preceded students asking one another for clarification. In one example (interaction 16), a student shared their thinking about how to convert between radians and degrees as a response to the teacher eliciting their thinking on a problem (i.e., asking “what you got”). Quickly after the teacher steps away, another student in the group asked the student “can you explain that to me?” which began a conversation that was a shift from the silent, individual work the group engaged in before the teacher interaction. Thus, in this case, a student sharing some of the details behind their work to the teacher likely led to the other student asking them to explain their thinking.

### Discussion and future work



By exploring the differences in small group discussion patterns before, during, and after teacher-group interaction through this exploratory approach, this study has provided insights on how non-lexical speech analytics can be used to better understand effective teacher-group interactions. Furthermore, our qualitative examples highlight the complex nature of the potential response that comes from quality teacher-group interactions. The implications of this study suggest that non-lexical tools like VAD can be used to explore the quality of teacher-group interactions and that quality teacher-group interactions can enhance student engagement and collaborative dialogue in small groups. In our ongoing work, we plan to expand how we code for teacher-group interactions to better account for the nuances and complexities found within some of the conferring interactions. Other approaches that include ASR will enable us to integrate both non-lexical and lexical processing methods to investigate moments of confusion, frustration, and uncertainty, which are instances where educators might be more likely to intervene. Through the knowledge gained from this study, we suggest further customizing the RMS thresholds values for individual groups (in this case detecting louder turns for quieter small groups) or using individual microphones as two potential solutions. In a future study, we hope to address how these changes to the VAD and how we collect audio data can potentially ameliorate these issues.

## Conclusion

To conclude, we learned a lot about the benefits of combining additional qualitative codes to computational analysis of classroom audio. We learned that when working with non-lexical information like turns of speech, context information like content logs, interaction IDs, and conferring interactions are extremely helpful in accounting for how some of these types of interactions effect important turn metrics. By using qualitative methods to code teacher-group interactions and speech analytics to interpret the relationship between teacher-group interactions and group discussion, we gained a deeper understanding of group discussions. Further qualitative coding is necessary to parse through the complexities and nuances of group discussion.

## References

- Dyer, E. B. (2016). Learning through teaching: An exploration of teachers' use of everyday classroom experiences as feedback to develop responsive teaching in mathematics (Doctoral dissertation, Northwestern University).
- Ehrenfeld, N., & Horn, I. S. (2020). Initiation-entry-focus-exit and participation: A framework for understanding teacher groupwork monitoring routines. *Educational Studies in Mathematics*, 103(3), 251-272.
- Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor. *Proceedings of the International Conference on Multimedia - MM '10*, 1459.
- Hudson, H., Parr, E. D., Dyer, E. B. (2021). Factors influencing teachers' use of groupwork in secondary math classrooms. Poster presentation at TN STEM Education Research Conference.
- Munson, J. (2019). After eliciting: Variation in elementary mathematics teachers' discursive pathways during collaborative problem solving. *The Journal of Mathematical Behavior*, 56, 100736.
- Palaguachi, C., Cox, E., & D'Angelo, C. (2022). Audio analysis of teacher interactions with small groups in classrooms. *General Proceedings of the 2nd Annual Meeting of the International Society of the Learning Sciences 2022*, (pp. 439 – 442). Hiroshima, Japan: International Society of the Learning Sciences.
- Sherin, B., & Star, J. R. (2011). Reflections on the study of teacher noticing. *Mathematics teacher noticing: Seeing through teachers' eyes*, 66.
- Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A New Era in Multimodal Learning Analytics: Twelve Core Commitments to Ground and Grow MMLA. *Journal of Learning Analytics*, 8(3), 10-27.

## Acknowledgements

This material is based upon work supported by the National Science Foundation (DRL-1920796). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

# A Framework for Collaborative Learning in Immersive Virtual Reality

Paul Hatch, University of Illinois Chicago, USA. phatch2@uic.edu  
Joseph E. Michaelis, University of Illinois Chicago, USA. jmich@uic.edu

**Abstract:** Collaborative learning in Immersive Virtual Reality (IVR) has the promise of becoming an essential component in learning environments by blending higher level learning activities with modern educational technologies. However, our understanding of how to implement collaborative learning in IVR is still developing. To aid researchers and designers in this area we created a framework based upon multiple theories of collaboration and interaction and tested it by conducting an initial systematic review. We reviewed 32 papers that incorporated collaborative activities into IVR activities and identified what processes and supports they implemented. We found that most implemented problem-solving, listening and discussion processes, and recreated real-world environments and objects to facilitate collaborative interactions. Importantly, our framework also helped identify meaningful gaps in collaborative learning in IVR, such as how the unique affordances of IVR are largely underutilized, and how the use of this framework may support further work in this area.

## Introduction

Collaborative learning has been long recognized as an effective method to enhance learning, develop communication and collaborative skills, build problem-solving skills, and provide significant cognitive benefits to articulation, conflict, and co-construction. (Brown & Campione, 1990; Won et al., 2023). Collaborative learning is effective if it occurs in coordinated, synchronous activities in which group members interactively construct knowledge through socio-cognitive activities such as engaging in rich interactions, explaining, and negotiating their ideas, shifting viewpoints and roles, engaging with others to resolve conflicts, and building shared knowledge situated in meaningful contexts (Kirschner et al., 2018). However, learners often have difficulties engaging spontaneously in collaborative learning activities without instructional guidance (Vogel et al., 2017).

There has been a recent surge in popularity and access to Immersive Virtual Reality (IVR) technologies (Won et al, 2023) and to platforms that can host multiple users to socialize (e.g., AltspaceVR, VRChat, and Horizon Worlds), work together (e.g., Spatial, and Horizon Workspaces) and learn together (e.g., ClassVR, and Engage). The use of virtual reality in education has been shown to facilitate crucial aspects of learning (Makransky & Peterson, 2021), but we still don't fully understand the affordances and limitations of how virtual reality can provide processes and supports for effective collaborative learning especially in immersive virtual reality (Lui et al., 2023). Towards that goal, this paper presents a new framework from which to view collaborative learning in IVR. We also used our framework to conduct a systematic literature review of current work on collaboration in IVR to inform designers and researchers how collaborative learning processes and supports are currently used. Thus, we uncover areas ripe for further study in collaborative learning in IVR and provide a framework to aid such work.

## The immersive virtual reality collaborative learning framework

There are many technological methods to access virtual worlds, which can be viewed as being part of a 'reality-virtuality continuum' (Milgram et al, 1994) that includes the use of desktop, large scale surround-screen projection (i.e., CAVE), augmented reality (AR) headsets or virtual reality headsets, each providing a different level of immersiveness. While the processes involved in interacting in a virtual environment may be very similar between these input technologies, it is apparent that the *affordances of collaboration* in virtual environments are technologically and physically distinct. This work focusses on Immersive Virtual Reality (IVR), where a fully immersive headset is used to access virtual environments, as it provides a fully unique interaction with the environment and fellow learners. For example, in an AR or CAVE collaboration the students can see and hear their collaborators with all the affordances of the real-world, whereas in a fully immersive headset the modes of communication, including verbal, gestural, and visual, are restricted to how the system interprets them.

Literature reviews have studied the use of virtual reality to support collaborative learning (Van der Meer et al., 2023), and the role of productive social interactions in technology-supported learning environments within VR (Won et al., 2023). Others studied the benefits of learning in IVR over desktop-based virtual learning (Cromley et al., 2023; Wu et al., 2020). However, there are no literature reviews on *collaborative learning using only IVR*, nor are there frameworks with which to view collaboration within IVR.

**Figure 1**  
*The IVR Collaborative Learning Framework*

Processes	Problem-solving		Listen & discuss		Create & explore		Training	
Supports	Scripts	Teacher-led	Assigned Roles	Reflection	Shared Summary	Environment	Tools & artifacts	

To build the framework (figure 1) from which to view the use of collaboration within IVR environments and understand (i) what types of *processes* or activities can lead to effective collaborative learning, (ii) how collaborative learning can be *supported* during those activities, we integrated multiple theories of collaboration.

## Collaboration Processes and Supports

*Collaborative Processes* can be used to effectively facilitate collaborative learning through the convergence of multiple strategies (Brenner et al., 2021). First, group *problem-solving* activities involve students negotiating roles and combining knowledge to devise solutions, thereby fostering a shared group dynamic that extends beyond individual cognition (Won et al., 2023). *Active listening and discussing* processes play a pivotal role in deepening learning and are crucial for cognitive growth through interactive dialogues (Bergmann and Sams, 2012). The integration of technology can potentially enhance this learning by actively prompting actions that lead to deeper engagement (Dede, 2009). *Creation and exploration* activities provide crucial facets of learning by allowing learners to engage in innovative thinking and deeper subject matter understanding through activities that involve constructing new ideas or manipulating existing ones (MacDowell & Lock 2022), which immersive virtual environments can amplify (Wu et al., 2020). Finally, structured *training activities* that adhere to cognitive apprenticeship models supported by scaffolding techniques and social learning, allow learners to integrate individual skills through collaborative learning (Glazer et al, 2005).

*Collaborative Supports* can be structured to guide and facilitate learner interactions and engagements, and reinforce effective collaborative learning (Dillenbourg, 1999). First, *collaborative scripts* are often at the core of interactive learning environments to provide a framework for structuring dialogues and interactions among learners, aid task-oriented communication, and ensure effective engagement, thereby enhancing effective group cognition (Dillenbourg, 1999; Stahl, 2006). The *teacher's role as a facilitator* helps guide and foster an environment conducive to social knowledge construction, and encourage learners to explore, discuss, and build knowledge collectively (Crook, 1994; Hatch et al., 2023; Hatch & Michaelis, 2023). Additionally, the strategic *assignment of roles* within collaborative learning enhances student interdependence and accountability, facilitating interaction and perspective-taking (Dillenbourg, 1999). *Reflection* serves as a critical tool in collaborative learning, particularly in technology-enhanced settings (Barron, 2003) and allows learners to contemplate their experiences and insights to their individual learning processes. The creation of *shared summaries* at the end of group tasks can effectively organize and solidify the group's collective knowledge (Scardamalia & Bereiter, 1994; Stahl, 2006). The collaborative *environment* or setting can reflect authentic contexts and practices, enabling learners to engage meaningfully and gradually transition from peripheral to full participation within a community (Lave & Wenger, 1991; Stahl, 2006). However, visually simpler environments minimize cognitive load and can increase the efficacy of learning in collaborative settings (Lui et al., 2023). Finally, *tools and artifacts* serve not only as practical aids but also as mediators of learning, fostering cognitive development through social interaction, facilitate authentic learning experiences, support meaningful communication, and aid in the formation of learners' identities within the community (Vygotsky, 1978).

## Method

Having established the IVR Collaborative Learning Framework, we conducted a systematic review of studies on the use of IVR for collaborative learning to find; (RQ1) *what are the collaborative processes used in studies on collaborative learning in Immersive Virtual Reality, and (RQ2) how are these processes supported in IVR?*

The review process was based on the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines (Page et al., 2020). A literature search was carried out using the ACM, IEEE Xplore, ISLS/CSCL, ERIC and Science Direct databases. The search query contained the following key elements: 1) collaboration, 2) immersive VR, and 3) learning or education, and limited studies to those published between 2013 and 2023 in English to capture recent trends in virtual reality learning research. We identified 2,072 publications using the search terms in October 2023 and screened them by browsing their title, abstract, and if necessary, their text body. Here, we excluded many publications that did not use headset-based VR, or where collaboration was not fully within the virtual environment. We examined the remaining 106 study publications

for the same eligibility by browsing the full text. The remaining 32 publications were then coded using the IVR Collaborative Learning Framework to identify the collaborative processes and supports. To understand *what and how* such processes and supports are used, we determined the frequency each was used, in what format or context they were used, and identified where one is highly associated with another.

## Findings

In table 1 we provide an overview of what and how collaborative learning processes were used, and what and how collaborative supports were included, as identified through the Collaborative Learning IVR Framework.

**Table 1**  
*What and How Processes and Supports Were Used*

Process	What and how used
Listen & Discuss (17/32)	Promotes interactive dialogue and critical thinking. Includes lectures with open discussion (7/17), video/simulation with discussion (5/17), modeled actions and open discussion (3/17), and partaking in game/challenge and discussing (2/17). Most were teacher-led (15/17).
Problem-solving (16/32)	Used for role negotiation, articulating reasoning, combining knowledge. Includes prompting collaborative problem-solving using games, puzzles and sequential challenges (6/16), domain knowledge (4/16), and accurately using domain tools (3/16).
Create & Explore (10/32)	Involves exploring or experimenting with environments, objects and attributes, including freely manipulating variables such as angles or chemical bonds (4/10), iterating and communicating ideas through building or drawing (3/10), and partaking in creative exercises and games (3/10).
Training (7/32)	Domain-specific collaborative training, usually in realistic environments (6/7) through live modeling and coaching (3/7), predefined steps (3/7), and role-switching (3/7).
<b>Support</b>	
Tools & Artifacts (28/32)	Variety of virtual objects used for collaboration including pens, whiteboards, and clipboards (7/28), domain-specific tools such as medical instruments, electronic gauges, and firefighting equipment (7/28), building blocks (3/28), manipulable shapes to explore and get feedback (6/28), and interactive puzzles and games that prompt collaborative talk and actions (5/28).
Environmt. (19/32)	Thematic settings to support authentic learning contexts, including a classroom or lecture theater (7/19), work sites such as a building site, ship, a surgery (7/19), and outdoor spaces (2/19).
Teacher-led (19/32)	Collaboration and discussions guided directly by teachers/facilitators, using slides (7/19), artifacts games and environments (9/19), or modeling actions (3/19).
Reflection (13/32)	Prompted feedback from participants on collaboration activity while, which includes post-activity surveys/journals (9/13) and verbal reflections while in virtual environment (4/13).
Assigned Roles (11/32)	Specific responsibilities or characters to facilitate interaction and perspective-taking. Roles included role-switching (4/11) and teacher/student pairing (3/11). They were often used to emulate an experimental test condition (5/11) or to improve collaboration skills (4/11).
Scripts (9/32)	Automated interaction steps aiding communication, coordination and engagement through audio/textual prompts (4/9), questions (3/9), or guiding a teacher or virtual assistant (2/9).
Shared Summary (8/32)	Collaborative summarization of topics/discussions while within IVR, to organize and solidify the group's collective knowledge, including visual presentations (4/8) and verbal summaries (4/8).

## Discussion

The processes most often employed by the studies, 'problem-solving' and 'listen & discuss' are notably both analogous to current instructional methods and therefore easier to implement. Many of these studies simply replicated existing classroom environments and tools and did not take advantage of the affordances of IVR to create unique environments. 'Creation and exploration' activities were less common possibly because it involves subjective and less quantifiable processes such as sketching, freely exploring attributes, and building and manipulating shapes without specific goals. However, collaborative learning theories emphasize how crucial this process is in supporting innovative thinking and deeper understanding (Barron, 2003; MacDowell & Lock 2022) and is a process that IVR is well positioned to amplify because of the affordances of IVR to work outside of the confines and limitations of traditional learning environments. Similarly, very few studies (6/32) used manipulable objects, where users can explore or communicate complex ideas, another key affordance of IVR.

In this study, we created a new framework with which to explore what processes and supports are used in studies on collaborative activities in IVR, and how collaborative learning is supported by those. We believe

this study has many implications for research and design of collaborative learning in IVR. First, our framework may be a valuable tool to evaluate other work, or to help design new collaborative IVR environments. Second, we identified crucial gaps in the literature, particularly regarding the use of creative affordances of IVR to enhance collaboration, such as the use of imaginative settings and manipulable objects for group construction, meaning making, exploration and communication. Such studies would help establish a more defined understanding of how IVR can support collaborative learning beyond traditional real-world methods. This research also has limitations. Primarily, we did not analyze the comparative efficacy of the studies, which warrants its own, longer study, nor did we analyze the papers' theoretical foundations. Also, due to space constraints we are not able to include a full list of the literature reviewed in the test of the framework, but it is available on request.

We hope the insights of this review may inspire researchers and designers to be resourceful of the yet unexplored affordances of IVR to support collaborative learning, particularly as the technology offers expansive creative exploration and collaborative interactions that can support learning in ways otherwise not attainable.

## References

- Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, 12(3), 307–359.
- Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. International society for technology in education.
- Brenner, C., DesPortes, K., Ochoa Hendrix, J., & Holford, M. (2021). GeoForge: Investigating integrated VR and personalized sites for collaboration in middle school science. *Information and Learning Sciences*, 122(7/8).
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243.
- Cromley, J. G., Chen, R., & Lawrence, L. (2023). Meta-Analysis of STEM Learning Using Virtual Reality: Benefits Across the Board. *Journal of Science Education and Technology*, 32(3), 355–364.
- Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *Science*, 323(5910), 66–69.
- Dillenbourg, P. (1999). Chapter 1 (Introduction) What do you mean by “collaborative learning”? *Collaborative-Learning: Cognitive and Computational Approaches*, Vol. 1.
- Glazer, E., Hannafin, M. J., & Song, L. (2005). Promoting technology integration through collaborative apprenticeship. *Educational Technology Research and Development*, 53(4), 57–67.
- Hatch, P., & Michaelis, J.E., Rahman Md A. (2023). PATHWiSE: An Authoring Tool to Support Teachers to Create Robot-Supported Social Learning Experiences During Homework. *Proceedings ACM HCI 7*.
- Hatch, P., & Michaelis, J.E. (2023). “Wait, wait, go back”: Investigating social supports for homework during co-design sessions with teachers. ISLS Annual Meeting. ISLS, Montreal, Canada.
- Kirschner, P., Sweller, J., Kirschner, F., & Zambrano, R. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported collaborative learning*, 13(2).
- Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- MacDowell, P., & Lock, J. (Eds.). (2022). *Immersive Education: Designing for Learning*. Springer.
- Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive VR. *Educational Psychology Review*.
- Milgram, P., Kishino, A.F. (1994). Taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* E77-D(1)
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge Building and Knowledge Creation. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 397–417). Cambridge Press.
- Stahl, G. (2006). Group cognition computer support for building collaborative knowledge.
- van der Meer, N., van der Werf, V., Brinkman, W.-P., & Specht, M. (2023). Virtual reality and collaborative learning: A systematic literature review. *Frontiers in Virtual Reality*, 4.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: A Meta-Analysis. *Educational Psychology Review*, 29(3), 477–511.
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard Press.
- Won, M., Ungu, D. A. K., Matovu, H., Treagust, D. F., Tsai, C., Park, J., Mocerino, M., & Tasker, R. (2023). Diverse approaches to learning with immersive VR a systematic review. *Computers & Education*.
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51(6), 1991–2005.

## Parsing the Use of Computational Concepts with Scratch Projects

Joey Huang, North Carolina State University, joeyhuang@ncsu.edu

**Abstract:** Studies in learning sciences have examined the learning of computational thinking through project-based learning. The application of computational concepts has been proven to be related to specific project types. Although studies have suggested the value of examining the relationship between computational concepts and project type, few have addressed the nuances regarding how different project types support specific concepts, as well as how educators could use to inform instructional design. This study examined students' group projects to understand how computational concepts associate with specific project types. The implications of the findings inform how educators can better design the instructional and learning objectives to facilitate computational thinking through project-based learning.

### Introduction

In recent years, extensive research has focused on teaching and learning through project-based learning (PBL) and computational thinking (CT) (Saad & Zainudin, 2022). Studies have explored PBL's application across diverse domains such as robotics, big data, game design, and programming (Chiu, 2020). Notably, there has been widespread adoption of PBL within computer science courses, which has focused on CT as a central concept in teaching and learning. Scratch is a prevalent tool to teach CT with PBL approach (Zhang & Nouri, 2019). Additionally, project types (e.g., games, animations, story) have been proven to demonstrate different uses of CT concepts (Park & Shin, 2019). Previous studies have found that games and animation support specific CT concepts, such as loops, user interaction for younger students (Maloney et al., 2008). Particularly, games projects have a higher number of uses on CT concepts like parallelism, conditionals, or data than projects like music or stories. Although studies have suggested the value of examining the relationship between CT concepts and project type, few studies addressed the nuances of how different project types support different CT concepts and how educators might use this knowledge in instructional design. Understanding how these CT concepts are associated with specific project types will help advance instructional design and learning objectives. It can provide a more holistic view by applying project design to facilitate the learning of CT concepts. In this study, I ask: 1. To what extent do students in small groups apply CT concepts differently across project types? 2. How do student groups utilize CT concepts across different types of projects?

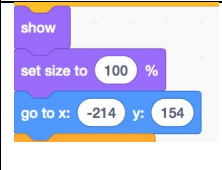
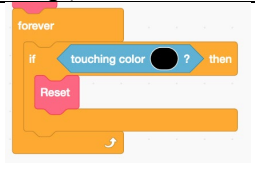
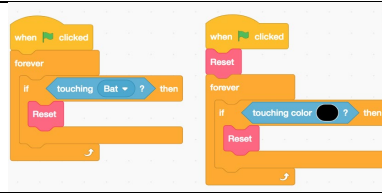
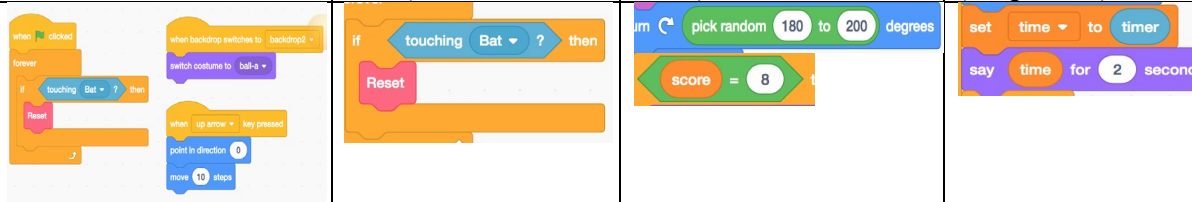
### Methods

The data were collected in a public middle school in a midwestern U.S. state that developed K-12 CS curriculum standards. They were collected as part of a five-week curriculum in an Introduction to Computer Science class. The data included four triad groups with a total of 12 students (Female = 4, Male = 8). The groups consisted of a mixture of novice to experienced students. The majority of students were Caucasian (n=8), two were Asian, and two were Hispanic. During the five-week curriculum, students worked in small groups to complete different types of projects. A total of 25 final projects (five from each group per project type) were collected and categorized into five project types - music, animation, interactive collage, story, and games. The projects were coded based on the seven concepts from Brennan & Resnick (2012)'s framework: sequences, loops, parallelisms, events, conditionals, operators, and data (see Table 1). Collaborating with a trained researcher familiar with Scratch, we analyzed the group projects together. We coded a total of 25 projects together. The group projects were exported as JSON files to examine the number of use of events (yellow blocks, i.e., when green flag, when this sprite clicked, broadcast message one), loops (repeat or forever blocks), conditionals (orange blocks, i.e., if-then, wait) (because the conditionals overlap with repeat or forever blocks, the analysis only included the ones outside these two blocks as conditionals), operators (green blocks), and data (variables blocks). In addition to these five CT concepts, we identified the use of sequences and parallelism by looking at the script of Scratch projects. Particularly, to identify as sequences, the blocks should include at least one event and at least two blocks after the event block, and parallelism indicates there were at least two identical triggers (i.e., when green flag clicked, when space key pressed, when this sprite clicked, when backdrop switches to x, when  $x > \text{"variable"}$ , or when I receive message 1) in one sprite. Descriptive statistics were calculated, and the one-way ANOVA was applied to determine differences in each of the seven CT concept across five design projects.

To understand how students applied CT concepts differently across project types, I detailed the post-hoc findings by providing a sample project of games to describe the use of the four CT concepts. Because these five project types were introduced in sequence during the implementation, I then examined the purpose and structure

of the blocks to understand if changes in students' use of CT concepts could be attributed solely to learning from the projects, or if the project type also had an impact.

**Table 1**  
*Computational Thinking Concepts (Brennan & Resnick, 2012)*

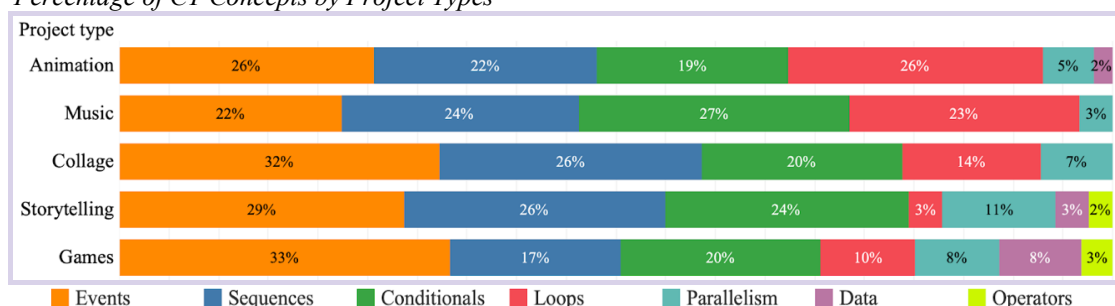
Concept	Sequence	Loops	Parallelism
<b>Description</b>	Identifying a series of steps for a task	Running the same sequence multiple times (in light orange)	Making things happen at the same time
<b>Example from Student Projects</b>			
<b>Events</b>	<b>Conditionals</b>	<b>Operators</b>	<b>Data</b>
One thing causing another thing to happen (yellow blocks)	Making decisions based on conditions (light orange blocks)	Support for mathematical and logical expressions (green blocks)	Storing, retrieving, and updating values (orange blocks)
			

## Results

### Findings of computational concepts by project type

The results were presented as percentages to show the proportion of each concept applied in each of the five projects. Figure 1 illustrates that events, sequences, conditionals were the dominant concepts across five projects (between 17% - 33%). Particularly, these three concepts were emphasized as part of the instructional objectives in animation projects, which was the first group project. In addition to these three concepts, loops were included in the instructional objectives in the first three project units – animation, music, and collage. The groups applied higher percentages of loops on animation (23%) and music projects (26%) compared to the other three project types (less than 14%). Parallelism was applied between 3% to 11% across five project types. Data and operators were considered as more advanced concepts (Weintrop et al., 2018) compared to the other five concepts and were introduced later in the project units – story and games. However, a small percentage (2%) of data concept was used in couple animation projects, which included score variables as an interactive element between users and the program. The findings provided a general overview of the CT concepts applied across the five project types.

**Figure 1**  
*Percentage of CT Concepts by Project Types*



In addition to analyzing the percentage of CT concepts by project type, I conducted a one-way ANOVA with frequency count to examine the differences in CT concepts across project types. Statistically significant differences were observed in four out of seven CT concepts, as determined by the one-way ANOVA. First, sequences ( $F(4,20) = 3.338, p = .030, \eta^2 = .400$ ). A Tukey post-hoc test revealed that the CT concept, sequences,

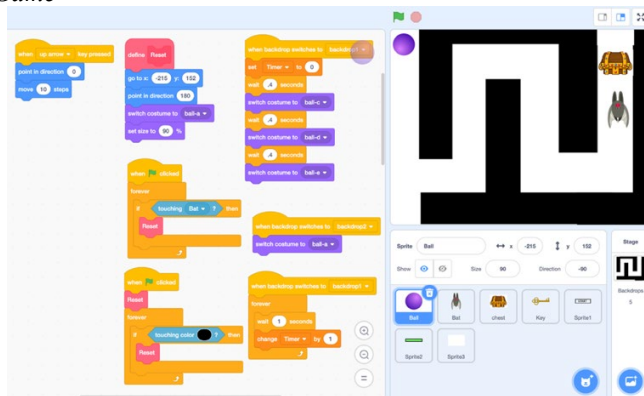
was statistically significantly different for the project type between games and the other two project types (i.e., animation and music). The effect size is large. Next, a Tukey post-hoc test showed that the CT concept, events, was statistically significantly different for the project type between games and other four project types: music, collage, animation, and story with a large effect size. Events ( $F(4,20) = 9.118, p = 0.000, \eta^2 = .646$ ). In addition to sequences and events, parallelism was statistically significantly different for the project type between games and other three project types: animation, music, collage, with a large effect size. Parallelism ( $F(4,20) = 8.004, p = 0.001, \eta^2 = .616$ ). Finally, the post-hoc test revealed that the CT concept – data, was statistically significantly different for the project type between games and other four project types: music, collage, animation, and story and with a large effect size. Data ( $F(4, 20) = 63.154, p = 0.000, \eta^2 = .927$ ). The results of the one-way ANOVA showed there were statistically significant differences between CT concepts and project types on four out of seven CT concepts: sequences, events, parallelism, and data. Additionally, the Tukey post-hoc tests showed that the games project was significantly different from the other four types of projects: music, collage, animation, and story on the CT concepts – data and events. Furthermore, the games project was significantly different from the three types of projects: animation, music, collage on the CT concept – parallelism. Finally, the games project was significantly different from the two types of projects: animation and music on sequences.

### Sequences, events, data, and parallelism in games project

Building on the quantitative results, I then examined the use of these four CT concepts (i.e., sequences, events, parallelism, and data) in games. Due to the limited space, I provided an example from a maze game, created by a student group, to illustrate how students applied the concepts. The group designed a maze game with a goal of opening a treasure box by passing two levels of the maze, where the player needs to pass the first level to receive the key for the treasure box. The group included various obstacles, such as a bat in Figure 2.1, in the game. The group used a variety of sprites to complete the game design. Events were required for every sprite to trigger the following actions (events - yellow blocks). Additionally, to ensure the game ran successfully, the group needed to consider the series of actions of each sprite (sequences), and these actions had to run simultaneously (parallelism) for the game to function. Due to the complexity of the blocks, the group applied the data function (pink block). The use of data blocks increased efficiency for reviewing and modularization for further adaptation and script changes.

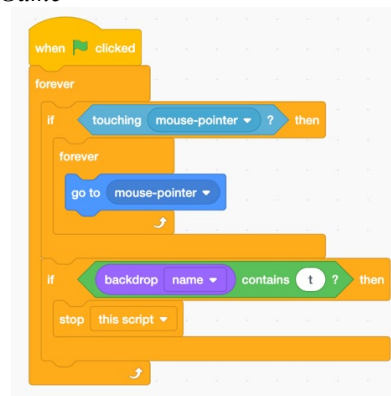
**Figure 2.1**

*The Use of Events, Sequences, Parallelism, and Data in a Maze Game*



**Figure 2.2**

*Use of Loops and Conditionals in a Maze Game*



Loops were a key learning objective across most units, prominently used in animation, music, and collage projects to repeat actions like costume changes and movements. In contrast, their application was less frequent and more complex in story and games projects. Particularly, loops in games included a conditional statement to specify a condition to stop the iteration. Figure 2.2 shows the group applied multiple conditional blocks, “if-then,” which are nested in a forever block, and another forever block was applied inside the if-then conditional block. This usage of loops was common in games projects. Loops is a challenging concept to teach in programming, especially for novice learners (Grover et al., 2016). In animation, music, and collage projects, groups predominantly used “wait” and “wait until” blocks, while “if-then” conditional blocks were more common in story and games projects. Games projects also incorporated “clone” blocks, enabling sprites to duplicate themselves during runtime. Although not a specified instructional goal, this advanced use of “clone” and “if-then” blocks showcases the groups' adept handling of complex conditional logic. This reflects the tailored application of loops



and conditionals across different projects, with games requiring a deeper grasp of these concepts for effective script execution (Park & Shin, 2019). In addition to loops and conditionals, parallelism requires the groups to orchestrate simultaneous events that trigger multiple executions of programs. For animation and music projects, the groups often used event blocks, such as “when green flag clicked,” “when the sprite is clicked,” or “when (left arrow) key pressed” for multiple sprites to achieve the goals of the projects. The groups applied parallelism, which consisted of multiple different types of events, such as “broadcast” and “when backdrop switches to (backdrop 1).” These events were applied under multiple sprites in collage, story, and games projects. Particularly, for games projects, many programs of sprites needed to proceed simultaneously to achieve the requirements of the project.

## Discussion

Building on previous studies, this study examines the nuances among CT concepts and their association with different project types. The results showed the distribution of the seven CT concepts across five project types — animation, music, collage, story, and games. Findings suggest that animation projects support the use of CT concepts like events and loops, while music projects support the use of conditionals (e.g., if-then, stop all). Collage and story projects demonstrated a higher number of sequences and events. Games projects featured the highest number of events among all project types, with conditionals and sequences being more frequent compared to the other four project types. Particularly, post-hoc results of the one-way ANOVA indicated a significant presence of four CT concepts: sequences, events, parallelism, and data.

The findings show that the complexity of projects increased progressively from the first unit (animation) to the last unit (games). Although it was not surprising to find a general progression in project complexity over time, certain CT concepts seemed to be associated with specific project types. In games projects, groups frequently applied conditional blocks nested with forever or repeat blocks (loops). Furthermore, the findings provide implications for instructional design. PBL has been widely employed in both formal and informal learning settings to teach CT (Lye & Koh, 2014). Tools like Scratch are designed with a “low floor” (easy to get started) and a “high ceiling” (complex projects) for students to engage their interests and create personal or group-meaningful projects. By understanding the use of CT concepts across project types, educators can better design the instructional and learning objectives to facilitate student learning. The findings provide insights for educators to develop and refine their project guidelines and curriculum to align instructional expectations and learning outcomes. Educators can consider project types to guide students in advancing specific CT concepts and refining their instructional goals.

## References

- Brennan, K., & Resnick, M. (2012, April). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada* (Vol. 1, p. 25).
- Chiu, C. F. (2020). Facilitating k-12 teachers in creating apps by visual programming and project-based learning. *International Journal of Emerging Technologies in Learning (iJET)*, 15(1), 103-118.
- Grover, S., Pea, R., & Cooper, S. (2016, February). Factors influencing computer science learning in middle school. In *Proceedings of the 47th ACM technical symposium on computing science education* (pp. 552-557).
- Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12?. *Computers in Human Behavior*, 41, 51-61.
- Maloney, J. H., Peppler, K., Kafai, Y., Resnick, M., & Rusk, N. (2008, March). Programming by choice: urban youth learning programming with scratch. In *Proceedings of the 39th SIGCSE technical symposium on Computer science education* (pp. 367-371).
- Park, Y., & Shin, Y. (2019). Comparing the effectiveness of scratch and app inventor with regard to learning computational thinking concepts. *Electronics*, 8(11), 1269.
- Saad, A., & Zainudin, S. (2022). A review of Project-Based Learning (PBL) and Computational Thinking (CT) in teaching and learning. *Learning and Motivation*, 78, 101802.
- Weintrop, D., Hansen, A. K., Harlow, D. B., & Franklin, D. (2018, August). Starting from Scratch: Outcomes of early computer science learning experiences and implications for what comes next. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (pp. 142-150).
- Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, 141, 103607.

# MOSAIC Protocol: Analyzing Small Group Work to Gain Insights into Collaboration Support for Middle School STEM Classrooms

Nga Hoang, University of Colorado, Boulder, nga.hoang-1@colorado.edu  
Jeffrey B. Bush, University of Colorado, Boulder, jeffrey.bush@colorado.edu  
Indrani Dey, University of Wisconsin-Madison, idey2@wisc.edu  
Emily Watts, University of Colorado, Boulder, emily.watts@colorado.edu  
Charis Clevenger, University of Colorado, Boulder, charis.clevenger@colorado.edu  
William R. Penuel, University of Colorado, Boulder, william.penuel@colorado.edu

**Abstract:** Collaborative learning is widely viewed as a tool for promoting educational equity and developing 21st-century skills. To support the improved facilitation of collaborative learning, this study applies the MOSAIC video analysis protocol to analyze student collaboration in small groups across 618 moments of support. The MOSAIC protocol provides insights into the types of support received, classroom conditions, and actions preceding and following support. Our analysis revealed a predominance of task-related support, with collaboration-focused support being notably less frequent. Activity design played a crucial role in accounting for the type of support provided.

## Introduction

Collaborative learning equips students with critical skills necessary for workforce preparation, including communication, social aptitude (Bower & Richards, 2006), and shared objectives (Johnson & Johnson, 2009) while also promoting equitable learning opportunities (Werner et al., 2004). However, students frequently encounter challenges managing their learning during small-group activities (Molenaar et al., 2010), and grouping students does not guarantee productive collaboration (Barron, 2003). Thus, a crucial need exists for targeted support to foster collaboration and improve collaborative learning experiences.

While there is a wealth of research on supporting individual learning (van de Pol et al., 2019; Webb & Palincsar, 1996), scaffolding the collaborative aspects of small-group work in Computer-Supported Collaborative Learning (CSCL) environments is less extensive, highlighting potential areas for further exploration. Artificial intelligence (AI) holds promise for enhancing online collaborative learning environments with features like task guidance and question-answering (Chen et al., 2019), though its application in physical settings is still emerging. For AI agents to be effectively integrated into face-to-face collaborative learning, a deeper understanding of current teacher support strategies in CSCL classrooms and the assistance needed is essential. Our study delves into *support moments* during small-group activities to understand the types of support received across different tasks in a collaborative unit and what student activity may have prompted it.

## Theoretical framework

Collaborative learning takes place when two or more students work to build a shared understanding and construct new knowledge together (Dillenbourg, 1999), through engagement in activities designed to mirror real-world complexities, leverage students' prior experiences, and foster a richer understanding by facilitating peer interaction (Bransford, 1999). Supporting students in their Zone of Proximal Development (Vygotsky, 1978) can enhance collaborative efforts by helping students accomplish together what they could not do alone.

We incorporate the Collaborative Problem Solving (CPS) framework (Graesser, 2019) to dissect the multifaceted nature of collaborative action in educational settings. CPS, involving small student groups tackling problems collectively, highlights skills such as establishing common ground, negotiation, coordination (Barron, 2003; Barron & Roschelle, 2009), knowledge building (Suthers et al., 2010), and effective communication, which are considered essential for collaboration. Further, we integrate CSCL principles, envisioning the AI as a dynamic, intelligent partner in the learning process. This perspective redefines the role of AI from a technological tool to an active collaborator in the learning environment, working synergistically with students and teachers.

## MOSAIC framework

We used the MOSAIC framework (Dey et al., 2024) to analyze specific instances of support within classroom interactions, focusing on *support moments* as the primary unit of analysis. A *support moment* starts when individuals outside the group (e.g., teachers, students from other groups, etc.) provide guidance or instructions to group members, and it concludes when they leave. The MOSAIC protocol (Table 1) examines details of the support moment and students' activity one minute before and after the support moment. Thirteen types of support

are identified during the support moment, with three specifics to collaboration: collaborating with others, including group members, and building group consensus. These categories are not mutually exclusive, allowing a single moment to capture multiple types of support.

**Table 1**

*Summary of the MOSAIC Coding Protocol for Analyzing Support Moments*

<b>General Information of the Support Moment</b>	
<ul style="list-style-type: none"> <li>• Lesson in the Unit (1, 2, 3, 4)</li> <li>• Sensor (environmental sensor, sound sensor, soil sensor, all sensors)</li> <li>• Who Provides the Support? (teacher, peer, researcher)</li> <li>• Who Initiates the Support? (teacher, students, peer, researcher)</li> <li>• Who the Support is Addressed to? (whole class, a small group, a student)</li> </ul>	
<b>Analyzing Student Behavior Before and After Support Moment</b>	
<ul style="list-style-type: none"> <li>• Students on task</li> <li>• Students off task</li> <li>• Students get stuck</li> <li>• Students express a need for direction</li> <li>• Students talk about collaboration</li> </ul>	<ul style="list-style-type: none"> <li>• Students share information</li> <li>• Students ask questions</li> <li>• Students offer ideas</li> <li>• Students build on other's ideas</li> <li>• Socializing</li> </ul>
<b>Type of Support During the Support Moment</b>	
<ul style="list-style-type: none"> <li>• Validation</li> <li>• Strategy for problem solving</li> <li>• Getting the right answer</li> <li>• Collaboration</li> <li>• Direction about the assignment</li> </ul>	<ul style="list-style-type: none"> <li>• Higher academic goal</li> <li>• "Get back to work" directive</li> <li>• Off-task support</li> <li>• Asks questions of the group</li> <li>• Group asks questions</li> </ul>

To develop an AI agent to support small-group collaboration, we seek to answer the following questions:

1. What types of support are provided to students during small group activities in a computer science (CS) based collaborative unit?
2. How does the collaboration support vary based on classroom conditions, such as the activity design, the provider, the initiator, whether the support was solicited, and student activity in the minute before the support moment?

## Methods

### Participants and context

This study was conducted within research-practice partnerships between the NSF Institute for Student-AI Teaming and two school districts in rural and suburban areas of the Western United States. Data collection included video recordings from four teachers across 17 classrooms of 7th and 8th graders, for a total of 302 consenting students. Students participated in these activities in small groups of 2 to 4 members, with teachers circulating to provide support.

The instructional context is the Sensor Immersion unit, an inquiry-based CS curriculum (Gendreau Chakarov et al., 2019) that strongly emphasizes collaborative skills through three kinds of small-group activities:

1. *Card-Sort (Lesson 1)*: Students engage in discussions to identify and understand the characteristics of effective explanatory science models.
2. *Program and Wiring (Lessons 2, 3, and 4)*: Students work with physical sensor systems (environmental, sound, and soil sensors) for the chosen project.
3. *Jigsaw (Lesson 4)*: Each student becomes an expert in one of three sensors and then shares their expertise within a small group of three, promoting an exchange of knowledge.

### Data analysis

We used the MOSAIC coding protocol to identify and examine 618 support moments from 205 videos, totaling approximately 21 hours of small group work. We used Krippendorff's Alpha for reliability, chosen for its flexibility in evaluating various metrics and multi-observer data (Hayes & Krippendorff, 2007). The average Krippendorff's alpha is 0.44, indicating moderate agreement due to the infrequent items in the dataset. The lead author performed expert coding, re-evaluating, and reaching consensus on rare items to improve reliability.

We used logistic regression to investigate the effects of activity type, type of support, and preceding actions on whether groups received collaboration support (1) or not (0), focusing on six categorical variables:

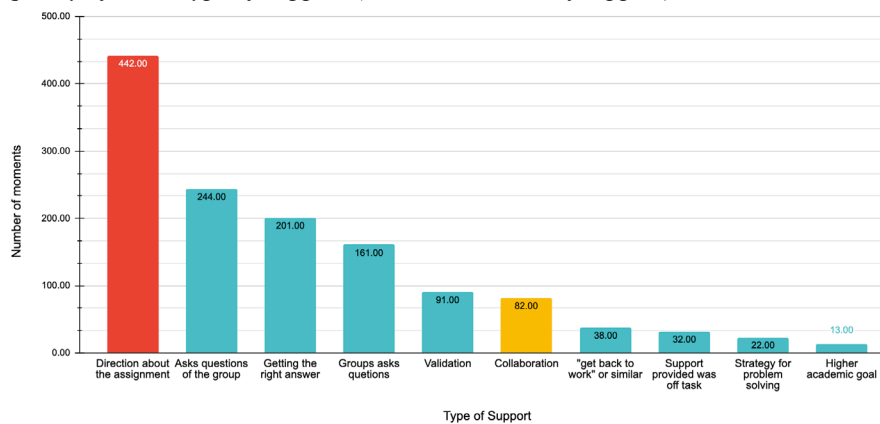
lesson activity, sensor type, support provider, initiator, recipient, and student activity before support. We incrementally built models, including only significant categorical variables adjusted by Sidak correction for conservatism.

## Results

To address our first research question, we analyzed the types of support provided during small group activities (Figure 1). The most frequent support type was *Direction about the assignment*, accounting for approximately 71.5% of the total support. This finding contrasts with the typical expectations for a high-structure classroom, where students usually require less process support (Mäkitalo-Siegl et al., 2011).

*Collaboration* support was found only 82 times, representing about 13% of all support moments and less than one-fifth as frequent as *Direction about the assignment* support. Given the curriculum's collaborative emphasis, this finding prompts consideration of whether teachers should provide greater collaboration in the classroom or if the result reflects a greater student need for task-related support. We also noticed that when the teacher *Asks questions of the group* or the *Group asks questions of the teacher*, the support types mirrored the overall support distribution, indicating that the act of asking and answering questions did not significantly alter the received support.

**Figure 1**  
*Frequency of Each Type of Support (n = 618 moments of support)*



We initially performed a regression model with six covariates to address our second research question. While model fit improved with the addition of covariates and both who initiated and who was addressed by the support had some significant covariation, the most significant results were related to lesson activity. This remained significant even when accounting for other covariates, revealing an association between activity design and collaboration support. Post-hoc tests showed that none of the coefficients showed statistically significant differences when compared to covariates within the same category other than the reference group (i.e., Lesson 2). Analysis of the odds ratios presented in Table 2 showed that collaborative support was significantly more likely (odds ratio 2.31,  $p < 0.01$ ) when students worked on the Lesson 4 jigsaw than when they programmed in Lesson 2. Lesson 2 was significantly less likely (odds ratio 0.12,  $p < 0.001$ ) to have collaboration support than other lessons.

**Table 2**

*Results from Logistic Regression to Explore the Impact of Lesson Activity on Collaboration Support*

Covariate	Categorical Variable 1: Lesson Activity				
	L1 Card sort	L2 Programming	L3 Programming	L4 Jigsaw	L4 Programming
Model 1	1.44 (-0.56)	<b>0.12***</b> <b>(-11.57)</b>	0.45 (-1.45)	<b>2.31**</b> <b>(-2.77)</b>	0.50 (-0.94)

pseudo  $R^2 = 0.04$ ;  $\chi^2 = 0.01$ ;  $p = 0.01$   
Exponentiated coefficients present odds ratios; z statistics in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## Discussion

This research highlights a significant reliance on support focusing on task direction, even in classrooms that implement a curriculum emphasizing collaboration. Considering that procedural support is closely tied to specific curricula, there is a promising opportunity for AI interventions or AI partners to alleviate the burden of routine

support tasks from teachers. This, in turn, would allow teachers to dedicate more effort to providing higher-level assistance, such as facilitating problem-solving and enhancing collaboration among students.

The MOSAIC protocol helped us identify *lesson activity* as the most significant covariate related to collaboration support. Students received more collaboration support during the jigsaw activity in lesson 4 than during the first programming and wiring task in lesson 2. The complexity of the jigsaw activity, requiring contributions from all team members, likely necessitated increased support for effective collaboration. This activity also provided a structure that assisted teachers in facilitating collaboration, thus fostering positive student interdependence (Johnson & Johnson, 2009). Surprisingly, the regression model revealed that lesson 2 saw students receiving less collaboration support, despite this being their first encounter with a complex form of collaboration (pair programming). This discrepancy may be attributed to the task's complexity; the introductory nature of the programming and wiring tasks might not have sufficiently challenged students to engage their collaborative skills.

Our study has highlighted several focal points and tensions in the quest to better support teachers and students in a collaborative learning environment. Even though our analysis had a rich data set of 618 support moments, our study's focus on a singular curriculum unit within middle school classrooms may limit the generalizability of our findings. Future research should aim to extend this analysis to other curricula and contexts, as well as refine and extend our regression model to encompass post-support activity to understand collaborative support dynamics better. Such insights are vital for developing AI agents that effectively support collaborative learning experiences.

## References

- Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, 12(3), 307–359.
- Barron, B. & Roschelle, J. (2009). Shared cognition. In Anderman, E. (Ed.). *Psychology of Classroom Learning: An Encyclopedia*, pp. 819-823. Detroit, MI: Macmillan Reference USA.
- Bower, M., & Richards, D. (2006). Collaborative learning: Some possibilities and limitations for students and teachers. In *23rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Who's Learning* (pp. 79-89).
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn (Vol. 11). *Washington, DC: National academy press.*
- Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The Role of Collaboration, Computer Use, Learning Environments, and Supporting Strategies in CSCL: A Meta-Analysis. *Review of Educational Research*, 88(6), 799-843. <https://doi-org.colorado.idm.oclc.org/10.3102/0034654318791584>
- Dey, I., Bush, J.B., Hoang, N. (2024) Analyzing Support Moments During Small Group Work. *Paper presented to AERA 2024 Annual Meeting Philadelphia, Pennsylvania, April 11–14, 2024.*
- Gendreau Chakarov, A., Recker, M., Jacobs, J., Van Horne, K., & Sumner, T. (2019). Designing a Middle School Science Curriculum that Integrates Computational Thinking and Sensor Technology. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 818–824. <https://doi.org/10.1145/3287324.3287476>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>.
- Johnson, D. W., & Johnson, R. T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, 38(5), 365–379.
- Mäkitalo-Siegl, K., Kohnle, C., & Fischer, F. (2011). Computer-supported collaborative inquiry learning and classroom scripts: Effects on help-seeking processes and learning outcomes. *Learning and Instruction*, 21(2), 257–266. <https://doi.org/10.1016/j.learninstruc.2010.07.001>
- Molenaar, I., Van Boxtel, C. A., & Sleegers, P. J. (2010). The effects of scaffolding metacognitive activities in small groups. *Computers in Human Behavior*, 26(6), 1727-1738.
- Werner, L. L., Hanks, B., & McDowell, C. (2004). Pair-programming helps female computer science students. *Journal on Educational Resources in Computing (JERIC)*, 4(1), 4.
- van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences*, 28(2), 206-239.
- Vygotsky, L. (1978). Interaction between learning and development. In *Mind and Society* (pp. 79–91). *Harvard University Press.*
- Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). *Macmillan Library Reference USA; Prentice Hall International.*

## Speech-based Automatic Classroom Activity Detection

Bahar Shahrokhian, Arizona State University, bshahrok@asu.edu  
David Zhou, University of Illinois Urbana-Champaign, david23@illinois.edu  
Kurt VanLehn, Arizona State University, Kurt.VanLehn@asu.edu

**Abstract:** Developing an automatic inexpensive, non-intrusive approach that can provide timely feedback to teachers in a noisy co-located classroom remains a difficult task. The first step is to detect class activities robustly in this noisy setting and then build a more fine-grained classifier on top of that. In this work, we trained a machine learning model that takes in speech data captured from low-cost microphones and recorders, while teachers were conducting the class using an orchestration system. We used acoustic, pre-trained representation(acoustic), and contextual features as input to a supervised machine learning algorithm to classify classroom activity into 3 categories (whole-class, group work, and others). The combination of all the features yielded the best result (F1-measure=0.91) on previously unseen data.

### Introduction

For professional development the amount of the time teacher spends with groups vs the whole class is important but providing this kind of feedback relies on human annotation of videos (Chen et al., 2020). The process of human annotation is intrinsically time-consuming and prone to mistakes. Annotation by humans can also be expensive since each coder needs training time, and adding to that, the time that it takes to reconcile any differences between the coders. To address teacher's needs, automatic classroom activity detectors are more practical since they are much cheaper and faster in comparison to human coders.

Such detectors are also essential for intelligent classroom orchestration systems such as our FACT (VanLehn et al., 2021). FACT sometimes sends messages to students while the teacher was addressing the whole class or while the teacher was talking to those students, so FACT needs to understand what the teacher is doing to suppress such messages.

So, the research question is how to automatically classify teacher's activity for real-time orchestration usage and after-class reflection during professional development activities. Here we focused on the acoustic features of teacher's speech because lexical and video can raise privacy concerns.

### Literature review

In recent years, there has been a growing interest in examining teacher's/classroom audio to study instructional practice within in-person or co-located classroom settings. Since this field of research is varied in many aspects, we found it useful to divide these works into 3 groups based on their approach to the annotation scheme.

### Classroom Activity Detection

These works are more focused on analyzing classroom audio, to detect who is talking, the student, teacher, both, or neither(silence) (Li et al., 2019). They usually strategically place the microphone at the front of the classroom to capture what is happening in the class. The coding scheme for most of these works is: "single-voice" (primarily lecture), "multi-voice" (primarily group discussion), "no-voice" (primarily silent work), or "other."

For example, in (Owens et al., 2017), authors developed a system called DART for detecting classroom activity (single-voice, multi-voice, no-voice, and others) using classroom noise levels. They trained and evaluated their binary decision tree using recorded audio in college STEM classrooms. Later Cosbey et al. in (Cosbey et al., 2019) applied deep learning and recurrent neural network approaches to the DART dataset without feature engineering. They reported single voice was the easiest to distinguish, and multi-voice and no-voice categories were more challenging to detect, especially for unseen instructors. They reported a significant (45.1%) frame-level error reduction for new classes and new instructors in comparison with the original DART paper. In (Li et al., 2019), authors used an attention mechanism to fuse acoustic and lexical features to classify classroom audio into teacher's or students' speech. They achieved the F1-measure of 0.87 for the teacher and 0.48 for the students. A comprehensive work by Slyman et al. (Slyman et al., 2021), used a webcam installed at the instructor station of STEM classes in a university to develop a 9-way model for classifying classroom activity. Their annotation scheme included *silence*, *other*, *group work*, *student* (student answer and ask questions), and *instructor* (instructor's announcement, lecture, ask and answer questions). They extracted acoustic features using OpenSMILE (Eyben & Schuller, 2015) and used a pre-trained embedding approach. They also included Mel-filterbanks features. They trained their model using DNN, DTCNN, GRU, and BI-GRU-based networks.

## Teacher Activity Detection

This second group is more focused on the teacher's pedagogical practice. They want to show that machine learning methods can be used to measure different teaching dimensions. Unlike the previous group, they are more focused on what instructional activity the teacher is engaged in. Typically, in these studies, the teacher wears a microphone during their instructional sessions. Their selected annotation schemes are tailored to the instructional activities deemed most relevant for each project. Our work aligns more with this group of studies.

For example, authors in (Wang et al., 2014) used a LENA device to record the teacher's speech. This device identifies each utterance as 1) teacher talk, 2) student talk, 3) overlapping speech, and 4) non-speech (silence or noise). Then two coders independently labeled each 30-second segment Lecture, Discussion, or Group Work. They used context features such as the percentage of teacher/student, overlapping talking time, silence time, and average volume. These features were provided as input to their RF classification algorithm as the independent variables. The classification accuracy for Lecture, Discussion, and Group Work is 0.881, 0.797, and 0.833, respectively, with overall classification accuracy of 84.37%. Authors in (Schlotterbeck et al., 2021), based their annotation scheme on the gross-grain version of COPUS (Smith et al., 2013) which includes Presenting, Administration, and Guiding/feedback. They recorded 4th-grade mathematics teachers using a Lavalier microphone and extracted acoustic, lexical, and contextual features from each 15-second segment, to detect the presence of mentioned teaching practices. They used an attention-based early fusion approach to merge the features. They reported over 0.88 of accuracy and 0.92 of AUC, outperforming acoustic and text-only models. In (Shahrokhian Ghahfarokhi et al., 2020) the authors employed a Lavalier microphone to record two teachers conducting middle school mathematics lessons. Their annotation scheme was tailored to support their orchestration system, involving the segmentation of audio into 'Whole Class,' 'Group Interaction,' and 'Admin' segments of varying lengths. Acoustic and time series features were extracted from each 30-second segment, resulting in a classification accuracy of 0.91 and a kappa value of 0.84. they also used a more fine-grained annotation scheme for their group interaction segments.

## Teacher Discourse Analysis

These works aim to analyze teacher discourse in a more detailed approach. The microphone is usually attached to the teacher and their coding scheme is usually inspired by pedagogical theories. These works may differ considerably based on the research context and pedagogical theories employed. For example, authors in (Donnelly et al., 2017) developed a model to automatically detect teacher's questions in a noisy classroom, based on the theory that questions are at the heart of a dynamic classroom. This work and others set the groundwork for more fine-grained question classifiers (Dale et al., 2022).

## Approach

### Data and human annotation

The data for this trial were collected in two different middle schools in California from June 6th to June 10th, 2022, as part of an evaluation of FACT (VanLehn et al., 2022). A total of five teachers were observed, comprising three females and two males. 20 class sessions were observed, each day, one teacher. Students worked on Chromebooks, one per student, running FACT. From the 4 math lessons supported by FACT, teachers chose different lessons for different classes. All the lessons started with individual and whole class work and then students spent most of the class time working in pairs to solve complex, atypical math problems. While students worked in groups, the teachers circulated carrying a tablet with FACT's dashboard. We recorded the audio and screen of both the teachers and students. The teachers' audio and screen were recorded using a Lavalier microphone plus a portable digital audio recorder and a commercial screen recorder running on their tablets respectively. The files were harvested after class and were saved on a secure machine, to preserve teachers' and possibly students' privacy.

The annotation scheme used is as follows:

- WCT (whole-class-teacher): The teacher is addressing the whole class. If the teacher is engaged in class-level questions and answers, it would still count as WCT. If the teacher is lecturing and then moves on to engaging in class-level QA, we made two different Whole class segments.
- GT (group-talk): The teacher is addressing a group or a student in a group. It also includes any talk with a student about the lesson's math content or FACT's technical issues. If the teacher moves from group to group, we make a new segment. (It is hard sometimes to know if the teacher is talking to a different group or just different students in the group. In this case, we followed our best guess, if no best guess was available, we made a new segment).

- OT (others-talk): The teacher isn't talking to the whole class or any groups or students. Either she's silent or talking to herself or a visitor in a non-distracting way. Anything else would be OT

The inter-rater agreement was considered acceptable with Cohen's kappa  $K = 0.81$ . 15% of the data was annotated by two human annotators and the segments in which they disagreed were discarded. We ended up with 821 labeled segments of different durations. The OT, WCT, and GT segments were 362, 248, and 181 respectively.

## Feature Extraction and training

Three sets of features were extracted: *Feature set 1* comprised acoustic features obtained through OpenSMILE (Eyben & Schuller, 2015) and Mel-scaled filter banks and Mel-Frequency Cepstrum Coefficients (MFCCs). Additionally, contextual features, such as segment duration, segment position relative to others, segment word count, and speech rate (# words in the segment/segment duration), as mentioned in (Donnelly et al., 2017), were incorporated into this feature set. *Feature set 2*) Pre-trained representations were extracted from the 24<sup>th</sup> layer of the Hubert large model (Hsu et al., 2021). Hubert models use a self-supervised approach, inspired by BERT, to learn high-level representation for audio inputs. *Feature set 3*) Combination of feature sets (1) and (2).

All the features were concatenated if needed, standardized, and put through PCA. These features were fed to the classifier models, for them to categorize each audio segment into WCT, GT, or OT. We used about 80% of the data for training classifiers such as KNN, random forest, multi-layer perceptron, additive logistic regression, and SVM. After random search, hyper-parameters for each of the above models were chosen during training using a grid search with 10-fold cross-validation. We achieved the best result using the Random Forest (RF) algorithm with 100 trees.

## Result and discussion

Table 1 compares the RF classifier's test results for all three feature sets. As shown, combining acoustic and pre-trained features yields the best result for predicting all 3 categories.

**Table 1**  
*Classification Performance per Label for all Feature Sets*

	Labels	Precision	Recall	F1-measure
Feature set 1	WCT	0.85	0.92	0.88
	GT	0.75	0.58	0.66
	OT	0.81	0.85	0.83
	Average	0.80	0.78	0.83
Feature set 2	WCT	0.89	0.84	0.86
	GT	0.91	0.81	0.85
	OT	0.84	0.91	0.8
	Average	0.88	0.85	0.83
Feature set3	WCT	0.94	0.92	0.93
	GT	0.91	0.83	0.87
	OT	0.89	0.94	0.91
	Average	<b>0.91</b>	<b>0.90</b>	<b>0.90</b>

Feature sets (2) and (3) performed more robustly across all 3 categories. However, we have to be mindful of the potential demand of extracting pre-trained features for computation resources, especially in typical co-located classrooms. Feature set 1, can be extracted from audio in nearly real-time without requiring extensive resources. However, they did not perform well in detecting group talk (GT). Hence, to build a more detailed classifier for analyzing teacher-student interactions, we need to improve beyond relying on these features for classification. This is one of the reasons we are planning to explore the impact and feasibility of adding multi-modal features that are available to us such as log, position or other privacy preserving data.

Any orchestration system that sends messages to students must ensure those messages are suppressed the students are supposed to be paying attention to the teacher to avoid distracting them. Hence, it is promising that the F1-measure of WCT is high in all 3 feature sets ( $\geq 0.86$ ).

## Contribution and future work

This work not only presents the development and performance of our best classifiers but also sets the groundwork for approaching a more fine-grained classifier for our application. While our initial motivation is to design more effective Orchestration Systems, our research could have implications for in-service teacher



professional development. For example, a system such as ours could perhaps replace the human annotators needed for video-based teacher reflection tools, such as CDA (Chen et al., 2020). This work demonstrates that privacy preserving acoustic features may be adequate for practical use.

One of our main challenges in this research and similar studies lies in the insufficiency of data, including those without annotations. Hence our future work will focus on investigating methods to broaden our dataset. We are planning to incorporate our previous trial data e.g. (Shahrokhian Ghahfarokhi et al., 2020), despite the minor variations in the annotation scheme. Also, using another source of related data such as teaching videos on YouTube could improve the generalizability of our classifiers. These steps could allow us to take advantage of deep learning approaches as well.

## References

- Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of the Learning Sciences*, 29(4–5), 642–680. <https://doi.org/10.1080/10508406.2020.1783269>
- Cosbey, R., Wusterbarth, A., & Hutchinson, B. (2019). Deep Learning for Classroom Activity Detection from Audio. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019-May*, 3727–3731. <https://doi.org/10.1109/ICASSP.2019.8683365>
- Dale, M. E., Godley, A. J., Capello, S. A., Donnelly, P. J., D’Mello, S. K., & Kelly, S. P. (2022). Toward the automated analysis of teacher talk in secondary ELA classrooms. *Teaching and Teacher Education*, 110, 103584.
- Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D’Mello, S. K. (2017). Words matter. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 218–227. <https://doi.org/10.1145/3027385.3027417>
- Eyben, F., & Schuller, B. (2015). openSMILE:). *ACM SIGMultimedia Records*, 6(4), 4–13. <https://doi.org/10.1145/2729095.2729097>
- Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhota, K., Salakhutdinov, R., & Mohamed, A. (2021). Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3451–3460.
- Li, H., Kang, Y., Ding, W., Yang, S., Yang, S., Huang, G. Y., & Liu, Z. (2019). *Multimodal Learning For Classroom Activity Detection*. <http://arxiv.org/abs/1910.13799>
- Owens, M. T., Seidel, S. B., Wong, M., Bejines, T. E., Lietz, S., Perez, J. R., Sit, S., Subedar, Z.-S., Acker, G. N., Akana, S. F., Balukjian, B., Benton, H. P., Blair, J. R., Boaz, S. M., Boyer, K. E., Bram, J. B., Burrus, L. W., Byrd, D. T., Caporale, N., ... Tanner, K. D. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences*, 114(12), 3085–3090. <https://doi.org/10.1073/pnas.1618693114>
- Schlotterbeck, D., Uribe, P., Jiménez, A., Araya, R., van der Molen Moris, J., & Caballero, D. (2021). TARTA: Teacher Activity Recognizer from Transcriptions and Audio (pp. 369–380). [https://doi.org/10.1007/978-3-030-78292-4\\_30](https://doi.org/10.1007/978-3-030-78292-4_30)
- Shahrokhian Ghahfarokhi, B., Sivaraman, A., & VanLehn, K. (2020). Toward an Automatic Speech Classifier for the Teacher. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12164 LNAI. [https://doi.org/10.1007/978-3-030-52240-7\\_51](https://doi.org/10.1007/978-3-030-52240-7_51)
- Slyman, E., Daw, C., Skrabut, M., Usenko, A., & Hutchinson, B. (2021). Fine-Grained Classroom Activity Detection from Audio with Neural Networks. *ArXiv*. <http://arxiv.org/abs/2107.14369>
- Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices. *CBE—Life Sciences Education*, 12(4), 618–627. <https://doi.org/10.1187/cbe.13-08-0154>
- VanLehn, K., Burkhardt, H., Cheema, S., Kang, S., Pead, D., Schoenfeld, A., & Wetzel, J. (2021). Can an orchestration system increase collaborative, productive struggle in teaching-by-eliciting classrooms? *Interactive Learning Environments*, 29(6), 987–1005. <https://doi.org/10.1080/10494820.2019.1616567>
- VanLehn, K., Shahrokhian, B., Viswahnathan, S., & Wetzel, J. (2022). *FACT: An automated teaching assistant for middle school math classrooms*.
- Wang, Z., Pan, X., Miller, K. F., & Cortina, K. S. (2014). Automatic classification of activities in classroom discourse. *Computers & Education*, 78, 115–123. <https://doi.org/10.1016/j.compedu.2014.05.010>

**Acknowledgement** Supported by NSF FW-HTF 1840051.

# Orchestrating Change: A Narrative Analysis of Teacher Practice in a Post-Pandemic, Technology-Enhanced Learning Environment

Yuxin Chen, University of South Alabama, [yuxinchen@southalabama.edu](mailto:yuxinchen@southalabama.edu)

**Abstract:** This study examines the role of reflective practice in a technology-enhanced K-12 environment, focusing on a teacher's orchestration of game-based learning. It investigates how the teacher adjusts curriculum elements and facilitation techniques in response to integrating technology, especially in the context of post-pandemic educational challenges. Employing narrative analysis, the research traces the teacher's journey from professional development to classroom implementation, culminating in a retrospective interview. The study highlights the nuances of instructional strategies and the importance of reflective practice in navigating the complexities of digital learning environments. The findings provide a nuanced understanding of the teacher's instructional practices, highlighting the complex and multifaceted aspects of teaching and learning in a science classroom. This study emphasizes the significance of understanding teachers' orchestration experiences to inform the design and integration of technology-enhanced learning environments and to better support instructors in facilitating effective, adaptive, and engaging learning experiences for their students.

## Introduction

Reflective practices offer a means to explore how teachers integrate technology and adjust practices to assess and refine their instructional approaches. Classroom orchestration is used as a metaphor to describe a teacher's responsibility to manage and regulate a classroom (Dillenbourg, 2013). However, the complexities inherent in technology-enhanced learning environments necessitate more empirical research to understand how teachers orchestrate the multifaceted dimensions within these complex settings. In particular, in the post-pandemic era, teachers encounter additional challenges in reverting to traditional practices. Incorporating reflective practices enables teachers to design, implement, and introspectively examine their instructional strategies, thereby enhancing their ability to orchestrate complex learning environments and adapt to new challenges by critically assessing their methods and interactions with students and technology (Blackley et al., 2017; Gibbs, 1988). The study employs narrative analysis and a case study method to explore a teacher's reflective practices and application of collaborative game-based learning in a K-12 setting. This study aims to explore the complexities of a teacher's orchestration practices, with a particular focus on instruction planning and adjustments during implementation, through two research questions:

- RQ1. How does a teacher plan and adapt curriculum elements to facilitate a collaborative game-based learning environment?
- RQ2. What strategies does a teacher utilize to adjust instructional practices during classroom orchestration?

## Theoretical framework

In collaborative learning environments, teachers' orchestration student learning across various social planes (e.g., individual, group, and whole class) while navigating multiple constraints, such as time and space (Dillenbourg et al., 2013). In recent years, the notion of orchestration has expanded to be on a continuum, ranging from a narrow emphasis on teachers' responsibilities to a more open-ended understanding that involves power and agency, the role of students, technology, and other aspects of a classroom (Sharples, 2013; Schwarz et al., 2018). Researchers recognize the significance of understanding teachers' perspectives, instructional needs, and perceptions of using orchestration tools (e.g., Holstein et al., 2019; Van Leeuwen et al., 2021). However, there is a need to focus on involving teachers in the planning stage, documenting their reflective modifications in orchestration, and highlighting their active adjustment of instructional strategies using technology to boost student learning (e.g., Kollar & Fischer, 2013; Prieto et al., 2011). When integrating new technologies in their classrooms, it's crucial to understand how teachers manage their orchestration process from pre-class design and real-time orchestration to post-reflection. Orchestration practices in the classroom can be more challenging when teachers incorporate collaborative learning. Approaches such as collaborative inquiry, commonly used in STEM education, promote students' critical thinking, creativity, and active engagement (Hmelo-Silver & Jeong, 2022). In particular, the integration of educational games into inquiry-based learning within science classrooms underscores the necessity for teachers to adapt their pedagogical approaches and facilitation strategies (Annetta, 2008). Reflective practice

enables teachers to assess their actions and instructional decisions, revealing the narratives they construct and experience, thereby offering insight into their unique teaching practices (Brookfield, 1995). Reflection-on-action' (Schön, 1983) is a type of reflection where teachers retrospectively analyze and articulate their decision-making after instructional events, often employing techniques such as questioning and thoughtful introspection (Smyth, 1992; Loughran, 2002). Focusing on 'reflection-on-action', this study aims to explore how a K-12 teacher orchestrates game-based learning in science classrooms, examining their reflective practices, instructional experiences, and the integration of the educational game through narrative analysis.

## Methods

### Learning context

Crystal Island: Lost Investigation (Rowe et al., 2011) is a single-player, narrative-driven educational game designed to encourage scientific inquiry centered microbiology. In the game, students assume the role of special agents assigned to investigate an infectious disease affecting a group of scientists on Crystal Island, a remote science outpost. A player is randomly assigned to diagnose either salmonellosis or influenza using one of three contaminated food sources: eggs, milk, or toast. Throughout their investigation, the player interacts with a variety of pedagogical agents in order to learn about the game's narrative, investigate diverse reading materials, and conduct virtual lab experiments. In addition, they utilize a diagnostic worksheet to record their diagnostic process, report their findings, and specify a treatment. Situated in a Midwest charter school, the study centers on Mr. Benson, the lead STEM teacher responsible for teaching all three classes comprising the entire STEM program, accommodating students from 6th to 8th grade. During the period from 2021 to 2022, the program witnessed a notable enrollment increase from 60-70 to 117 students. With eighteen years of STEM education experience, Mr. Benson served as the lead science teacher for 13 years. Upon reaching out to Mr. Benson in Fall 2021, he expressed readiness to support the research, noting students' limited collaboration opportunities due to increased enrollment and the transition from remote to in-person learning. Consent forms were provided to both the teacher, students, and their guardians.

### Study design

The study was divided into three phases: 1) professional development (PD) sessions; 2) implementation of the game unit within the teacher's STEAM program; and 3) retrospective interview sessions. During Phase 1, four PD sessions lasting 30-40 minutes each were conducted to enhance Mr. Benson's skills in integrating the game into the curriculum. Video data captured interactions between Mr. Benson and the researcher, leading to collaborative instructional decisions. In Phase 2, camera operators collected two types of video data in classrooms: student interactions in pairs and footage of Mr. Benson, capturing his orchestration practice during lessons. Phase 3 comprised video-stimulated recall interviews using clips from classroom footage to prompt Mr. Benson's reflection on his actions.

### Narrative analysis

Narrative analysis, a qualitative research approach prioritizing human subjectivity and individual voices, empowers participants to share and interpret their experiences, bridging dualistic boundaries between researchers and participants (Cortazzi, 1994). Participants' accounts offer valuable insights beyond personal storytelling, as they not only reveal the hidden motivations shaping their perspectives and decisions, but also illuminate their underlying values and identity positioning (Dennis, 2016). In this study, narrative analysis was used to interpret, reconstruct, and present Mr. Benson experiences and reflection when orchestrating game-based learning in his classrooms. To ensure validity and reliability, multiple video data sources, including PD sessions, classroom implementation, and the retrospective interview, were utilized to construct Mr. Benson's narrative. PD sessions facilitated discussions on instructional objectives and implementation plans, which were further examined during the semi-structured interview. Classroom video recordings confirmed his facilitation strategies, strengthening the validity of his reflections. Reflexivity played a central role in shaping my research approach. This involved organizing and labeling video data daily, creating content logs to capture key moments, maintaining a researcher's journal for documentation, and engaging in discussions with data collection volunteers to validate and broaden perspectives.

### A narrative about Mr. Benson's orchestration practices

The narrative analysis findings address RQ1 by presenting instructional decisions and respond to RQ2 by reflecting on orchestration actions during implementation, along with relevant recommendations for orchestration.

## Decision-making for orchestrating collaborative learning

In 2021, Mr. Benson transitioned to a STEAM program format for his science classes, accommodating nearly 120 students, double the previous year's population, leading to the division of students into three classrooms. Due to the tightening of Mr. Benson's schedule, there are limited opportunities for engaging in collaborative inquiry projects. Consequently, our collaboration in designing the orchestration aimed to address this issue. In addition, Mr. Benson aimed to reframe the game's original purpose to connect it with the pandemic, leveraging students' interest in health and well-being, in alignment with the curriculum theme of 'humanity.' As a result, it was decided that students would work in pairs to play the game, promoting collaborative inquiry, while being seated in groups of four to ensure peer or cross-group support. To familiarize students with pair work and facilitate learning of basic scientific terminology and microscope skills, Mr. Benson prepared them by conducting experiments with Petri dishes to cultivate bacterial samples in pairs before implementation. Mr. Benson emphasized that a key outcome he envisioned for students through the game was the ability to transfer their newly acquired scientific terminology to real-world situations reflected in the game (Pedaste et al., 2015). He wanted them to understand practical applications of science, pondering questions like, "Where are microscopes being used? Who is using them? Why are those tools and methods of science being used to solve real problems?"

Another decision involved the design of physical workbooks. These workbooks serve as paper-based scaffolds, providing structure, guiding problem identification and hypothesis generation, and facilitating active participation in the inquiry process. Later in the interview, Mr. Benson noted the instrumental role of physical workbooks in enabling students to write, discuss, and generate ideas, expressing their significant utility in understanding students' learning processes and planning his instruction accordingly.

## Adaptive instruction during orchestration

The section outlines three scenarios in which Mrs. Benson adjusted her teaching methods to address the challenges that arose during classroom orchestration. In the first few days, Mr. Benson decided to redirect the use of Accountable Talk (Michaels et al., 2008) in pair discussion. Accountable Talk was introduced to students as a principal method for sharing, exchanging ideas, and engaging in reasoned arguments and discussions. However, students noted that the provided sentence starters sounded too formal and structured. A major reason is that students who were affiliated with the elementary school in the same system were already accustomed to similar discussion norms, such as 'I want to build off of...' and 'Could you say more...'. Thus, Mr. Benson encouraged students to adapt Accountable Talk in their own way. However, he also noted that Accountable Talk sentence starters were particularly useful for students who encountered communication challenges, proving especially beneficial for students with Individualized Education Programs or social needs to enhance their learning and participation in discussions.

A second scenario showcasing Mr. Benson's adaptive instruction was highlighted through his support for a specific student: Alex and Jason. During a check-in, Mr. Benson fostered Alex's understanding of the human body's response to various pathogens by incorporating the student's personal experience with COVID vaccines into the discussion. Inspired by this engaging encounter, I prompted Mr. Benson to elaborate more, leading to his reflection on his educational philosophy and instructional strategies. Mr. Benson, motivated by a 'power of respect' training he had attended, adopted the concept of 'disrupting education' to challenge and transform traditional educational norms, thereby promoting student ownership as a means to boost learning. For Alex, who faced challenges with a learning disability and memorization, Mr. Benson reasoned, "if I could get him to talk about his own experience and teach me, he would understand it a lot better."

The third scenario occurred while supporting a pair of students, Liam and Peter. They approached Mr. Benson with a complex question about cultivating viruses in a Petri dish. Mr. Benson openly admitted, 'I don't know that' encouraging them to research independently. Recalling the interaction, he noted, "it was a moment where I wanted them to see that there is a lot to learn. We do not have all the answers, but we can use tools to find those answers." After learning that a virus can grow in a Petri dish with a cell culture, Liam and Peter realized the necessity of living host cells for viral replication and presented their investigation to the class at Mr. Benson's request. Mr. Benson also pointed out that external resources, including internet searches, can enhance classroom activities and encourage student-led inquiry beyond the predefined instructional design.

## Redesigning instruction for enhanced orchestration

During the reflection, Mr. Benson looks forward, contemplating how he might refine his instruction to better his orchestration. Mr. Benson expresses his ambition to employ an orchestration tool that would enable him to track student progress and assess their knowledge of game-based learning material. In line with Mr. Benson's attention

to groups and students who need help, he envisioned a dashboard tool that could swiftly identify learners with similar struggles, offering insights for tailored group formation. This feature would enable Mr. Benson to quickly reorganize students into new groups for collaborative problem-solving or to pair them with advanced peers who can serve as peer tutors. Regrouping is particularly crucial for short-term projects where immediate adjustments are more practical.

## Conclusion

Understanding teachers' narratives can unravel the dynamics of technology-enhanced learning environments. Through the lens of Mr. Benson's experiences, this study provides a window into the lived experience of teaching in technology-enhanced environments. Utilizing narrative analysis, I explore his trajectory from setting goals to classroom implementation and reflective practice. Teachers play a crucial role in leading student-centered, technology-advanced learning environments. This study extends beyond the current literature focus on facilitation and orchestration tools, offering an in-depth look at a teacher's reflective journey in orchestrating K-12 classrooms. It sheds light on how teachers, like Mr. Benson, navigate the integration of new technologies, adapt their curriculum, and refine instructional strategies amid the evolving landscape of post-pandemic education. By focusing on the nuances of a teacher's journey in implementing game-based learning, this study offers insights into the dynamic orchestration required for effective technology integration.

## References

- Annetta, L. A. (2008). Video games in education: Why they should be used and how they are being used. *Theory into practice*, 47(3), 229-239.
- Blackley, S., Sheffield, R., Maynard, N., Koul, R., & Walker, R. (2017). Makerspace and reflective practice: Advancing pre-service teachers in STEM education. *Australian Journal of Teacher Education (Online)*, 42(3), 22-37.
- Brookfield, S. (1995). The getting of wisdom: What critically reflective teaching is and why it's important. *Becoming a critically reflective teacher*, 1-28.
- Cortazzi, M. (1994). Narrative analysis. *Language teaching*, 27(3), 157-170.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485-492.
- Dennis, B. (2016). Myths of monoculturalism: narratively claiming the other. *International Journal of Qualitative Studies in Education*, 29(9), 1069-1085.
- Gibbs, G. (1988). Learning by doing: A guide to teaching and learning methods. *Further Education Unit*.
- Hmelo-Silver, C. E., & Jeong, H. (2022). Synergies Among the Pillars. In *Handbook of Open, Distance and Digital Education* (pp. 1-16). Springer Singapore.
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-Designing a real-time classroom orchestration tool to support teacher-AI complementarity. *Journal of Learning Analytics*, 6(2), 27-52.
- Loughran, J. J. (2002). Effective reflective practice: In search of meaning in learning about teaching. *Journal of teacher education*, 53(1), 33-43.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in philosophy and education*, 27, 283-297.
- Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., ... & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47-61.
- Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, 21(1-2), 115-133.
- Schon, D. A. (1983). *The reflective practitioner: How professionals think in action* (p. 1983). New York: Basic Books.
- Schwarz, B. B., Prusak, N., Swidan, O., Livny, A., Gal, K., & Segal, A. (2018). Orchestrating the emergence of conceptual learning: A case study in a geometry class. *International Journal of Computer-Supported Collaborative Learning*, 13(2), 189-211.
- Sharples, M. (2013). Shared orchestration within and beyond the classroom. *Computers & Education*, 69, 504-506.
- Smyth, J. (1992). Teachers' work and the politics of reflection. *American educational research journal*, 29(2).
- Van Leeuwen, A., Knoop-van Campen, C. A. N., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12. *Journal of Learning Analytics*, 8(2), 6-21.

# Toward Productive Multivocality in AI Development: Excavating Ethics Concerns among an Interdisciplinary Team

Grace (Yaxin) Xing, University at Buffalo, yaxinxin@buffalo.edu  
Sanaz Ahmadzadeh Siyahrood, Georgia Institute of Technology, ssiyahrood3@gatech.edu  
X. Christine Wang, University at Buffalo, wangxc@buffalo.edu  
Jinjun Xiong, University at Buffalo, jinjun@buffalo.edu

**Abstract:** Adapting the *productive multivocality* framework (Suthers et al., 2013), we engage an interdisciplinary team, which designs AI tools for children with speech and language challenges, in developing guidelines for ethical AI development and deployment. In the initial phase of this 5-year study, we investigated ethical concerns about AI in general and the AI tools being designed in particular across the varied discipline teams. Employing thematic analysis, descriptive statistics, BERT topic modeling, and WordCloud tool, we analyzed survey data and focus group interviews of 13 researchers in Speech-Language Pathology/Learning Sciences, Human-Computer Interaction, Multimodality, and Core Technology. Our analysis uncovered prevalent unease about AI, along with nuanced and varying degrees of concerns regarding the AI tools under development across different disciplines. The findings inform our broader study and highlight the potential role of productive multivocality in fostering responsible and equitable AI development and its eventual implementation.

## Introduction

As artificial intelligence (AI) increasingly permeates our daily lives, the imperative for ethical AI development and deployment becomes paramount, especially in educational contexts (The White House, 2023). While various ethical guidelines for AI exist, researchers challenge the efficacy of current AI ethics tools, pushing for methods rooted in best practices from different sectors to combat structural injustices (Ayling & Chapman, 2022; Heilinger, 2022). To address this gap, we are conducting an ethical study by following the IEEE Standard 7000-2020 (Olszewska, 2020) to establish comprehensive ethical guidelines for the tool development and deployment at the National Science Foundation (NSF)/Institute of Education Sciences (IES) jointly funded AI Institute for Exceptional Education (hereafter, *AI4ExceptionalEd*; <https://www.buffalo.edu/ai4exceptionaled.html>).

Our study adapts the *productive multivocality* framework (Suthers et al., 2013). Rooted in Bakhtin's (1981) seminal work on multivocality, the framework promotes knowledge expansion through interdisciplinary collaboration by transforming individual disciplinary contributions into a cohesive, interdisciplinary synthesis (Oshima et al., 2015). We apply its five core principles to help us move toward productive multivocality: valuing diverse perspectives, promoting open dialogue, engaging in reflective practice, co-constructing knowledge, and respecting complexity (Suthers et al., 2013). This paper concerns the first phase of our study, which focuses on excavating ethical concerns among the *AI4ExceptionalEd*'s interdisciplinary team and investigates three overarching questions: (1) What are ethical concerns about AI in general across the team members' disciplines? (2) What are ethical concerns about the team's targeted AI tools across the team members' disciplines? and (3) How are these two types of ethical concerns related to each other?

## Methods

Employing a case study approach, we investigated the complex ethical considerations confronting the interdisciplinary team at *AI4ExceptionalEd*. Supported by NSF and IES, *AI4ExceptionalEd* aims to mitigate the exacerbated shortage of Speech and Language Pathologists (SLPs) due to the COVID-19 pandemic by creating two advanced AI technology suites: the AI Screener and the AI Orchestrator. The Institute's unique focus on AI tools for young children, SLPs, and educators as end-users, combined with the diverse expertise of a large interdisciplinary team, makes it a "telling case" (Mitchell, 1984, p. 239) for examining responsible and equitable AI development and deployment. The ethical study's first phase began in September 2023, five months after the Institute's launch, capitalizing on a moment when team members' varying perspectives were most pronounced due to early cross-disciplinary interactions and collaborations.

## Participants

We recruited participants from four key disciplinary fields (SLP/Learning Sciences, HCI, Multimodality, and Core Technology) at *AI4ExceptionalEd*. Fourteen agreed to participate in individual surveys, and 13 continued

participating in the focus group interviews. The largest group among the participants was the SLP/Learning Sciences (LS) with six members, followed by HCI and Core Technology fields, each with three participants, and the smallest group was Multimodal with just one researcher.

## Data collection and analysis

Data was collected through individual surveys and focus group interviews organized within disciplines. The survey consisted of five open-ended questions about their experiences, attitudes, and concerns about AI in general. Structured interview questions focused on ethical concerns related to the AI Screener or/and the AI Orchestrator depending on which one(s) participants focused on. The interviews were conducted via Zoom and the length ranged from 20 to 90 minutes.

Survey data was organized in an Excel sheet, while interview recordings were transcribed and analyzed from CSV files after thorough data cleaning. This preparation involved standardizing text by removing irrelevant characters and converting it to lowercase for accurate word frequency analysis. We integrated thematic analysis, descriptive statistics, BERT topic modeling, and WordCloud tool in comprehensive qualitative and quantitative analysis of the survey and interview data. While the thematic analysis was conducted to identify recurring and divergent themes, descriptive statistics were used to show the general patterns of responses. BERT topic modeling was used to visualize key themes and topics within the data. Additionally, WordCloud was used to represent keyword frequency and conceptual relationships.

## Findings and discussion

### AI experiences, attitudes, and ethical concerns

There were varying experiences with AI across teams. The Core Technology team had the most extensive experiences and knowledge (e.g., “working on several topics in deep learning” “working on a fundamental theory for AI” and “decades of experience”). Although the Multimodality and the HCI teams had a similar amount of AI experiences as the Core Technology team, their work focused more on human-centered AI applications (e.g., “computer-child interaction” “social robotics” and “detecting cognitive states of learners”). The SLP/LS team had the least AI experiences, positioning themselves more as domain experts rather than technologists.

Individually, the participants' attitudes toward AI varied widely ranging from ambivalence to optimism. About 35% of participants had mixed feelings, recognizing both AI's strengths and weaknesses; 21% were neutral, viewing AI as a tool or a "superpower baby" shaped by humans; and 43% were positive, stressing the potential benefits of AI. Collectively, the Core Technology team was the most optimistic about AI's potential, followed by the HCI and Multimodality teams while the SLP/LS team was decidedly more mixed about AI's potential.

Connecting participants' AI experiences and their attitudes toward AI showed that individuals (mostly SLP/LS team members) new to AI adopted a balanced stance being “open” toward its “positive” effects while also concerned about its “negative” outcomes or potential “dangers.” But they often overlooked human agency and their abilities to address some of the problems. In contrast, those with rich AI experience (e.g., Core Technology, HCI, Multimodality teams) recognized AI as a tool that is controlled and reshaped by humans. They exhibited a pragmatic optimism, marveling at what AI could achieve, yet being aware of its limitations and social responsibilities. The enhanced familiarity with AI is associated with a more nuanced perspective that acknowledges the technology's promise alongside its ethical and societal implications, which is consistent with Ehsan et al.'s (2021) findings. This underscores the imperative of incorporating a broad spectrum of multidisciplinary expertise and multivocality in the development of AI to ensure a well-rounded and informed approach.

Regarding ethical concerns about AI in general, the most frequently cited concerns were “Disinformation and Misuse” (29%) and “Lack of Understanding of AI's Limitations” (29%), highlighting the risks of misinformation and overestimation of AI's capabilities. Other concerns included “Ethical and Legal Aspects” (21%), “Lack of Human Touch and Social Skills” (14%), “Over-Reliance of AI” (14%), and “Bias and Discrimination” (14%), pointing to the need for ethical and legal frameworks in AI regulation.

### Ethical concerns about AI screener/AI orchestrator

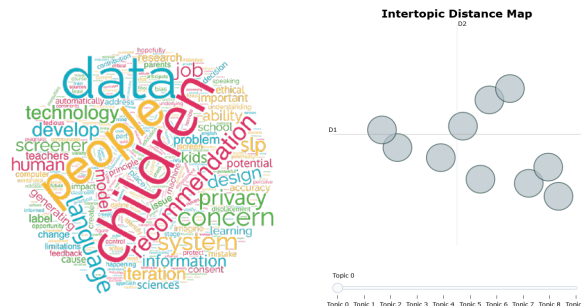
Our analysis of participants' potential ethical concerns about the AI Screener/AI Orchestrator revealed a collective agreement on fundamental issues as well as distinct insights reflecting their discipline-associated viewpoints.

#### Common themes

The WordCloud in Figure 1 prominently features terms like “children,” “people,” “data,” “recommendation,” “language,” “concerns,” “privacy,” and “system.” The accompanying Intertopic Distance Map suggests a thematic

interconnectedness among these subjects. Also, each circle representing a topic has the same size. This suggests that all important topics were given equal importance or consideration in the analysis.

**Figure 1**  
*Word Cloud and BERT Topic Modeling Results*



A deeper analysis of the interview transcripts yields four prevalent themes: (1) Data Privacy, Consent, and Ownership: All teams emphasized data privacy and the secure handling of sensitive information, e.g., young children's multimodal data (e.g., facial, behavioral, vocal, and semantic information). They raised questions about the extent and nature of informed consent in classrooms where not all may consent, as well as data access, ownership, and rights as children grow older. They were also concerned about the potential harm of AI rigidly labeling or categorizing children; (2) Human Oversight and Final Responsibility: There was a consensus across teams that AI should support, instead of replacing, decision-making by professionals, who should maintain final authority and accountability; (3) Transparency and Educating about AI's Limitations: Participants across teams advocated clear communication about AI's functionality and limitations, pushing for critical assessment of AI tools by end-users as well as educating public to prevent misunderstandings; and (4) Equity, and Fairness: The SLP/LS and Multimodality teams consistently expressed concerns about AI's potential to address or exacerbate inequalities. They stressed the need to ensure equitable access for marginalized groups and minimize bias, especially in language and speech assessments of bilingual and multilingual children.

Although different teams shared similar concerns or "voices," their contributions or "productivity" in addressing these concerns could vary significantly. For example, while the SLP/LS team and the Multimodality team both focused on equity and fairness, the former might provide insights into the pedagogical and practical implications, whereas the latter might offer more specific technological solutions. This variation is crucial as it allows for a more integrated approach to AI development and deployment.

### Distinct themes

The Core Technology team emphasized AI's environmental impact and suggested smaller computing models to minimize carbon footprint. They also stressed the unrealistic *expectations* people often have of AI, emphasizing that individuals should understand that AI, much like *humans*, can make *mistakes*. The HCI team was optimistic about the AI Screener/AI Orchestrator's role in transforming SLP's jobs but discussed concerns about potential "job displacement" or SLP's "over-reliance" on these tools. The Multimodality team warned about privacy concerns and misuse of AI by educators, while the SLP/LS team stressed the need for safeguards against unintended AI outcomes in language education, advocating for features that give users more control and flexibility. They also highlighted the importance of equitable AI integration in schools.

The multivocality approach underscores that the distinct themes raised by each team are not merely separate strands but integral parts of a complex mosaic, reflecting the multifaceted nature of AI's implications. For instance, the Core Technology team's emphasis on the environmental impact of AI operations resonated with broader ethical considerations, intersecting with the HCI team's focus on job displacement and societal values, as both teams were essentially considering the long-term implications of AI on society and the planet. Similarly, the SLP/LS team's call for user-empowering features like a "redo" button dovetailed with the Multimodality team's apprehensions about data misuse, as both advocated for greater control and ethical oversight.

### **Connections between concerns about AI in general and the specific AI tools**

The AI Screener and AI Orchestrator, designed to assist SLPs and young children with speech and language challenges, bring to the fore specific ethical concerns that resonate with the broader discourse on AI. Issues such as ethical issues, disinformation, and a lack of understanding about AI's limits are not unique to this tool but are part of a wider conversation about responsible AI deployment. For example, the risk of misinterpretation by



teachers and parents using the AI Screener's results could lead to unintended consequences, highlighting the need for clear communication and education about AI's capabilities and limitations.

However, these tools also offer chances to alleviate apprehensions about AI in general. The AI Screener and AI Orchestrator aim to augment, not supplant, SLP expertise, assuaging fears of job loss or diminished human skills. By automating mundane tasks, they could allow SLPs to focus on more complex cases, potentially improving intervention quality. Involvement in AI creation can dispel some of the mystique surrounding the technology, fostering a sense of agency and comprehension that might lessen ethical worries. As one participant shared, "My concerns are lessened by my involvement in the creation process; understanding its mechanics provides reassurance." Nonetheless, it's vital to avoid becoming complacent, assuming we need only worry about general AI ethical issues and not those specific to one's own team's creations.

## Conclusions and implications

The discourse surrounding AI, especially when involving vulnerable groups like young children, is multi-layered. Ethical concerns articulated by interdisciplinary teams encompass a spectrum of issues, including disinformation, misuse, and lack of understanding of AI's limitations, underscoring the urgency for robust ethical and legal frameworks. Shared themes such as data privacy, informed consent, human oversight, and equity are juxtaposed with unique concerns like environmental sustainability, potential job displacement, and the amplification of biases within the AI Screener/Orchestrator framework. These general and specific ethical issues are interrelated, as the broader concerns about AI's societal impact are echoed in the nuanced apprehensions regarding the AI Screener/Orchestrator, highlighting the importance of integrating diverse expertise to navigate the ethical landscape of AI development and application.

Based on the findings, we strongly advocate for increased transparency in AI development, workforce training, proactive bias reduction, stringent data protection, efforts to reduce AI's ecological footprint, and increasing public's understanding of AI's capabilities and boundaries. Such a comprehensive strategy is critical for AI's ethical and responsible deployment in schools. The implications of this study highlight the essential role of multivocality in AI development, advocating for a collaborative, interdisciplinary approach that embraces diverse ethical perspectives to ensure responsible and equitable integration of AI technologies.

## References

- Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? *AI Ethics*, 2, 405–429. <https://doi.org/10.1007/s43681-021-00084-x>
- Bakhtin, M. (1981). Discourse in the novel. In M. Holquist (Ed.), *The Dialogic Imagination* (pp. 259-422). Austin: University of Texas.
- Ehsan, U., Passi, S., Liao, Q. V., Chan, L., Lee, I., Muller, M., & Riedl, M. O. (2021). The who in explainable AI: How AI background shapes perceptions of AI explanations. ACM CHI. <https://doi.org/10.48550/arXiv.2107.1350>
- Heilinger, J. C. (2022). The ethics of AI ethics. A constructive critique. *Philosophy & Technology*, 35(3), 61. <https://doi.org/10.1007/s11023-020-09517-8>
- Mitchell, J. (1984). Typicality and the case study. In R. Ellen (Ed.), *Ethnographic Research: A guide to general conduct* (pp. 237 - 241). London: Academic Press.
- Olszewska, J. I. (2020). IEEE recommended practice for assessing the impact of autonomous and intelligent systems on human wellbeing. *IEEE Standard*, 7010-2020. <https://doi.org/10.1109/IEEESTD.2020.9084219>
- Oshima, J., Oshima, R., & Fujita, W. (2015). A multivocality approach to epistemic agency in collaborative learning. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, S. Ludvigsen (Eds), *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL '2015), Exploring the Material Conditions of Learning* (Vol. 1, pp. 62–69). Gothenburg, Sweden: International Society of the Learning Sciences. <https://doi.org/10.22318/cscl2015.146>
- Suthers, D. D., Lund, K., Rosé, C. P., & Teplov, C., (2013). Achieving productive multivocality in the analysis of group interactions. In *Productive multivocality in the analysis of group interactions* (pp. 577-612). Boston, MA: Springer.
- The White House. (2023, October 30). *Fact sheet: President Biden issues executive order on safe, secure, and trustworthy artificial intelligence*. The White House.

## Acknowledgments

We thank all the participating members of the AI4ExceptionalEd team for their time and support for this project.

# Symposia



## Perspectives from the Field: Scaffolding Peer Critique and Feedback

Christian Schunn, University of Pittsburgh, schunn@pitt.edu  
Omid Noroozi, Wageningen University and Research, omid.noroozi@wur.nl  
Wenli Chen, Nanyang Technological University, wenli.chen@nie.edu.sg  
Qianru Lyu, Nanyang Technological University, nie20.lq@e.ntu.edu.sg  
Siew Cheng (Aileen) Chai, Nanyang Technological University, aileen.chai@nie.edu.sg  
Xinyi Li, Beijing Normal University, l\_xinyi@126.com  
Justice Walker, University of Texas at El Paso, jtwalker@utep.edu  
Sayed Mohsin Reza, Pennsylvania State University Harrisburg, skr6024@psu.edu  
Alex Acquah, University of Texas at El Paso, aacquah@miners.utep.edu  
Andi Scarola, University of Texas at El Paso, adscarola@miners.utep.edu  
Amanda Barany, University of Pennsylvania, amanda.barany@gmail.com  
J. Bryan Henderson, Arizona State University, jbryanh@asu.edu  
Annabel Stoler, Boston University, ajstoler@bu.edu  
Eve Manz, Boston University, eimanz@bu.edu  
Jinzhi Zhou (Co-chair), Indiana University, zhoujinz@iu.edu  
Cindy E. Hmelo-Silver, Indiana University, chmelosi@indiana.edu  
Danielle Murphy (Co-chair), Rutgers University, dm880@scarletmail.rutgers.edu.  
Zachary Ryan, Indiana University, zryan@iu.edu  
Christina Stiso, Indiana University, cstiso@indiana.edu  
Qiuyu Lin, Rutgers University, qiuyu.lin@rutgers.edu  
Joshua Danish, Indiana University, jdanish@indiana.edu  
Ravit Golan Duncan, Rutgers University, ravit.duncan@gse.rutgers.edu  
Clark A. Chinn, Rutgers University, clark.chinn@gse.rutgers.edu  
Leslie Rupert Herrenkohl (Discussant), University of Michigan, leslierh@umich.edu

**Abstract:** This symposium serves as a platform for in-depth discussions about how we scaffold students in giving and receiving peer critique and feedback. While peer critique is recognized as an effective practice for learning in diverse disciplines, it poses challenges for both students and teachers. This interactive session will discuss the complexities of peer critique in various contexts, scaffolds that we employed to address these challenges, and lessons learned.

### Introduction

Peer learning, as a broad concept, and peer critique, in particular, are considered effective learning practices (Noroozi & De Wever, 2023). Such practices provide opportunities for students to engage in vetting, revising, and refining their scientific reasoning, as well as improving their performance in writing (Yu & Schunn, 2023). Ford (2008) argued that science knowledge is advanced through a dialectic between scientists who construct claims and the community that critiques them. Thus, critique at the communal level is crucial to the production of well-defended knowledge, as well as helping learners value this social aspect of science knowledge building (Ford, 2008). Empirical research has revealed that peer feedback facilitates students' collective construction of accurate knowledge in scientific inquiry, empowers students to develop epistemic agency, and helps them gain new perspectives on their own work (Tasker & Herrenkohl, 2016; González-Howard & McNeill, 2020). However, there has not yet been substantial uptake of this vital practice of critique in the process of knowledge construction in science education, despite calls to emphasize critique more strongly in research and instruction (González-Howard & McNeill, 2020).

In contrast to science education, the practice of peer feedback is well accepted in some disciplines, particularly for science writing (Noroozi & De Wever, 2023). Studies have explored learning benefits from providing peer feedback (Yu & Schunn, 2023), different types of peer feedback (Wu & Schunn, 2023), and students' epistemic beliefs towards peer feedback (Noroozi, 2022). Despite the wealth of insights derived from studies of peer feedback in writing, it remains challenging for students to engage in the practice productively.

To bridge these gaps, research that explores how to productively scaffold students' engagement in peer critique and feedback is critical. More attention is needed to address ways to support students in providing high-quality peer feedback. Although existing interventions do engage students in critique, students need more help in

using the critique to improve their work (Danish et al., 2021). This symposium brings in scholars from diverse disciplines, including science, data science, and writing to delve into the intricacies of peer critique and feedback. We will discuss challenges that teachers and students face with the practice, including questions of relationality and risk. We will share effective scaffolds, with a particular emphasis on the role of technological tools and different strategies that enable critical meta-epistemic understandings and adeptness in the practice.

The symposium will be organized in a structured poster format, with the following agenda: 1) 2 mins introduction and framing by the chair; 2) 1 min each for firehose presentations of the seven presenters (7 mins total); 3) 30 mins for the posters (with three 10 min cycles); 4) commentary by our discussant Leslie Rupert Herrenkohl (10 mins); 5) general Q & A with the audience (11 mins). Presenters will address the following questions:

- 1) In what context are you engaging students in peer critique?
- 2) What challenges do students face with giving and receiving peer critique?
- 3) What scaffolds do you provide?
- 4) What are the big lessons you have been learning about engaging students in peer critique?

## **Paper 1: Lessons learned from scaffolding at scale within SWORD/Peerceptiv**

Christian Schunn

*Contexts of peer critique.* Peer critique has been quickly expanding in research and practice, heavily supported by broad availability of laptops and internet access, particularly at the university level, but also by significant growth in widely-used web-based systems that implement a structured approach to peer critique. These broadly used web-based systems presents new opportunities for research, particularly in studying both typical practice and the impact of web-based peer critique on various aspects of student engagement and learning across a very broad range of contexts and contents. Schunn and colleagues have taken advantage of available data from SWORD/Peerceptiv, which spans over one hundred institutions around the world (some secondary, many undergraduate, and some graduate courses), hundreds of courses (in a wide range of disciplines), thousands of assignments, and hundreds of thousands of students.

*Challenges in giving and receiving.* On the giving side, students rarely provide incorrect information, but they often avoid commenting upon issues for which they are uncertain. As a result, with respect to disciplinary knowledge, they prefer to discuss problems of communication (i.e., in terms of what they did not understand) rather than weaknesses of disciplinary content (i.e., in terms of what is correct). This pattern in turn means that the disciplinary content issues that most students struggle with are also the issues that fewest students address in their comments. On the receiving side, students are more likely to make revisions to their arguments when they receive feedback from multiple peers and they all comment on the same issue. Thus, multi-peer feedback is better than single peer feedback. However, receiving a large amount of feedback (a natural consequence of multi-peer feedback) sometimes produces improvements in later student performance, but sometimes produces reductions in later student performance, with a net mean effect of zero. Students need to have a certain amount of peer feedback literacy to know how to produce effective feedback and how to benefit from peer feedback.

*Scaffolds provided.* The SWORD/Peerceptiv system scaffolds critique in multiple ways. First, the critique process is highly structured: comments are interleaved with ratings, ratings are anchored with detailed descriptors, comments are given detailed comment prompts. Second, reviews are randomly assigned and critique is double-blinded, enabling honest critique. Third, reviewers are held accountable to providing accurate ratings and helpful feedback through grading algorithms and a feedback-on-critique step, the latter being particularly important.

*Lessons learned.* First and foremost, providing feedback is the strongest learning opportunity, as it is a kind of constructive learning task, whereas receiving feedback is often a kind of passive learning. It is possible to do even better by arranging for interactive learning via peer critique, but it requires going beyond the typical setup of online peer critique (e.g., collaborative critiquing or multiple rounds of interaction between author and reviewer). Second, multiple forms of scaffolding are important to encourage in-depth engagement in providing feedback. Third, random-assignment generally works well, but learning can be further optimized by matching author-reviewer pairs, for both motivational and cognitive reasons.

## **Paper 2: Scaffolding essay writing with argumentative peer critique scripting**

Omid Noroozi

*Contexts of peer critique.* Peer learning, as a broad concept, and peer critique, in particular, are regarded as vital instructional approaches that have the potential to enhance students' learning processes and outcomes within

educational environments (Noroozi & De Waver, 2023). In the past decade, we have integrated scaffolded peer learning methodologies, encompassing peer critique, peer feedback, peer assessment, peer review, peer interaction, peer dialogue, and more, into our practical educational settings. We have especially leveraged technological advancements to develop online argumentative peer critique scripting, which effectively directs students toward engaging in constructive and argumentative peer interactions. This, in turn, has proven instrumental in aiding students in attaining their desired learning outcomes. Additionally, our research has delved into the impact of these argumentative peer critique scripting on a diverse array of learning processes and outcomes, specifically in the context of essay writing across various domains within the life sciences.

*Challenges in giving and receiving.* Delivering effective peer critique encompasses several crucial elements, including offering affective statements (e.g., praise or compliments), pinpointing the identified problem(s), detailing the arguments and reasoning behind these identified problems, and proposing constructive solutions to tackle the identified issues (i.e. feedforward). However, when providing feedback, students frequently omit the arguments and reasoning behind the identified problems due to a variety of psychological, emotional, epistemological, and cognitive factors. This can result in an emphasis on aspects that may not be beneficial for their peers' learning. On the receiving side, a significant challenge is the limited uptake of feedback and the neglect of peers' comments. Several factors may contribute to this, including a lack of thorough reflection of the feedback comments, mistrust in peers' knowledge, as well as the lack of justified arguments of the feedback received. This suggests that students must possess a degree of peer learning literacy with emphasis on argumentation and reasoning, enabling them to both provide justified feedback and evaluate and reflect the feedback received through argumentative analysis.

*Scaffolds provided.* The design of our online argumentative peer critique scripting offers scaffolds in a multifaceted manner: To start, the critique process is intricately structured, featuring explicit question prompts that cover various elements of the argumentative essay. Additionally, it incorporates multiple randomly selected peer raters with diverse domain-specific knowledge for each essay, addressing concerns related to trust within the peer critique process. Lastly, students are compelled to furnish justifications for each identified issue within their peers' essays.

*Lessons learned.* Primarily, our online approach to scripting argumentative peer critique has proven effective in enhancing students' peer feedback processes, ultimately resulting in improved argumentative essays. We can further enhance the process by motivating students to engage in thoughtful reflection and analysis of the feedback they receive, and then nurturing dialogical interactions between feedback providers and recipients. Our aim is not solely to offer support for addressing the immediate argumentative essay tasks at hand (first-order scaffolding: tools for living) but also to enhance students' peer learning literacy to self-regulate their peer learning experiences, enabling them to tackle new and complex challenges in similar situations (second-order scaffolding: tools for learning). Achieving this necessitates features such as contingency, gradual reduction of support (fading), and the transfer of responsibility.

### **Paper 3: Promoting peer feedback literacy for secondary schoolers: Design-based research from Singapore**

Wenli Chen, Qianru Lyu and Aileen Chai and Xinyi Li

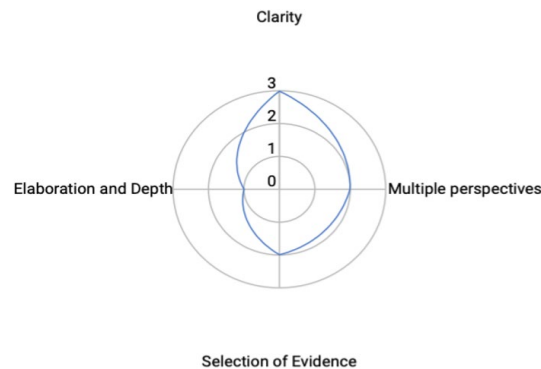
*Contexts of peer critique.* Critical thinking, communication and collaboration are important 21<sup>st</sup> Century competencies for Singapore K-12 students. We introduce peer feedback activities into Singapore secondary classrooms and teachers encouraged students to provide constructive feedback for peers and support possible improvements (Tan & Chen, 2022). How to promote students' peer feedback literacy in everyday classroom settings becomes a key question for us. Design-based research is applied to fine-tune the design of peer feedback activities.

*Challenges of giving and receiving.* Secondary schoolers face challenges in both giving and implementing peer feedback. The challenge during giving peer feedback lies in understanding the reviewed content. Students reported that they were often unsure about what the content conveyed, which made it difficult for them to provide evaluation (positive or negative), not to mention providing suggestions for improvements. The challenge during receiving feedback includes emotional feelings (confused, hurt, frustrated), understanding the feedback as well as acting upon the feedback. Some students said that they felt hurt, confused and frustrated when receiving questions or negative evaluations. They also had difficulties decoding the feedback comment, which directly led to obstacles to acting upon feedback as the receivers.

*Scaffolds provided.* We try to address the challenges in both peer feedback giving and receiving stages. For feedback giving, we provided scoring rubrics and sentence openers to scaffold the construction of peer feedback content. The scoring rubrics provide clear definitions of how low, middle and high performance looks

like in every scoring dimension. Students are asked to provide quantitative evaluation and qualitative feedback for each dimension (see Figure 1). Meanwhile, the qualitative feedback is facilitated by the sentence openers adapted from the Perkins (2003), encouraging students to identify their strengths and weaknesses (e.g., “I appreciate...”; “It’s not clear that...”), ask for clarification (e.g., “What do you mean by...”) and suggest possible improvements (e.g., “You may modify...”).

**Figure 1**  
*Quantitative Evaluation*



For feedback receiving, we invite students to reflect on the feedback and work on act-upon plans, informed by the feedback literacy model (Carless & Boud, 2018). Students are encouraged to reflect on the feedback in four aspects: “Our understanding of the feedback is...”, “We agree/disagree that...”, and “We want to clarify/discuss that...”. In the end, students should think forward and plan their actions: “Our plan of action is...”, “We need to add/contribute...”, and “We should learn more about...”. This reflection process allows students to have more chances to make judgments, manage affect and finally, take action.

*Lessons learned.* Providing flexible scaffolding is needed for real-world peer feedback practices. During the design-based research practice, the researchers and school teachers have been closely collaborating to maximize the effect of peer feedback activities and iteratively fine-tune the scaffolding design. We learned that an effective scaffolding for real-world teaching and learning should have its characteristics of a low floor, high ceiling and wide wall (Chen et al., 2021). Another lesson to learn is that peer feedback is a complex multi-faced practice that is not limited to cognitive activity but also social and emotional activities. Engaging students in peer feedback should take into consideration their “readiness” in various aspects.

## **Paper 4: Sandboxing data science: Tensions in peer critique to support epistemic agency in computing-based teaching and learning**

Justice Walker, Sayed Mohsin Reza, Alex Acquah, Andi Scarola and Amanda Barany

*Contexts of peer critique.* As data science continues to make significant inroads in pre-college (K-12) education (Jiang et al., 2022), there are a growing number of questions about how to spur productive participation and learning in these areas, and especially among intellectually and demographically diverse learner groups. In computing education, data science has shown promise in supporting such outcomes as domain mastery and literacy in activities that emphasize epistemic practice (Schanzer et al., 2022). While efforts along this frame have been important in bridging increasingly relevant genres of computing and technology to learning environments before college, much of what we know about these areas depend on activities and data sources that are highly curated—thus undermining the potential for learners to engage in acts that are agentic and situated in authentic contexts (Barany et al., 2023). This is often because data science and computing are complex domains and it can be difficult for educators to support epistemic agency along lines of inquiry that vary in connection with individual learners and their myriad interests, cultural histories and/or sociopolitical concerns. In our project, we characterize this approach as sandboxing in data science and it involves learners using python in the Google Colab environment to data mine (or scrape) large sets of data from the popular social media platform, Twitter or, more recently, X in order to answer research questions they developed along their personal interests (e.g., soccer star popularity), cultural histories (e.g., language use in STEM public discourse), and/or sociopolitical concerns (e.g., public sentiment about animal testing). Peer critique is a promising frame within which to support sandboxing in computing-based data science. This is because peer critique provides a sustainable way to encourage knowledge building couched in authentic epistemic practices. In this NSF-funded project, Coding Like a Data Miner

(CLDM), we share exploratory insights on a knowledge building activity—referred here as the “*debugging circle*” and report on the ways learners provided and used peer critique on inquiry projects that varied in terms of the research questions pursued, epistemological strategies and bugs encountered.

*Challenges in giving and receiving.* Our early observations suggest that for open-ended inquiry activities that involve intersecting disciplines (in our case computing and data science), peer critique can be a viable method for supporting epistemic agency among learners and in connection to their varying values and priorities. We also observed that peer feedback reveals ontological and interpersonal tensions that exist in learning environments where there are several plausible paths for success along an inquiry (e.g., competing feedback that is conducive to success and/or social dynamics that privilege or subordinate ideas independent of their merits). These two observations create tensions that, if unmitigated, can exacerbate persistent issues of equity regarding who gets to contribute to domain knowledge, and whose epistemic orientations are valued.

*Scaffolds provided.* Our debugging circle is a knowledge building activity meant to give study participants a chance to showcase their data science project progress and get feedback about their implementation strategies (e.g., parsing out complex qualitative data from tweets) and technical challenges (e.g., how to debug code to create a data visualization). The debugging circle involved learners using a shared GoogleSheet to populate computer screenshots and descriptions of issues they encountered and then discussing collectively how they might solve identified problems. The exercises were facilitated by instructors who prompted and modeled clarifying questions and feedback.

*Lessons learned.* Open inquiry brings along with it a whole host of pedagogical (for teachers) and epistemic (for learners) challenges. In CLDM, we were able to observe moments where scaffolds can be conducive to peer feedback and knowledge building. Accompanying these supports were also tensions, including the various ways learners select and privileged approaches when there are multiple solutions to a goal or problem, and the social dynamics that accompany whose ideas are taken up, or not. Both these observations have significant implications in shaping learner participation in knowledge building, and subsequent outcomes. We discuss these tensions in relation to broader issues of equity and empowerment in computing and data science fields that emphasize epistemic agency.

## **Paper 5: PAL: A Utilization of educational technology to scaffold peer critique**

J. Bryan Henderson

*Contexts of peer critique.* I have built a scaffold for peer critique called *PAL Mode* that is part of a larger digital student response platform that I have been developing for the last decade called *Braincandy* (Henderson & Lewis, 2020). The scaffold guides students through specific steps to navigate peer disagreement. Research suggests that middle school is a time that presents serious challenges to students’ interest and motivation to engage in science (Carlone et al., 2014). This makes middle school an important context for data collection, especially when considering that it will likely take time for students to develop habits of mind that value critique and epistemic vigilance. I regularly attend middle school classrooms where I provide instructional coaching during class time on how to use the Braincandy PAL Mode to promote classroom critique.

*Challenges in giving and receiving.* Student likelihood to talk in class is often tied to the confidence they have in possessing “correct” answers. Hence, a challenge in promoting student willingness to critique lies in the enculturation of student mindsets where speaking out loud is relevant to all stages of learning, and not just reserved for when there is confidence that thinking is in its “final form.” Meanwhile, a challenge in the receiving of critique is a propensity for students to interpret critique as something more than just the specific idea that is being questioned.

*Scaffolds provided.* PAL Mode for the Braincandy platform utilizes *Rapoport’s Rules of Argument* (Popova, 2014) to scaffold student talk moves as they argue with others. Under PAL Mode, a series of four answer boxes will appear on student Braincandy screens. Each successive answer box will not open until the previous answer is submitted. The first three answer box prompts for PAL Mode are currently worded as so:

[Paraphrase] **In your own words, summarize your discussion partner’s position.**

[Align] **What are things your position and your discussion partner’s position have in common?**

[Learn] **What are things you might have learned from your discussion partner? If you don’t feel you’ve learned anything new, ask your partner a question about something you’re confused about.**

Once these three preliminary questions are answered, a fourth answer box then appears to allow students to submit their final argument into the Braincandy system. The prompt for this final argument is: Now, considering where you might agree with your discussion partner and what you might have learned from your discussion partner, please explain your current position AND explain why other positions are less preferable. The purpose of this



final prompt is encouraging not only construction of why students think their argument is correct, but also critique of why certain alternative ideas made privy to them through PAL Mode were not compelling.

*Lessons learned.* Even when students encounter differing interpretations, their propensity is to focus on the interpretation they find preferable, as opposed to why they find the alternative interpretations less preferable (Henderson et al., 2014). Hence, the purpose of the final PAL Mode prompt is to encourage not only construction of why students think their argument is correct, but also critique of why certain alternative ideas made privy to them through PAL Mode were not as compelling. In other words, it can be beneficial to provide students repeated reminders to not only articulate why they think right answers are right, but also articulate why wrong answers are wrong.

## Paper 6: Critique and risk in elementary science modeling activity

Annabel Stoler and Eve Manz

*Context of peer critique.* Our work is situated in a larger project focused on incorporating uncertainty into elementary children's modeling and investigation work, with the aim of supporting more meaningful and conceptually productive science practice around modeling and investigation in classroom science. We conceptualize science knowledge as a communal accomplishment developed through cycles of construction, critique, and stabilization of models and their relations, driven by uncertainty in those relations. In classroom instruction that we co-design with teachers, students often present and discuss ideas using explanatory models (e.g., drawing a model that uses molecules to explain why a bottle of water breaks when the water freezes). Modeling is used for purposes that require evaluation and critique: to establish disagreement, to develop puzzles to orient further work, to revise ideas based on what has been learned, and to put pieces together to develop consensus explanations. From our perspective, modeling, and the evaluative activity associated with it, involves both *epistemic* and *relational* activity. By *epistemic*, we mean the ways in which students are positioned with opportunities to construct knowledge and make decisions about how to use this knowledge (Miller et al., 2018). By *relational*, we refer to the ways that children relate to each other as individuals and as part of a learning community (Krist & Suárez, 2018).

*Challenges in giving and receiving.* We have found that both students and teachers experience challenges with the critique involved in community knowledge development. Students' epistemic challenges include engaging with another person's idea and understanding their thinking (rather than stating one's own idea), focusing evaluation on the explanatory features of the model, and revising models in light of new information or critique. Relational challenges include a sense of risk in publicly sharing ideas and the negotiation of status. Further, *teachers* may find critique and disagreement risky and may sense tensions in their epistemic and relational goals, since they feel a need to elicit student ideas in ways that promote a safe and supportive classroom culture. Student status is often inequitably distributed in the classroom, such as when some students are positioned with greater authority in class discussions. It can be challenging for teachers to incorporate opportunities for peer feedback that disrupt inequitable structures and provide opportunities for all learners to engage in the practice of modeling and critique.

*Scaffolds provided.* Our co-design team of teachers and researchers have tried several strategies to support students' epistemic and relational work during modeling. We focus on how scaffolds can shape activity productively, make important aspects explicit, and provide students with tools to draw on, *without* overly narrowing students' heterogeneous ideas and ways of expression. Scaffolds include explicitly framing purposes and activities of modeling conversations (e.g., to understand others' ideas and develop questions for further work) and providing sentence frames. Further, teachers have implemented and explored different structures for student critique, such as providing an anonymous or teacher-constructed model, or listing similarities and differences without evaluation.

*Lessons learned.* Evaluation is a key part of engaging with others' ideas for communal knowledge development, but brings challenges and tensions. When teachers attempted to minimize opportunities for students to engage in model evaluation in order to mitigate risk felt by students, epistemic engagement and co-construction of ideas decreased, and, often, teachers felt they needed to step in to do epistemic work relational management. We have identified several strategies that support students to evaluate models in ways that attune to relational concerns. These include explicitly framing models as tools to collectively think with (rather than to evaluate one person's idea), highlighting how models can change over time as students gather more information, and the teacher positioning herself as a co-thinker alongside students. In the context of these and other strategies, we have observed students supporting each other's epistemic and relational interactions, such as by revoicing a peer's idea, drawing attention to parts of the model that they learned something from, trying to mitigate each other's risk, and crediting each other for helping them deepen their understanding. Our developing focus is how teachers and

students can work with critique as an activity within an overall enterprise of community knowledge development among people who value each other— where students are engaged in “thinking-with” models and people (de la Bellacasa, 2017).

## Paper 7: Engaging students in gallery walks within MEME

Jinzhi Zhou, Cindy E. Hmelo-Silver, Danielle Murphy, Zachary Ryan, Christina Stiso, Qiuyu Lin, Joshua Danish, Ravit Golan Duncan and Clark A. Chinn

*Contexts of peer critique.* It is important for students to recognize the value of critique in model refinement and revision (Ford, 2008). However, students need support and practice in giving and receiving critique, to fully appreciate its role in enhancing modeling practices. The project *Scaffolding Explanations and Epistemic Development for Systems* (SEEDS) designed and investigated an instructional approach with integrated software tool that scaffolded learners as they developed, critiqued, and refined scientific models using evidence (Danish et al., 2021). The software tool, called Model and Evidence Mapping Environment (MEME), scaffolds students’ use of evidence and epistemic ideals (criteria) when creating and revising models. While students understand the importance of making partial models more accurate or complete, they often need motivation to revise their models. In one study, we implemented a “gallery walk” activity in a 5<sup>th</sup>/6<sup>th</sup> grade classroom. In small groups, students used MEME to provide feedback on other groups’ models, then responded to their peers’ feedback. This activity helped students apply criteria to evaluate models, identify opportunities for improving their models, and make their ideas about modeling and evidence explicit.

*Challenges in giving and receiving.* When giving peer critique, students need support in using criteria consistently and on a regular basis to evaluate models. Students may privilege their preexisting idiosyncratic beliefs surrounding criteria for good models. Moreover, students need scaffolding in learning to provide constructive peer critiques (Tasker & Herrenkohl, 2016). It remains challenging for students to develop the ability to provide critical and specific feedback on peers’ models within a short period of time. In addressing peer critique, students do not effectively leverage general or ambiguous feedback. They need additional support and opportunities to communicate with peer reviewers to better understand the critiques.

*Scaffolds provided.* MEME incorporated a comment function to scaffold students’ peer critique. Within MEME, students accessed peers’ models and provided comments on any specific components (e.g., entity, process, outcome) using the “comment box”. They could select the most relevant criterion for their comment from a menu that includes options aligned with the class-shared criteria for model evaluation. The criteria, developed and discussed with students, played a central role in model creation and evaluation. Moreover, students were provided with guidelines for what counts as constructive critique. They were encouraged to show appreciation for peer work, be critical and positive, and provide specific suggestions. When addressing peer critique, students could carefully review each comment pertaining to specific components and make revisions as needed. The MEME “comment box” also featured an option “question or response,” enabling students to seek clarifications from their peers when necessary.

*Lessons learned.* First, peer critique is productive for model revision. Our study highlighted the significance of giving and receiving peer critique in fostering collaborative criteria-based model evaluation, iterative model revision, and productive group discussions about modeling. Second, promoting the development and use of collective epistemic criteria among students is critical for evaluating peers’ models effectively. MEME features and other instructional materials scaffolded students’ adoption of criteria for giving peer critique. However, we need to design further scaffolds to encourage students to provide more critical and specific comments to promote model revision.

## References

- Barany, A., Reza, S., Johnson, M., Barrera, A., Badreddin, O., Fuentes, C., & Walker, J. T. (2023). Towards the design of a culturally relevant curriculum for equitable, data mining-based CS education. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023*, pp. 1498-1501. International Society of the Learning Sciences.
- Calabrese Barton, A., & Tan, E. (2019). Designing for Rightful Presence in STEM: The Role of Making Present Practices. *Journal of the Learning Sciences*, 28(4–5), 616–658.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325.
- Carlone, H. B., Scott, C. M., & Lowder, C. (2014). Becoming (less) scientific: A longitudinal study of students’ identity work from elementary to middle school science. *Journal of Research in Science Teaching*, 51(7), 836–869.

- Chen, W., Tan, J. S. H., & Pi, Z. (2021). The spiral model of collaborative knowledge improvement: an exploratory study of a networked collaborative classroom. *International Journal of Computer-Supported Collaborative Learning*.
- Danish, J., Vickery, M., Duncan, R., Ryan, Z., Stiso, C., Zhou, J., Murphy, D., Hmelo-Silver, C., Chinn, C. (2021). Scientific model evaluation during a gallery walk. *Proceedings of the 15<sup>th</sup> International Conference of the Learning Sciences – ICLS 2021*. International Society of the Learning Sciences.
- de La Bellacasa, M. P. (2017). *Matters of care: Speculative ethics in more than human worlds* (Vol. 41). University of Minnesota Press.
- Ford, M. (2008). ‘Grasp of practice’ as a reasoning resource for inquiry and nature of science understanding. *Science & Education*, 17(2-3), 147-177.
- González-Howard, M., & McNeill, K. L. (2020). Acting with epistemic agency: Characterizing student critique during argumentation discussions. *Science Education*, 104(6), 953-982.
- Henderson, J. B., & Lewis, A. (2020). Integrating Technology: Integrating Technology to Support Classroom Argumentation. *Science Scope*, 43(8), 16-19.
- Henderson, J. B., Osborne, J., MacPherson, A., & Szu, E. (2014). A new learning progression for student argumentation in scientific contexts. *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research for Evidence-Based Teaching and Coherence in Learning*, 7, 26–42.
- Jiang, S., Lee, V. R., & Rosenberg, J. M. (2022). Data science education across the disciplines: Underexamined opportunities for K-12 innovation. *British Journal of Educational Technology*, 53(5), 1073-1079.
- Krist, S., & Suárez, E. (2018). Doing science with fidelity to persons: Instantiations of caring participation in science practices. *Proceedings of the 13th International Conference of the Learning Sciences*, 1, 424–431.
- Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching*, 55(7), 1053–1075.
- Noroozi, O. (2022). The role of students’ epistemic beliefs for their argumentation performance in higher education. *Innovations in Education and Teaching International*. 1–12.
- Noroozi, O. & De Wever, B. (Eds.). (2023). *The power of peer learning: Fostering students’ learning processes and outcomes*. Cham: Springer International Publishing.
- Perkins, D. (2003). *King Arthur's round table: How collaborative conversations create smart organizations*. John Wiley & Sons.
- Popova, M. (2014, March 28). How to Criticize with Kindness: Philosopher Daniel Dennett on the Four Steps to Arguing Intelligently. Retrieved October 17, 2019, from Brain Pickings website: <https://www.brainpickings.org/2014/03/28/daniel-dennett-rapoport-rules-criticism/>
- Schanzer, E., Pfenning, N., Denny, F., Dooman, S., Politz, J. G. Lerner, B. S., Fisler, K. & Krishnamurthi, S. (2022). Integrated data science for secondary schools: Design and assessment of a curriculum. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1* (pp. 22-28).
- Tan, J. S. H., & Chen, W. (2022). Peer feedback to support collaborative knowledge improvement: What kind of feedback feed-forward? *Computers & Education*.
- Tasker, T. Q., & Herrenkohl, L. R. (2016). Using peer feedback to improve students’ scientific inquiry. *Journal of Science Teacher Education*, 27(1), 35–59.
- Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology*, 73, 102160.
- Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, 147, 107857.

# Stepping Outside the Ivory Tower: How Can We Implement Multimodal Learning Analytics in Ecological Settings, and Turn Complex Temporal Data Sources into Actionable Insights?

Bertrand Schneider (chair), Harvard University, [bertrand\\_schneider@gse.harvard.edu](mailto:bertrand_schneider@gse.harvard.edu)

Richard Davis, Swiss Federal Institute of Technology, [richard.davis@epfl.ch](mailto:richard.davis@epfl.ch)

Roberto Martinez-Maldonado, Monash University, [roberto.martinezmaldonado@monash.edu](mailto:roberto.martinezmaldonado@monash.edu)

Gautam Biswas, Vanderbilt University, [gautam.biswas@vanderbilt.edu](mailto:gautam.biswas@vanderbilt.edu)

Marcelo Worsley, Northwestern University, [marcelo.worsley@northwestern.edu](mailto:marcelo.worsley@northwestern.edu)

Nikol Rummel (discussant), Center of Advanced Internet Studies (CAIS) gGmbH and Ruhr - University Bochum, [nikol.rummel@rub.de](mailto:nikol.rummel@rub.de)

**Abstract:** The field of Multimodal Learning Analytics (MMLA) is expanding, allowing researchers to capture rich, fine-grained data on learning processes in a variety of learning environments. High-quality process data can open the door for new insights on how people learn, creative interventions to support them, and build the foundation for personalized learning platforms. There is, however, a growing recognition that there is a lack of ecological implementations in MMLA (Cukurova, Giannakos & Martinez-Maldonado, 2020): the vast majority of projects are lab-based, which limit the generalizability and impact of multimodal sensing in education. This symposium brings together researchers who have used MMLA methods in the wild. The panel will discuss obstacles to the use of multimodal data in real-world settings, share lessons learned from current projects, and propose productive next steps for the field to become more ecologically relevant.

## Introduction

Multimodal Learning Analytics (MMLA; Blikstein, 2013) is a field of research that involves analyzing and interpreting data from diverse sources to understand and improve the process of learning and instruction. The term "multimodal" refers to the various modes of communication (such as speech, text, eye-movement, gestures, facial expressions, etc.) and interaction (like physical artifacts, technological interfaces, etc.) that are taken into account. The aim of MMLA is to provide meaningful insights into the learning process by integrating and examining data from multiple dimensions. It seeks to understand how different modalities contribute to learning, how they intersect and influence each other; and helps in designing effective learning environments. It applies various methods and techniques such as machine learning, data mining, and artificial intelligence to process, organize, and interpret complex data. MMLA has substantial implications for personalizing learning experiences, improving teaching methodologies, providing real-time feedback, and promoting successful learning outcomes.

Researchers, however, are recognizing that there is a need for more ecological validity and impact from MMLA: "there is a clear need for further work in the implementation of MMLA systems in authentic spaces where learning occurs (e.g., homes, classrooms, museums), an endeavor that is already acknowledged as challenging (Baker, Ocumpaugh, & Calvo, 2015). It is clear that this line of work has not yet reached its full capacity, and proper in situ setups hold the potential to bridge data quality and ecological validation..." (Cukurova, Giannakos & Martinez-Maldonado, 2020). In a literature review on the scalability of MMLA, Yan, Zhao, Gasevic and Martinez-Maldonado (2022) found that more than half (51%) studies were conducted in laboratory settings. In a related field (Multimodal Collaboration Analytics, MMCA), Schneider et al. (submitted) have reviewed 147 studies that have used multimodal sensing to capture collaborative processes. They found that only 24 (16%) of them took place in ecological settings.

In short, there is growing evidence that MMLA can be helpful in capturing learning processes in controlled environments; but there is a need to generalize these results to practice. This symposium brings together five perspectives to discuss the challenges of using MMLA in ecological settings. Each researcher presents a project below, with lessons learned and proposed solutions to facilitate the use of multimodal data in education. The symposium will discuss a wide range of issues, such as data privacy, ethics, validity, setting up complex data collection pipelines, data fusion (i.e., synchronizing and integrating different data streams), data analysis (i.e., finding signal in the noise), the role of theory in MMLA, replication in educational research, and more. The outcome of the symposium is to draft a preliminary list of Grand Challenges in MMLA and a research agenda to address them.

## Contribution #1: Multimodal learning analytics to support learning and teaching in constructionist learning environments

Bertrand Schneider

### Project description

Constructionist learning environments, such as makerspaces and digital fabrication labs, are ideal training grounds for cultivating 21st century skills. These environments are inherently student-centered and project-based, through tasks that carry real-world relevance. However, the very nature of such open-ended spaces, where each student follows a unique learning trajectory, presents significant challenges, making the measurement of learning progress complex. To tackle this issue, we have instrumented our makerspace with eight cameras capturing both pose and gaze data (Figure 1; the left image shows the 3D pose and gaze data mapped onto a floorplan of the makerspace; the right side shows the field of view of a camera where the same two students are collaborating while another one is working at the laser cutter). This approach has generated millions of observations, offering an alternative glimpse into students' learning processes and interactions. Our current challenge is to make sense of this rich dataset and use it to enhance both learning and teaching. We are considering different data-driven interventions to support learners and teachers. This includes the development of a dashboard (Guillain & Schneider, 2021), which, despite its potential usefulness, has been underutilized, as well as explorations in the use of GPT-based feedback systems (Sung, Guillain & Schneider, 2022). These efforts aim to translate that data into practical tools and feedback that can improve teaching and improve students' learning outcomes using multimodal sensing data.

**Figure 1**

*The Makerspace is Equipped with 8 High Resolution Cameras used to Extract Multimodal Data.*



### Challenges

We had to overcome several challenges during this project. The first was to capture accurate data from the makerspace. We first tried to build our own platform based on Microsoft Kinect sensors. Unfortunately, cleaning the data took a significant amount of time because we had to connect different datasets, remove duplicates, connect tracks, and manually clean up corner cases. We then moved to a self-contained package (opentrack.org), which is now deprecated and resulted in a dataset that was too noisy to be usable. Finally, we partnered with the Montessori Wildflower schools and used their 3D reconstruction algorithm, which worked well in our setting ([github.com/WildflowerSchools/poseconnect](https://github.com/WildflowerSchools/poseconnect)).

The next major challenge was to persuade the institutional review board (IRB) and students that decreasing data privacy would result in new insights and increased learning outcomes. We were careful to frame this project as a way to improve the quality and frequency of formative feedback for learners and improve data-driven decision making for teachers. We specified that the data would never be used for summative assessment, or impact students' learning experience. Additionally, we did not collect audio data to avoid a situation where students would feel spied on. We devised an opt-out procedure, where students could ask us to discard their data. Finally, we showed them the final anonymized dataset (i.e., the stick figures on the left side of Figure 1). This resulted in an environment where students were comfortable having their data recorded.

Once the data is generated, however, you need to analyze it and make sense of the results. Using theory (Wise & Shaffer, 2015) helped us generate meaningful metrics and validate them (Chng, Seyam, Yao & Schneider, 2022). A final challenge was to use quantitative methods on such a small dataset (~20 students). To produce meaningful results, we had to run the same course 2-3 times over 2-3 years. This generated a dataset of 40-60 students, which took considerably longer to collect than some other types of studies.

## Lessons learned

Even though this project is ongoing, we learned several lessons that apply to other learning environments and MMLA implementations: 1) if possible, find partners who are working on the same problem (instead of reinventing the wheel); 2) work with students to understand their level of comfort with data collection tools, and which measures need to be taken to create a safe and trusted learning environment; 3) avoid collecting data that impacts privacy (e.g., speech), especially if it's not crucial for answering research questions; 4) as much as possible, adopt participatory design methodologies to create useful platforms for learners and teachers (e.g., Guillain & Schneider, 2021); 5) don't underestimate the technical know-how and infrastructure required to collect and process multimodal data; 6) focus on formative (and not summative) assessment; 7) use theory to guide data analysis (e.g., Chng, Seyam, Yao & Schneider, 2022); 8) replicate results across different cohorts, especially when working with small sample sizes.

## Contribution #2: Obstacles facing adoption of (MM)LA for formative assessment in higher education

Richard Lee Davis

### Project description

What will it take to bring MMLA out of the research world and into practice? What obstacles and challenges should we expect to face? We have explored these questions by carrying out a series of qualitative studies investigating the adoption (or lack thereof) of learning analytics tools in higher education. In interviews with instructors and teaching assistants we have uncovered a set of needs that indicate that there is a place for learning analytics tools in classroom practice. At the same time, we have identified a number of concerns and constraints that help explain why adoption in practice has been low, and which suggest that the path to adoption is substantially more difficult than previously acknowledged. We have advanced a theoretical framework to support the adoption of LA in practice and anticipate how these insights might inform efforts to utilize MMLA for formative assessment in higher education.

### Challenges, lessons learned, next steps

We have identified obstacles to both adoption and use of LA in higher education. To help make sense of the obstacles to adoption, we have developed the TACT framework (Technology Adoption Costs and Tolerances) (Davis et al., 2023) which theorizes that teachers' willingness to adopt new LA technologies is a function of two things: the adoption costs of a technology and the teacher's tolerances to those costs. When an LA tool fits with a teacher's existing practices and meets perceived needs, tolerances to adoption costs are high. Otherwise, tolerances, and prospects for adoption, are low.

MMLA systems are likely to impose very high adoption costs because they typically require new forms of classroom instrumentation and generate novel types of data. To ensure that teachers are willing to tolerate these costs, it is essential for MMLA systems to mesh with teachers' existing practices and meet their real needs. At least initially, this will restrict uses of MMLA to specific types of learning environments (open-ended spaces) and to specific types of teachers (those already using formative assessment in their teaching). Even when these conditions are met, care must be taken to ensure that adoption costs are as low as possible. An advantage of human-centered methods, like those used by Martinez-Maldonado in the development of AIAugmentTeam, is that they help keep costs low.

Even after adoption, obstacles to the use of MMLA in higher education remain. Teachers in our interviews expressed concerns about how introducing LA tools into exercise sessions might degrade environments of trust and safety that they had worked hard to cultivate. At least in the context of our study, students were perceived as placing high value on rights to anonymity and privacy, and teachers worried that LA tools would be perceived as violating these rights (Cai et al., 2023). These concerns are likely to be exacerbated by the multiple sources of data collected by MMLA systems. As both Schneider and Martinez-Maldonado explain, care must be taken to ensure that the data collection process is transparent, that students have the option to opt-out of data collection, and that the benefits of these systems are clearly communicated.

### Next steps for the project and for the field

Co-designing MMLA tools with teachers and students provides the clearest path to use in classrooms, as it helps break down obstacles to both adoption and use. By tailoring the tools to a specific context and set of needs, adoption costs are lowered and tolerances to those costs are raised. And by building the tools together with stakeholders, fears about malicious uses of data can be defused since the design and implementation is made

transparent through the co-design process. This makes co-design ideal for development and small-scale deployment of MMLA tools in higher education.

An important open question is whether co-design methods can be extended to handle medium- and large-scale applications of MMLA for formative assessment, or whether a different approach is needed. When the number of potential users grows into the hundreds, or even thousands, traditional co-design approaches are no longer viable due to the amount of time and care needed. There is a need to identify strategies that bring the same benefits of co-design, otherwise adoption of MMLA for formative assessment will suffer.

### Contribution #3: AIAugmentTeam: Multimodal teamwork analytics in immersive healthcare simulation

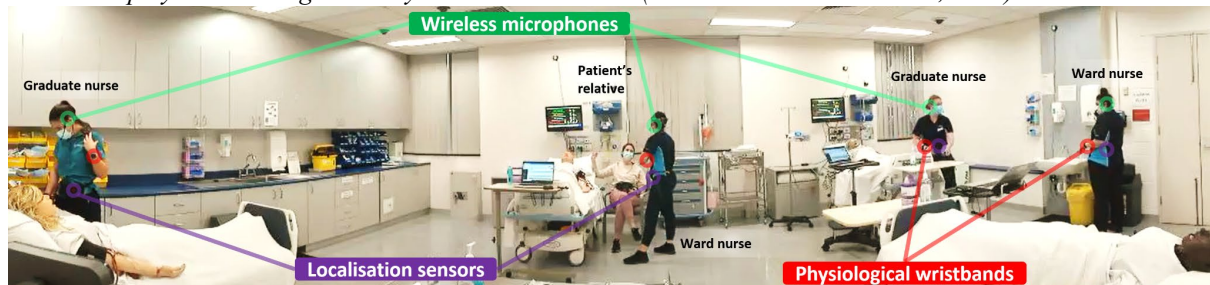
Roberto Martinez-Maldonado

#### Project description

Advancements in MMLA and Generative AI (GenAI) are revolutionizing how we understand and improve collaborative learning among students. These technologies significantly enhance our ability to support the development of teamwork skills and the reflective practices of both students and teachers, especially in situations where learning is not necessarily mediated by computers. However, currently, only a few MMLA tools offer practical feedback to students and teachers to aid in this reflection (Yan et al., 2022). **AIAugmentTeam** is an MMLA platform designed to give actionable feedback on team interactions. The system includes: (i) *a data capture platform* that works with various sensors (like microphones and position trackers), physiological wristbands, and teacher annotation tools; (ii) *a data analysis system* that turns raw data into meaningful insights, including using GenAI to automate transcription and analyze team dialogues; and (iii) *human-centered interfaces* for teachers to give augmented feedback during team sessions. To date, the tool is currently tailored for immersive team simulations within healthcare settings and has been employed by 620 students and 18 teachers in real-world classrooms from 2021 to 2023. In 2023 alone, five teachers have adopted it as a regular analytics tool. According to the most up-to-date MMLA literature reviews (Yan et al., 2022), our implementation represents the most extensive MMLA study to date that completes the learning analytics loop by offering students and their teachers direct, group-based feedback via MMLA-enabled visual interfaces.

**Figure 2**

*Sensors Deployed in the High-Fidelity Medical Simulation (Martinez-Maldonado et al., 2023).*



#### Lessons learned

We synthesized a set of lessons learned from a large human-centered MMLA study conducted in-the-wild (i.e., a deployment that is as naturalistic as possible) in the context of nursing education (e.g., see Figure 2). This study took place over three years, with three key phases. The first focused on data collection only, the second on using an MMLA dashboard for classroom reflection, and the third on enhancing teachers' reflection and leadership with a co-designed orchestration/analytics tool. Our lessons learned have been detailed elsewhere (Martinez-Maldonado et al., 2023) and are summarized as following: 1. *Human-Centered Design, Teaching and Learning*: Teachers partnering with researchers in the design process of MMLA systems leads to better alignment with teaching practices and learning goals. 2. *Human-Centered MMLA and Research Innovation*: Involving teachers and students in the design process helps validate MMLA interfaces and improves the logistics of MMLA research studies. 3. *Consenting and Participation Strategies*: Explaining complex MMLA studies to students in person rather than providing excessive technical details about sensors and analytics helps in gaining informed consent. 4. *Data Privacy and Sharing*: Students are willing to share their multimodal data for learning purposes if their privacy is ensured. Some see the benefit in making their data available for others' learning or for teachers to improve learning tasks. 5. *Technological Sustainability*: A lightweight microservices-based architecture that

allows for easy attachment and detachment of various sensors can enhance long-term technical sustainability. 6. *MMLA Appropriation in the Classroom*: Embedding sensing capabilities in the classroom, empowering users, training teachers in system usage and data interpretation, and minimizing the need for technical support can maximize the adoption and effective use of MMLA technology.

### Suggestions for the field

Based on our first-hand experiences in deploying MMLA in-the-wild, we have synthesized the following suggestions for researchers and practitioners. 1. *Collaborative Design with Teachers and Students*: Effective use of sensor data in education requires close collaboration between teachers, students, and technology developers. This collaboration can ensure that the data and technology align with educational goals and teaching methods. Involving teachers and students in designing these systems helps address practical challenges and makes the data more meaningful and useful in real classroom settings. 2. *Acknowledging Data Limitations and Empowering Teachers*: Data from sensors can be imperfect or incomplete. MMLA systems should avoid making automatic decisions based on this data. Teachers need control over these systems and should be informed about the reliability of the data. This also highlights the importance of training teachers to understand and use this technology effectively. 3. *Prioritizing Safety and Privacy*: Introducing advanced technology in classrooms raises privacy and surveillance concerns. Teachers and students should be aware of how their data might be used and have control over it. Guidelines for data privacy and user consent are crucial, especially for sensitive information. Systems should be designed to allow users to manage their own data, including the option to delete it after educational use.

## Contribution #4: Multimodal learning analytics in embodied learning environments

Gautam Biswas

### Project description

Embodied Learning builds on the demonstrated value of play or game-based learning in supporting the learning of domain knowledge and collaboration processes. In embodied learning, students are immersed in a mixed-reality environment, and this allows them to playfully explore science phenomena, such as the rules of particle behavior in solid, liquid and gas and the photosynthesis process through collective embodied activity (Tu, et al, 2019). Frameworks for analyzing embodied cognition, such as the Learning in Embodied Activity Framework (LEAF) framework have developed methods that account for collective activity without erasing and replacing the individual's role as part of the collective. LEAF supports the synthesis across individual and sociocultural theories of embodiment and thus provides a more robust account of how the body can play a role in both individual and collective cognition and learning (Danish, et al, 2020).

Currently, research teams use a combination of interaction analysis and qualitative coding of teacher and student interactions to examine patterns in the learning processes during the embodied play activities (Davis, et al, 2019). To support our embodied learning research team, we have now deployed our multimodal learning analytics (MMLA) pipeline to facilitate data collection from multiple cameras and microphones, posyx data for tracking student movements, and simulation log data that maps student movements and actions into the evolving science simulation (Davalos Anaya, et al, in press). In addition, we are combining state-of-the-art deep learning methods and human-in-the-loop learning to perform some of the interaction analyses online and capture events of interest as students enact a scenario, and then provide this information back to teachers and students to enhance classroom teaching and learning experiences. Currently, we have developed and applied methods for motion and gaze tracking for groups of 3-4 students enacting a scenario, the teacher, and other students in the classroom environment who support the group in their enactments. We are also developing face tracking algorithms that capture students' affective states, with the goal of extending the socio-cognitive framework in LEAF to a socio-cognitive-affective framework for embodied and collaborative learning.

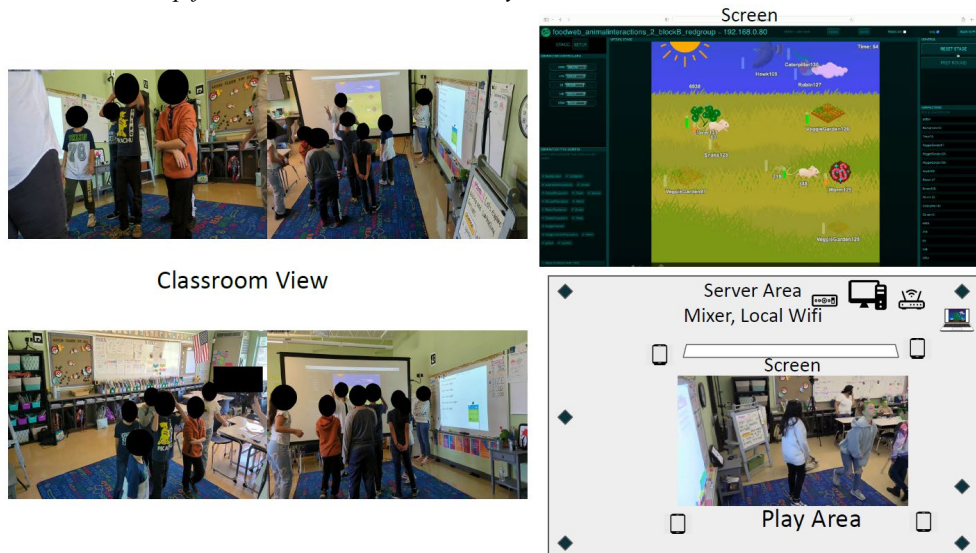
### Challenges and lessons learned

In this work supported by the NSF AI Institute on Engaged Learning, we have faced a number of computational, logistic, and IRB-related challenges in deploying our MMLA pipeline called Chimera Py in a constrained and noisy classroom environment (Figure 3). ChimeraPy is a minimal setup distributed streaming platform, is optimized for high throughput multimodal data transfer. Its architecture is designed to handle multiple challenges, such as time-aligned data collection and scalable multimodal analysis that overcomes the computational complexity of running multiple deep learning algorithms online by developing a reconfigurable distributed computing architecture. Besides, we have had to deal with additional issues, such as coherent treatment of signals



from heterogeneous sensors, and the requirements for fusing large volumes of multimodal data. To address privacy issues, it advocates for in-network de-identification of sensitive data like video or GPS before internet transmission. Furthermore, in classroom environments space is constrained, configurations differ from classroom to classroom, and setup and takedown of the equipment has to happen quickly so as not to disrupt class scheduling and day to day classroom activities.

**Figure 3.**  
*Classroom Setup for Data Collection and Analysis.*



Additional challenges that we are working on include human in the loop training of deep learning algorithms to study gaze cohesion and shifting from video that includes multiple actors (Zhou, et al, in review), the diarization of speech and the use of LLMs to summarize student conversations, and link the movements, gaze events, and utterances to generate a comprehensive summary of students' enactments in a scenario.

### Suggestions for the field

Collecting, aligning, and analyzing multimodal data itself is a challenge, and it becomes an even greater challenge when the data has to be processed online to support feedback to learners as well as provide actionable insights to teachers to support their debriefing activities. Two key suggestions in this regard are: (1) the need for close collaboration between learning science and AI researchers to ensure that the analyses and inferences made from the data can be linked to educational theories; and (2) the emphasis on active learning that includes human in the loop few-shot approaches and in-context learning in training and fine tuning the deep learning models for multimodal analysis. Other suggestions include the development of robust infrastructure pipelines to support data collection and analyses in classroom environment, and building in ethical considerations into

## Contribution #5: Collaboration analytics in K-12 and higher education space

Marcelo Worsley

### Project description

Over the past decade, learning analytics researchers have developed a host of dashboards, toolkits, and algorithms that can help researchers and educators leverage video and audio data. However, we very seldom consider ways that learners might want to leverage this data to support self-regulated learning, metacognition, or reflection. The Building Literacy in N-Person Collaboration (BLINC) project (Worsley, Anderson, Melo & Jang, 2021) addresses this opportunity by providing a platform whose capabilities align with the ways that students want to learn about their collaboration practices. The BLINC architecture allows participants to use web-enabled devices with a microphone and, optionally, a camera to collect collaboration analytics in near real-time. The platform provides metrics about: distribution of speech, question annotation, sentiment analysis, topic annotation and keyword detection. Most metrics provided through BLINC are customizable to the user's needs. Additionally, the BLINC platform provides user anonymity by representing data at the group level.

## Challenges

Developing this platform is requiring us to overcome a number of ongoing challenges. One of our early challenges was to think about a hardware infrastructure that would allow for reliable use across different contexts, and that could scale quickly. To address this challenge, we employ a *bring your own device* capability where participants use their own smartphones, computers, or tablets to utilize the platform. This eliminates the need for specialized hardware that can be hard for organizations to acquire and deploy. At the same time, the bring your own device capability, together with some backend decisions gives learners more ownership over when and how their data is collected. Students exercise control over when their data is being recorded.

Another significant challenge is student concerns about privacy and anonymity. As noted previously, we represent data at the group level, and avoid labeling any of the data with names. However, surveys among college students suggested that many of them would be concerned about having a video recording device used within the BLINC platform. We address this concern in a number of ways. First, participants can decide if they want to enable the video component of the platform at the start of each collaboration session. Secondly, we have integrated a cartoonification of the video data to obfuscate participant identities. We are currently in the process of testing if this approach will sufficiently assuage concerns around the use of video. While video is an optional feature, the addition of video information can help document contextual information about the collaboration, and support the extraction of additional relevant features (e.g., attention, facial expressions, head pose, gestures).

An additional set of challenges that we will briefly mention is with regard to the constantly changing landscape of artificial intelligence. Research teams around the world are constantly creating new innovations, and pushing the boundaries on what we can extract from various modalities. From a platform development perspective, we want to ensure that our tool is doing the best that it can to accurately represent student engagement and participation, but also want to maintain a stable and reliable platform.

## Lessons learned

One of the main lessons learned within this work is the importance of building around the needs and desires of the stakeholders. In particular, we have centered the needs of students and made significant effort to ensure that the data, data representations, and interface aligns with their goals. At the same time, we have observed the importance of understanding the technological constraints/pain points of participants. While many technologies can work well in the context of small scale laboratory spaces, the challenge of scaling to large numbers of users in ecological settings can add significant constraints. For the BLINC project, we found success in using students' smartphones, as opposed to trying to deploy specialized microphone arrays. This constraint did result in some shifts in the quality of data that we could collect, and required us to find some additional approaches for giving people the data that they wanted to see without utilizing state of the art technology. The other major lesson that we have learned along the way has been the importance of not forgetting about the contextual nature of data and how it is interpreted. Our research participants have done a great job of reminding us that the shifts in context can result in the same piece of data being seen in a completely different light. Hence, part of our job is to provide participants with the pertinent data, and not be too quick to draw decontextualized inferences about what that data might be suggesting about the nature of the collaboration.

## Suggestions for the field

Engaging participants in the process has been integral to our design process. This has meant thinking broadly about how the platform might be used, as well as connecting with participants that might have some very specific use cases. Part of being able to follow a user-centered design process is having a technical architecture that features flexibility, with components that can be quickly customized or swapped out. For BLINC this involves having a web-based API and a collection of services that we can connect to.

An additional suggestion is to acknowledge that perfection is infeasible, while also acknowledging where your tool might have gaps. There is no artificial intelligence tool on the market that works perfectly. The MMLA tools that we employ will not be an exception to that trend. In our case, we avoid making specific recommendations about how people should collaborate because of known shortcomings in our ability to effectively capture and represent the context. Instead, we invite learners to practice careful reflection and metacognition with regard to their data.

## Discussion

The five contributions of this symposium highlight key challenges in the use of MMLA in real-world settings. Beyond technical obstacles (implementation, data collection, storage, data fusion, analysis, data-driven interventions), one of the foremost challenges is the sensitive nature of data privacy. MMLA researchers must

navigate the complex landscape of ethical considerations, ensuring the protection of student information while leveraging data to enhance learning experiences. As suggested above, adopting a user-centered approach that involves educators, students, and stakeholders in the design process, can foster analytics tools that not only inform but also empower users. This requires analytics platforms to be intuitively understandable and seamlessly integrated within the existing pedagogical frameworks, ensuring that insights are directly translated into practical strategies for enhancing teaching effectiveness and student learning outcomes. This symposium and its panelists will explore these questions in depth, in collaboration with the ISLS community.

## References

- Blikstein, P. (2013, April). Multimodal learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 102-106).
- Cai, Z., Davis, R. L., Tormey, R., Dillenbourg, P. (2023). Learning Analytics Beyond Traditional Classrooms: Addressing the Tensions of Cognitive and Meta-Cognitive Goals in Exercise Sessions. *Manuscript submitted for publication*.
- Chng, E., Seyam, M., Yao, W., & Schneider, B. (2022). Toward Capturing Divergent Collaboration in Makerspaces using Motion Sensors. *Information and Learning Sciences*, Vol. 123 No. 5/6, pp. 276-297.
- Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, 51(5), 1441-1449.
- Danish, J. A., Enyedy, N., Saleh, A., & Humburg, M. (2020). Learning in embodied activity framework: A sociocultural framework for embodied cognition. *International Journal of Computer-Supported Collaborative Learning*, 15, 49-87.
- Davalos, E.A., Timalcina, U., Zhang, Y., Wu, J., Fonteles, J.H. & Biswas, G. (2023, in press). ChimeraPy: A Scientific Distributed Streaming Framework for Real-time Multimodal Data Retrieval and Processing. *IEEE Big Data*, Sorrento, Italy, December.
- Davis, B., Tu, X., Georgen, C., Danish, J. A., & Enyedy, N. (2019). The impact of different play activity designs on students' embodied learning. *Information and Learning Sciences*, 120(9/10), 611-639.
- Davis, R. L., Olsen, J. K., Courtois, J., Jermann, P., Tormey, R., Dillenbourg, P. (2023). Dashboards do not support the teaching styles of most instructors: How do we move forward? *Manuscript submitted for publication*.
- Guillain, L. V., & Schneider, B. (2021). Facilitators First: Building a Tool With Facilitators to Foster a More Collaborative Makerspace Community Through Movement Traces. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 533-539).
- Martinez-Maldonado, R., Echeverria, V., Fernandez-Nieto, G., Yan, L., Zhao, L., Alfredo, R., ... & Gašević, D. (2023). Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction*. In press.
- Schneider, B., Fang, X., Kang, A., & Chejara, P. (submitted). Advances in Collaboration sensing: A Comprehensive Review of Multimodal Metrics and their Connection with Collaborative Constructs from 2000 Onwards. *Computer-Supported Collaborative Work*.
- Sung, G., Guillain, L., & Schneider, B. (2023). Can AI Help Teachers Write Higher Quality Feedback? Lessons Learned from Using the GPT-3 Engine in a Makerspace Course. In *Proceedings of the 17th International Conference of the Learning Sciences* (pp. 2094-2095).
- Tu, X., Danish, J., Georgen, C., Humburg, M., Davis, B., & Enyedy, N. (2019). Examining how scientific modeling emerges through collective embodied play.
- Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, 2(2), 5-13.
- Worsley, M., Anderson, K., Melo, N., & Jang, J. (2021). Designing Analytics for Collaboration Literacy and Student Empowerment. *Journal of Learning Analytics*, 8(1), 30-48.
- Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022). Scalability, sustainability, and ethicality of multimodal learning analytics. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 13-23).
- Zhang, M, Fonteles, J.H., et al (2023, in review). Exploring Artificial Intelligence Supported Interaction Analysis. *4th Annual ISLS Conference*, Buffalo, NY.

## Bridging Social Annotation Practice with Perspectives from the Learning Sciences and CSCL

Xinran Zhu (Co-chair), University of Pennsylvania, xrzhu@upenn.edu  
Yeonji Jung (Co-chair), University of Memphis, yeonji.jung@memphis.edu  
Bodong Chen (Co-chair), University of Pennsylvania, cbd@upenn.edu  
Daniel Hickey (Discussant), Indiana University Bloomington, dthickey@indiana.edu  
Grant Chartrand, Indiana University Bloomington, gchartra@indiana.edu  
Remi Kalir, University of Colorado Denver, remi.kalir@ucdenver.edu  
Justin Hodgson, Indiana University Bloomington, hodgson@indiana.edu  
Chris Andrews, Independent Researcher, chris@chrisandrews.com  
Alyssa Wise, Vanderbilt University, alyssa.wise@vanderbilt.edu  
Hong Shui, University of Minnesota Twin Cities, shui0003@umn.edu  
Pingting Chen, Independent Researcher, paintingcpt@gmail.com  
Rukmini Manasa Avadhanam, University of Illinois Chicago, rukminia@illinois.edu

**Abstract:** Social annotation has emerged as an important approach to supporting students' social interaction and collaborative knowledge building in the classroom. Despite great interest among practitioners and a growing body of literature, social annotation activities are often guided by practical intuitions rather than informed by theories of learning and technology-supported collaboration. To strengthen social annotation practice, more work is needed to explore the systematic application of rich theories of learning and collaboration in this context. The proposed hybrid symposium aims to engage learning scientists, CSCL researchers, and stakeholders in productive dialogues to explore the integration of social annotation as a complex practice that can benefit from meaningful application of theories, explicit consideration of learning constructs, and careful design of technological and analytical support. The symposium will both contribute to social annotation practice in the classroom and help learning scientists and CSCL researchers in achieving broader impacts in the education system.

### Introduction

Annotation covers a broad territory. It has been construed in many ways: as link making, as path building, as commentary, as marking in or around existing text, as a decentering of authority, as a record of reading and interpretation, or as community memory. (Marshall, 1998, p.40)

Annotation, the practice of adding notes, comments, and other representations of ideas to text, is an important part of human cognition that supports reading, writing, and scholarship (Marshall, 1997). For example, readers actively engage with textual content by annotating printed books, with their annotations serving a multitude of functions such as procedural signals, recall cues, and attention traces (Marshall, 1997; O'hara & Sellen, 1997). Annotation can be dynamic—it evolves as readers interact with texts by adding new meanings, which reflects the evolving thinking processes and cultural contexts of its different readers (Liu, 2005; Marshall, 1998). *Web annotation* is a genre of information technology that offers an interactive way for users to engage with digital content, allowing them to add, share, and collaborate on annotations directly over web resources (W3C Web Annotation Working Group, 2016). Since the pioneering Annotea project (Kahan & Koivunen, 2001), various tools, such as NB (Zyto et al., 2012), Hypothesis, and Perusall, have emerged to support web annotation and its application across varied fields. More recently, these tools have undergone notable improvement with the development of more adaptable forms such as public web versions, browser plugins, and Learning Management System (LMS) integrations, further broadening their application in education.

*Social annotation* in this symposium refers to the application of web annotation technologies in educational settings to support student interaction around course materials and with each other. In higher education, social annotation has been widely adopted as an online discussion activity where students collaboratively read and annotate course readings (Sun et al., 2023). Unlike a post in a traditional discussion forum, a student annotation anchors a discussion to their original context, making the discussion more specific and focused (Sun & Gao, 2017). A growing body of research has investigated the design and implementation of social annotation activities in classrooms (Andrews et al., 2019; Chen, 2019; Hollett & Kalir, 2017; Zhu et al., 2023). Studies suggest that social annotation, utilized across education levels, could help in processing domain-specific

knowledge, promoting argumentation and literacy skills development, supporting assessment, and connecting online learning spaces (Zhu et al., 2020).

Despite growing interest in both research and practice, many existing social annotation activities remain generic and not clearly driven by theories of learning or collaboration. For instance, instructors may routinely mandate students to annotate and reply for a certain number of times during a week. This approach appears to insufficiently align the affordances of web annotation with possible scenarios of productive student engagement with course content and with each other. A recent literature review indicates that the majority of current studies in social annotation do not explicitly state learning theories they employ and are only implicitly informed by theories mentioned in literature review or activity design (Sun et al., 2023). The lack of theoretical grounding in the design of social annotation activities may lead to a misalignment between the design activity and target learning outcomes, limited understanding of the mechanisms that drive effective learning in social annotation, and inconsistent implementation of social annotation activities across different contexts.

To strengthen social annotation practice in the classroom, more work is needed to explore the systematic application of rich theories of learning and technology-supported collaboration in the social annotation context. Work in this fertile area is already happening. For example, Andrews et al. (2019) investigated the use of an *expansive framing* framework in an undergraduate course, with the goal of enhancing generative collaborative learning in social annotation activities. Expansive framing encourages students to relate their immediate learning experiences to broader contexts and future applications and offers support for productive conversation and knowledge transfer. In a separate study, Zhu et al. (2023) developed a scaffolding framework with predefined participation roles (including facilitator, synthesizer, and summarizer) for learners to play each week to improve their social interaction and cognitive engagement. These efforts have demonstrated the prospects of infusing theories of learning and collaboration in the design of social annotation activities. However, these studies are only scratching the surface and more work is needed to bridge theories of learning and computer-supported collaborative learning (CSCL) with social annotation as a complex educational practice shaped by a constellation of factors including teachers, learners, technologies, and even the open web. How can we build on theorizations of social annotation as conversation along cognitive, social, cultural, and political dimensions (e.g., Kalir & Dean, 2018; Marshall, 1997) to pursue new ways of conceptualizing social annotation in learning spaces? How can theories that are actively explored in the learning sciences inform this pursuit? At the same time, how may CSCL theories that think carefully about individual minds, small groups, larger communities, digital objects, and collaboration processes (Stahl & Hakkarainen, 2021) shed light on the social processes around annotation artifacts? How could technological affordances offered by web annotation technologies be leveraged to support CSCL practices such as establishing a joint problem space, communicating with each other, and creating a shared knowledge space (Chen & Lin, 2020; Jeong & Hmelo-Silver, 2016)? It is time to expand from prior efforts by engaging learning and CSCL researchers in conversations with designers and participants of social annotation activities to collaboratively explore this rich space. This exploration can generate more theoretically robust and practically impactful designs of social annotation and hereby enhance social annotation experiences in the classrooms. Given the increasingly broader reach of social annotation, this work will be extremely meaningful for learning scientists and CSCL researchers who are interested in achieving tangible change in the education system and large-scale environments (Wise & Schwarz, 2017).

To this end, the symposium aims to delve into the nuanced process of integrating social annotation in authentic learning settings in order to initiate conversations within the learning sciences community to advance social annotation research and practice. Our goals are threefold: (1) Advancing theory use – We seek to encourage a more sophisticated application of learning theories in social annotation, which involves translating abstract theoretical concepts into pedagogical strategies by understanding how theories can be more effectively applied in practical settings; (2) Connecting learning constructs with technology affordances – We aim to precisely map the affordances of social annotation technologies—such as interactivity around artifacts and visibility of student ideas (Hennessey, 2011; Pifarré, 2019)—with specific learning constructs, and thereby create a more targeted and effective learning experience; (3) Exploring design complexity – The symposium will discuss the complexities involved in social annotation practice, including disciplinarily specific learning outcomes, instructor design decisions, learning analytics applications, and new technologies targeting higher-order learning skills.

To achieve these goals, this symposium has involved five teams from multiple institutions who have been actively exploring social annotation from distinct perspectives. Each presentation is charged to respond to the following questions, with the hope of stimulating rich conversations within the learning sciences and CSCL research communities and between the research communities and people who are actively involved in social annotation:

1. What learning theories can productively inform the research and design of social annotation practice?
2. What learning constructs are supported by the affordances provided by social annotation technologies?

3. How can we develop robust methodologies to investigate and assess the learning occurring within social annotation activities?

## Personal versus professional authenticity in social annotation

Daniel Hickey and Grant Chartrand

We explore a crucial aspect of annotation that deserves systematic consideration. This concerns how learners *frame* their annotations (i.e., contextualize, as in Goffman, 1974). Social annotation naturally lends itself to “personal” framing where learners are pushed to find connections with individually relevant people, places, topics, and times, beyond the boundaries of the course. In practice, framing is closely related to the more pragmatic principle of *authenticity*, as popularized in the essay by Brown et al. (1989) which introduced many to situated cognition.

Personal authenticity was explored in a 2021 expert consensus study report on computing education from the National Academies of Sciences, Engineering, and Medicine by leading learning scientists (including Barbara Means, Victor Lee, and Mimi Ito). They juxtaposed personal authenticity with *professionally* authentic experiences that are widely used to frame learning in STEM contexts and beyond. The report argued forcefully that professionally authentic experiences often marginalize learners from non-dominant backgrounds and that *all* learners can benefit when *each* learner engages in personally authentic experiences (e.g., Calabrese-Barton & Tan, 2019).

Personal authenticity is entirely consistent with the situative design principles for *expansive framing* in Engle et al. (2012). This is important for at least two reasons. The first is the five compelling explanations that Engle and colleagues presented to explain why expansive framing should support generative learning that transfers readily and widely (c.f., Hickey, 2022). The second reason is that expansive framing and its precursor principles for *productive disciplinary engagement* (Engle & Conant, 2002) are being extended to support equity and inclusion (see especially Agarwal & Sengupta-Irving, 2019, systematically reviewed in Freedman et al., in review).

We are motivated by the fact that the design principles for expansive framing have yet to be widely taken up by others; our own efforts to promote expansive framing in social annotation and beyond (e.g., Hickey et al., 2020) have had limited impact. We suspect that this is because many assume that situative theories of learning (i.e., Brown et al., 1989) call for “real world” (i.e., professionally authentic) experiences (e.g., Herrington, 2014) and because of Engle’s problematic contrast with “bounded” framing.

Our presentation will first summarize the theory and practice of expansively framed, personally authentic social annotations. This will include using *socio-political uncertainties* (Agarwal & Sengupta-Irving, 2019; Hickey & Quick, 2020) that invite minoritized learners to position themselves as having unique expertise. The presentation will then summarize supporting evidence from discourse analysis in case studies of three fully online courses. These include secondary computing, undergraduate learning theories, and graduate learning & cognition.

## Identifying linguistic, cognitive, and social indicators of undergraduate students’ social annotation

Remi Kalir, Justin Hodgson and Chris Andrews

The use of social annotation as a learning activity in undergraduate education can productively support students’ textual analysis, collaboration, and knowledge production (Kalir et al., 2020; Morales et al., 2020). In the context of composition and literature courses, social annotation can aid students’ reading strategies and writing skills as “readerly additions” (Davis & Mueller, 2020) augment mentor texts and enhance peer discourse (Hodgson, Kalir, & Andrews, 2023). This paper reports initial findings from a broader, large-scale study of undergraduate students’ reading and writing practices to identify the prevalence of linguistic, cognitive, and social qualities in student writing as evidenced by social annotation.

Through a research-practice partnership, the English Department at a large public research university in the Midwest implemented social annotation activities in all standard sections of a required composition course for first-year students. During the spring 2021, fall 2021, and spring 2022 semesters, over 50 course sections each term—enrolling on average over 1,000 students—used the social annotation tool Hypothesis to read and discuss texts. We examined student social annotation of three texts selected because each appeared in the most sections across the three semesters (S21, F21, S22): A chapter of Gloria Anzaldúa’s *Borderlands/La Frontera* was annotated in 40 sections by 772 students; a selection from Jeffrey Jerome Cohen’s *Monster Theory: Reading Culture* was annotated in 37 sections by 695 students; and a chapter of John Berger’s *Ways of Seeing* was

annotated in 33 sections by 633 students. Following data collection, we used Linguistic Inquiry Word Count (LIWC; Pennebaker et al., 2015), a validated text analysis tool, to calculate the frequency of linguistic features found in 3,868 student annotations of Anzaldúa’s text, 3,425 annotations of Cohen, and 3,029 annotations of Berger. Table 1 reports the average frequency of key linguistic, cognitive, and social language indicators as demonstrated in undergraduate students’ social annotation.

**Table 1**  
*Average Linguistic Frequencies of Students’ Social Annotation (S21, F21, S22)*

Text	Analytic	Clout	Authenticity	Tone	Cognition	Social Processes
Anzaldúa	47.72	53.91	35.55	32.61	17.08	17.38
Cohen	55.68	51.64	45.73	26.60	17.85	11.96
Berger	55.03	62.84	42.70	33.26	16.67	15.44

Across three texts and three semesters, our results indicate that undergraduate students’ social annotation demonstrated a mix of analytical thinking and conversational discourse, expressive confidence in line with other studies of online learning (e.g., Moore et al., 2021), and moderate levels of emotional authenticity and tone. When responding to varied texts and peers, cognitive features of student writing were evident in approximately 17% of all annotation text, a frequency higher than similar analyses of students’ online discussion (e.g., Zhu et al., 2019). This analysis of first-year students’ writing is an exploratory account of how regularly linguistic, cognitive, and social attributes appear in over 10,000 instances of social annotation. As data were collected prior to the widespread use of generative artificial intelligence writing tools, our results also provide descriptive insight about students’ authentic online language use and sense-making when jointly interacting with texts and peers.

## Supporting actionable social annotation through learning analytics

Yeonji Jung and Alyssa Wise

Effective social annotation relies on students actively engaging with shared materials and co-developing ideas with peers. However, challenges such as inconsistent engagement, low-quality annotations, and hesitation in tool use hinder its potential (Novak et al., 2012). These issues manifest in students’ tendencies to focus on irrelevant details, produce repetitive annotations, and accept information uncritically, stemming from difficulties in interacting with related information and identifying the parts needed for attention (Ghadirian et al., 2018; Novak et al., 2012).

To address the challenges, this study uses a learning analytics approach to promote effective learning through social annotation (Zhu et al., 2020). Through extensive human-centered design activities with students and instructors, the student-facing analytic tool was developed with the primary objective of meeting the need for timely guidance in identifying areas for meaningful contributions to the social annotation activities. The tool provided individualized analytic-driven suggestions about where they could contribute to their social annotation tasks (e.g., “Buzz! Check out this active conversation”, “Connect with someone new”), featuring three tool characteristics: integration into existing learning tools, direct paths to action, and alignment of analytics with learning activity timing. This tool was implemented in a fully asynchronous course to 91 students twice a week for five weeks. Students received different versions of the analytics depending on their participation status (whether they had started participating in the learning task or not) and the time of the week (early or late).

Using student access data and interview responses, this study examined how students engage in the learning tasks of social annotation throughout the week and whether and how their use of analytics might make a difference to their existing learning routines. Findings showed that while opening the analytics promptly, students used the analytics in different ways for social annotation, either backward or forward in their learning routines. In some cases, students did not initially use the analytics as a precursor to the annotation tasks. Instead, they later used them to review and stay connected with the completed tasks. However, lower open rates identified in backward use raise questions about connection between task completion and importance attributed to analytics. In other cases, students used the analytics as a proactive tool to prepare for upcoming annotation tasks, even including other course assignments, gaining a prospective idea of what would happen in the annotation tasks before starting them. Several students, who followed an all-at-once learning routine, tended to make comments based on the analytics while reading, particularly when the analytics suggested intriguing questions or relevant content. In a few cases, some students took a combined approach, using the analytics as an aid when encountering

difficulties in understanding readings and identifying relevant spots to contribute, which was identified as a critical need in the participatory design process. In this case, they focused on reading to extract the main ideas or clarify complex aspects, further helping them generate ideas and make comments on ongoing conversations. The variation in student use of analytics depending on their learning status indicates that aligning the delivery of analytics with students' existing routines may promote timely access, but it is not enough. This was coupled with students' reported challenges that while they opened their analytics right away, this was not the time when they normally did annotation tasks, so they did not use analytics directly. This highlights potential areas for future work, considering customizing the timing of analytics delivery to ensure that students receive timely and relevant feedback in their own timelines of engaging in social annotation tasks. This study suggests the potential of using analytics to foster social annotation learning for students to find particular areas for contribution and enhance engagement.

## Supporting knowledge synthesis in social annotation activities

Xinran Zhu, Bodong Chen, Hong Shui and Pingting Chen

The knowledge-creation perspective grounded in CSCL literature (Paavola & Hakkarainen, 2014; Paavola & Hakkarainen, 2021) emphasizes the role that student-created artifacts play in mediating learning and collaboration. This viewpoint conceptualizes learning as the process of creating and progressively refining shared artifacts through interactions with peers. From this perspective, the design of social annotation activities should consider two key aspects: facilitating learning and collaboration by acknowledging annotations as valuable artifacts and nurturing the continuous development of students' ideas that emerge from engagements on social annotation platforms. In this presentation, we introduce a design research project, named Knowledge Synthesis, that aims to foster productive interaction and related knowledge practices in social annotation activities by recognizing these aspects, as part of a multi-year research-practice partnership.

The Knowledge Synthesis project tackles a key challenge in social annotation practices: the limited opportunities for successive idea refinement beyond the initial round of annotations. Typically, social annotation activities end once students have completed their annotations, neglecting the potential of these contributions to facilitate ongoing collaborative learning. Consequently, ideas generated from the annotations remain isolated from other learning activities, thereby restricting their further development and hindering their integration into broader knowledge practices. This project aims to address this challenge by proposing a *knowledge synthesis intervention*. Knowledge synthesis is an important form of human cognition that involves skillfully and strategically weaving together diverse strands of information to foster conceptual innovation, generate novel knowledge, and design creative solutions (Deschryver, 2014; Morabito & Chan, 2021; Scardamalia & Bereiter, 2014). This intervention recognizes students' annotations as objects open to continuous development, engaging students to connect, analyze, and expand upon their ideas through the synthesis processes. Meanwhile, the synthesis products can be integrated into other learning events, enriching the overall learning experiences. Concepts related to knowledge synthesis, such as "rise above" in Knowledge Building (Scardamalia & Bereiter, 2014), have been recognized across various fields for their role in fostering individual growth and collaboration. Despite its recognized importance, there is still a notable gap in understanding how knowledge synthesis operates within CSCL environments and how it can be explicitly supported to become a pivotal element in knowledge creation.

The knowledge synthesis intervention involves a web application developed by the research team, named the Synthesis Lab, which retrieves students' social annotation data and provides explicit scaffolds to guide students' synthesis making process. The workflow within the tool aims to achieve two primary goals: categorizing peers' ideas into Conceptual Building Blocks (Morabito & Chan, 2021), and developing a synthesis of the discourse. The application provides a structured workspace for students to decompose the complex synthesis task into smaller building blocks, such as distilling, connecting, analyzing, and rising above ideas generated from the annotations. To effectively incorporate this tool into classroom settings, pedagogical support informed by the CSCL literature has been designed. One such design involves collaborative scripts that scaffold student participation. This is achieved by assigning a pair of students each week as discourse facilitators. These facilitators were asked to promote deep thinking in annotations and foster engagement in peer responses. Additionally, they used the Synthesis Lab to connect and synthesize student ideas from the annotations before in-person class discussions, with particular focus on key themes, disagreement, and confusions. This synthesis was then used to mediate further in-person discussions or group projects, thereby enhancing the overall collaborative learning experience. Through empirical implementation of the design in a graduate classroom, we examine how students perceive knowledge synthesis as a part of their collaborative learning, how they synthesize the student-created artifacts, and how knowledge synthesis mediates ongoing interaction and knowledge creation throughout the learning experience.



In this project, CSCL theories were used to guide both technological and pedagogical designs, particularly in how it conceptualizes learning as the process of creating and developing shared artifacts. A key aspect of this alignment is the recognition of the mediational role played by student-created artifacts in their collaborations, such as their annotations and syntheses. This approach emphasizes the importance of digital artifacts not just as byproducts of learning activities but as living components in the ongoing collaborative discourse process, shaping further learning and application. Additionally, investigation of this intervention also demonstrates the potential to further expand theories of learning and CSCL designs.

## **Exploring instructor use of social annotation in undergraduate online courses**

Rukmini Manasa Avadhanam

Research on social annotation in higher education online learning has increased exponentially in the past two decades (e.g., Novak et al., 2012; Zhu et al., 2020). However, this rich body of literature mainly studied the evaluation of social annotation tools and their effectiveness on student-related measures, and very few studies discuss instructors' perspectives and their use of social annotation. There needs to be more knowledge about the processes and challenges instructors face in using and implementing social annotation in undergraduate online courses. The lack of studies on instructor perspectives on social annotation makes it challenging to understand the teaching, assessment, and participation strategies that effectively achieve the course objectives, improve student learning outcomes, and engage students in learning. It is also important to understand instructors' design and pedagogical processes as they use social annotation tools to facilitate collaborative learning in online learning environments and their processes to facilitate learning through student annotations.

Through a rich, descriptive, in-depth qualitative case study, this study delved into instructors' processes behind the thoughtful and intentional design of social annotation activities that enhance CSCL in undergraduate students. This study aims to understand how and why instructors use social annotation to achieve their pedagogical goals, the processes behind the thoughtful and intentional design of social annotation activities for their online classes, and their perception of how it impacts student learning experiences. The findings of this study illustrate rich descriptions of instructor design and implementation processes of five instructors teaching online courses in two modalities, asynchronous and synchronous. The instructors were mainly from the schools of social sciences and humanities, whose teaching philosophies included social constructivism and active student dialogue and conversations to enhance learning. They learned about social annotation tools and tried the tools with zest as they identified that their course objectives aligned well with the use of social annotation. Instructors strongly voiced their dislike for threaded discussion boards and shared how they noticed visible student dialogue and critique as they used social annotation tools like Hypothesis. Thematic analysis of qualitative data sources also elaborates that instructors use social annotation tools to create an authentic, collaborative learning community for student discussion and to ensure student perspectives are more visible. Instructors' design and pedagogical processes, like providing guiding prompts, participation-based assessment strategies, and instructor participation to further student discussion, are also evident. They also indicated the differences in synchronous and asynchronous modalities of online courses, how they influence student participation in social annotation, and the added challenge of designing and facilitating the conversations. The study's implications indicate how there should be more focus on instructor use of learning technologies, support them institutionally with professional development, and communities of practice.

## **Significance of the symposium**

In this symposium, we engage learning and CSCL researchers in conversations with designers and participants of social annotation activities to collaboratively explore the integration of social annotation as a complex practice in education. Studies presented at this symposium approach learning and CSCL theories from different angles, including examining students' reading strategies and writing skills demonstrated in social annotation through linguistic, cognitive, and social language indicators (Kalir et al.), exploring the role of personal authenticity in the framing of annotations from a socio-political lens (Hickey & Chartrand), applying learning analytics to enhance student engagement in social annotation (Jung & Wise), understanding the role of annotations as knowledge artifacts in mediating ongoing collaborative learning processes (Zhu et al.), and investigating instructors' perspectives in implementing social annotation in higher education classrooms (Avadhanam). Collectively, this symposium demonstrates meaningful application of theories, explicit consideration of learning constructs, support for instructor decision making, and careful design of technological and analytical support. This hybrid symposium will not only contribute to social annotation practice in the classroom but also help learning scientists and CSCL researchers in achieving broader impacts in the education system.

## References

- Agarwal, P., & Sengupta-Irving, T. (2019). Integrating power to advance the study of connective and productive disciplinary engagement in mathematics and science. *Cognition and Instruction*, 37(3), 349–366
- Andrews, C. D., Chartrand, G. T., & Hickey, D. T. (2019). Expansively framing social annotations for generative collaborative learning in online courses. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *Proceedings of the 13th International Conference on Computer Supported Collaborative Learning - CSCL 2019* (pp. 33–40).
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Calabrese Barton, A., and Tan, E. (2019). Designing for rightful presence in STEM: Community ethnography as pedagogy as an equity-oriented design approach. *Journal of the Learning Sciences*, 28(4–5), 616–658.
- Chen, B. (2019). Designing for networked collaborative discourse: an UnLMS approach. *TechTrends*, 63(2), 194–201. <https://doi.org/10.1007/s11528-018-0284-7>
- Chen, B., & Lin, F. (2020). Representational affordances for collaborative learning in technology-enhanced environments. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (1st ed., pp. 513–531). Routledge. <https://www.routledge.com/Handbook-of-Learning-from-Multiple-Representations-and-Perspectives-1st/Van-Meter-List-Lombardi-Kendeou/p/book/9780367001179>
- Davis, M., & Mueller, A. (2020). The places of writing on the multimodal page. In P. R. Powell (Ed.), *Writing changes: Alphabetic text and multimodal composition* (pp. 103–122). Modern Language Association of America.
- Deschryver, M. (2014). Higher order thinking in an online world: Toward a theory of web-mediated knowledge synthesis. *Teachers College Record*, 116(12), 1–44.
- Engle, R. A., Lam, D. P., Meyer, X. S., & Nix, S. E. (2012). How does expansive framing promote transfer? Several proposed explanations and a research agenda for investigating them. *Educational Psychologist*, 47(3), 215–231.
- Freedman, E. B., Hickey, D. T., Schamberger, B., Chartrand, G. T., Harris, T., & Luo, M. Q. (in review). Promoting equity through productive disciplinary engagement and expansive framing. *Educational Psychologist*.
- Ghadirian, H., Salehi, K., & Ayub, A. F. M. (2018). Social annotation tools in higher education: a preliminary systematic review. *International Journal of Learning Technology*, 13(2), 130–162.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Hennessy, S. (2011). The role of digital artefacts on the interactive whiteboard in mediating dialogic teaching and learning. *Journal of Computer-Assisted Learning*, 27(6), 463–586.
- Herrington, J., Reeves, T. C., & Oliver, R. (2014). *Authentic learning environments* (pp. 401–412). Springer.
- Hickey, D. T. (2022). Productive disciplinary engagement and expansive framing: The situative legacy of Randi Engle. In M. McCaslin & T. Good (Eds.) *Routledge Online Encyclopedia of Education*.
- Hickey, D. T., Chartrand, G. T., & Andrews, C. D. (2020). Expansive framing as pragmatic theory for online and hybrid instructional design. *Educational Technology Research and Development*, 68, 751–782.
- Hickey, D. T., & Quick, J. D. (2020). A modest feature for repositioning minoritized online students to support disciplinary engagement and achievement. In M. Gresalfi, & I. S. Horn (Eds.), *Proceedings of the International Conference of the Learning Sciences* (pp. 2435–2436).
- Hodgson, J., Kalir, J., & Andrews, C. (2023). Social annotation: Promising technologies and practices in writing. In O. Kruse, C. Rapp, C. Anson, K. Benetos, E. Cotos, A. Devitt., & A. Shibani (Eds.), *Digital writing technologies in higher education: Theory, research, and practice* (pp. 141–155). Cham, Switzerland: Springer Nature.
- Hollett, T., & Kalir, J. H. (2017). Mapping playgrids for learning across space, time, and scale. *TechTrends*, 61, 236–245.
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, 51(2), 247–265. <https://doi.org/10.1080/00461520.2016.1158654>
- Kalir, J. H., & Dean, J. (2018). Web annotation as conversation and interruption. *Media Practice and Education*, 19(1), 18–29. <https://doi.org/10.1080/14682753.2017.1362168>
- Kalir, J., Morales, E., Fleerackers, A., & Alperin, J. (2020). “When I saw my peers annotating:” Student perceptions of social annotation for learning in multiple courses. *Information and Learning Sciences*, 121(3/4), 207–230.

- Kahan, J., & Koivunen, M. R. (2001). Annotea: an open RDF infrastructure for shared Web annotations. In *Proceedings of the 10th International Conference on World Wide Web*. 623–632.
- Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of documentation*, 61(6), 700–712. <https://doi.org/10.1108/00220410510632040>
- Marshall, C. C. (1997). Annotation: from paper books to the digital library. *Proceedings of the second ACM international conference on digital libraries*. 131–140. <https://doi.org/10.1145/263690.263806>
- Marshall, C. C. (1998). Toward an ecology of hypertext annotation. *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*. 40–49. Association for Computing Machinery Press.
- Moore, R. L., Yen, C. J., & Powers, F. E. (2021). Exploring the relationship between clout and cognitive processing in MOOC discussion forums. *British Journal of Educational Technology*, 52(1), 482–497.
- Morabito, J. S., & Chan, J. (2021). Managing context during scholarly knowledge synthesis: Process patterns and system mechanics. *Proceedings of the 13th Conference on Creativity and Cognition*. 1–5.
- Morales, E., Kalir, J., Fleerackers, A., & Alperin, J. (2022). Using social annotation to construct knowledge with others: A case study across undergraduate courses. *F1000Research*, 11:235, <https://f1000research.com/articles/11-235/v2>
- National Academies of Science, Engineering, and Mathematics (2021). *Cultivating interest and competencies in computing: Authentic experiences and design factors*. National Academies Press.
- Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education*, 15(1), 39–49.
- O’Hara, K., & Sellen, A. (1997). A comparison of reading paper and on-line documents. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 335–342. <https://doi.org/10.1145/258549.258787>
- Paavola, S., & Hakkarainen, K. (2014). Trialogical approach for knowledge creation. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 53–73). Springer.
- Paavola, S., & Hakkarainen, K. (2014). Trialogical Approach for Knowledge Creation. In S. Tan, H. So, & J. Yeo (Eds.), *Knowledge creation in education*. Springer. [https://doi-org.proxy.library.upenn.edu/10.1007/978-981-287-047-6\\_4](https://doi-org.proxy.library.upenn.edu/10.1007/978-981-287-047-6_4)
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015 operator’s manual* (pp. 1–23). Pennebaker Conglomerates.
- Pifarré, M. (2019). Designing a dialogic technology-enhanced pedagogy to support collaborative creativity. In N. Mercer, R. Wegerif, & L. Major (Ed.), *The Routledge international handbook of research on dialogic education* (pp. 704–725). Routledge.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. R. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). Cambridge University Press.
- Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In U. Cress, C. Rosé, A. Wise, & J. Oshima (Eds.), *International handbook of computer-supported collaborative learning* (pp.23–43). Springer.
- Sun, Y., & Gao, F. (2017). Comparing the use of a social annotation tool and a threaded discussion forum to support online discussions. *The Internet and Higher Education*, 32, 72–79. <https://doi.org/10.1016/j.iheduc.2016.10.001>
- Sun, C., Hwang, G. J., Yin, Z., Wang, Z., & Wang, Z. (2023). Trends and issues of social annotation in education: A systematic review from 2000 to 2020. *Journal of Computer Assisted Learning*, 39(2), 329-350.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12, 423–467.
- W3C Web Annotation Working Group. (2016). available at: [www.w3.org/annotation/](http://www.w3.org/annotation/)
- Zhu, M., Herring, S. C., & Bonk, C. J. (2019). Exploring presence in online learning through three forms of computer-mediated discourse analysis. *Distance Education*, 40(2), 205–225.
- Zhu, X., Chen, B., Avadhanam, R. M., Shui, H., & Zhang, R. Z. (2020). Reading and connecting: using social annotation in online classes. *Information and Learning Sciences*, 121(5/6), 261–271.
- Zhu, X., Shui, H., & Chen, B. (2023). Beyond reading together: Facilitating knowledge construction through participation roles and social annotation in college classrooms. *The Internet and Higher Education*, 59, 100919.
- Zyto, S., Karger, D., Ackerman, M., & Mahajan, S. (2012). Successful classroom deployment of a social document annotation system. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1883-1892.

# Posters



# Examining Scratch Programming: Exploring its Capacity to Cultivate Computational Thinking-aligned Learning Opportunities

Faiaz Rahman, Kathy L. Schuh  
faiaz-rahman@uiowa.edu, kathy-schuh@uiowa.edu  
University of Iowa

**Abstract:** In 1980, Papert introduced the world to computational thinking (CT) and the concept was later expanded upon by Wing. The current study examined the development of CT in college students as they participated in iterative block-based programming activities. Data from 16 students in an online course revealed differences in CT presence between student-created Scratch codes and discussion posts. Collaborative assignments supported CT *practices* like *testing and debugging* and CT *perspectives* like *expressing, connecting, and questioning*.

## Introduction and grounding

Seymour Papert (1980) introduced the world to the concept of “Computational Thinking” (CT) (p.182). Later, Wing (2011) expanded upon the concept, defining it “*the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively executed by an information processing agent*” (2011, p.1). While several tools have previously been used to explore CT, Scratch, a block-based programming language, has gained recent popularity as a tool for CT exploration. Despite extensive research (Chou, 2020; Grover & Basu, 2017; Zhang & Nouri, 2019), two gaps exist in literature. First, most studies primarily involve K-12 students, with less emphasis on college students. Second, these studies focused on assessing codes and concepts, with lower emphasis on collaboration among students. This study addressed these gaps by examining CT development in college students as they engaged in collaborative Scratch-based design activities. The core research questions (RQ) were,

1. In what ways do CT concepts, practices, and perspectives manifest in college students as they participated in iterative Scratch-based programming activities?
2. What is the role of collaboration experienced during Scratch-based programming activities on the development of CT concept, practices and perspectives in college students?

Grounded in constructionism (Papert, 1980), this study conceptualized CT using Brennan and Resnick's (2012) CT framework that categorizes CT into three dimensions: (a) *concepts, practices*, and *perspectives*, which includes connecting, questioning, and expressing as components.

## Methods

In Spring 2020 and 2021, 18 college students participated in an 8-week online participatory learning course at a midwestern R1 University. Deidentified data generated by these students were used for this study. Students utilized Scratch to design four applications and completed discussion board posts. Posts were qualitatively analyzed by the primary investigator (PI) and a second coder (SC), with over 90% agreement in excerpt selection. Excerpts were categorized into 14 CT characteristics based on Brennan and Resnick's (2012) framework. Scratch codes were examined visually for evidence of CT *concepts, practices, and perspectives*.

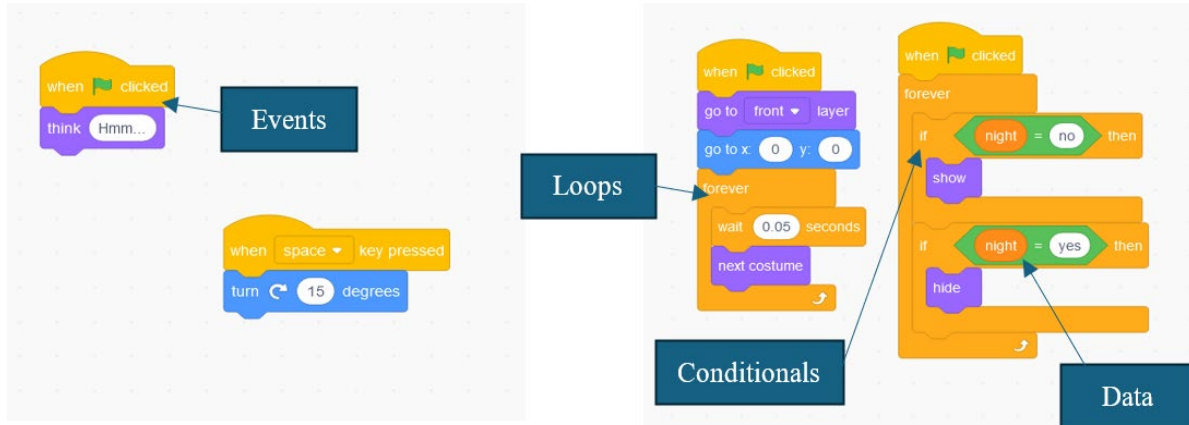
## Findings

RQ1 examined how CT manifested in students while working on Scratch projects. Figure 1 shows codes for 2 sprites created by student 117 during their first and third build respectively. In their first build, 117 created an alphabet learning application. Within the sprite “*Characters 1*” the student demonstrated an understanding of the CT *concept events*, whereby if a green flag is clicked, a character will display a thought bubble that said “Hmm” The third iteration of the project revealed CT *concepts of loops, data and conditionals*. In “*Sprite17*” (two animated birds) the left forever block will repeatedly wait for 0.05 seconds and change costumes (depicting movement of the birds) in a loop and the same sprite contains an *if...then* block (the block indicating understanding of *conditionals*) where the birds are shown *conditionally* based on the yes/no parameters of a student created variable named “night” (variable creation indicating understanding of *data*).

RQ2 focused on the role of collaboration in the development of CT in college students. The majority of evidence connected to this question was found in the discussion posts. 177 excerpts were extracted, of which 43% were indicators of CT *practices*, 18% were indicators of CT *concepts* and about 38% were indicators of CT *perspectives*. CT *practices* was the most prominent indicators present in the discussion posts. For example, student

**Figure 1**

“Characters 1” Sprite Codes in 117’s First Build (left) and “Sprite17” Codes in 117’s Third Build (Right)



114 stated, “I added a wolf sprite which I found license free through Google search.” This is an example of the CT *practice, reusing and remixing*. Meanwhile, in the context of *being incremental and iterative*, student 112 wrote, “This week, I decided to continue building upon my original work from last week and make it more interactive for the viewer.” This comment suggests the student iteratively enhancing their projects with new features. Next, in the context of CT *perspectives*. Student 113 commented, “I am thinking this could be Dracula returning home at dawn from a fun night out on the town.” This is an indicator of the CT *Perspective, expressing*. *Connecting* was also prominent in the posts. For example, after exploring student 106’s Scratch build, student 103 stated, “However, at the end the sprite stays at the top of the cage somewhat hidden. When I tried to click it to move it, it started playing the script from the beginning.” To this, 106 replied, “Thanks for the feedback. I was struggling on my end to make the sprite return to where it belongs.” In a subsequent build, student 106 seemed to have fixed the positioning of the sprite (making it less hidden). In the context of the final factor, *questioning*, student 115 stated, “My biggest concern is time/resources to allow students the opportunity to create builds. When do students get the time to experiment/play with Scratch?” This is indicative of the student questioning the “taken-for-granted” (Brennan & Resnick, 2012, p.11).

Concerning RQ1, Scratch codes show more CT concepts, and discussion posts highlight practices and perspectives. Analyzing both is crucial for understanding CT manifestation. Regarding RQ2, collaborative assignments enhance CT practices like testing and debugging, as well as foster the expression of CT perspectives through idea sharing, peer connections, and questioning Scratch’s functionalities.

## Implications

Findings in the study can inform learning scientists about the ways in which certain CT characteristics like *perspectives*, appearing muted in Scratch codes, can manifest during peer-peer discussions among college students. In addition, the findings could also inform educators about the effectiveness of collaborative activities in the development of CT in college students.

## References

- Brennan, K., & Resnick, M. (2012, April). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American educational research association*, Vancouver, Canada (Vol. 1, p. 25).
- Chou, P.-N. (2019). Using scratchjr to foster young children’s Computational thinking competence: A case study in a third-grade computer class. *Journal of Educational Computing Research*, 58(3), 570–595. <https://doi.org/10.1177/0735633119872908>
- Grover, S., & Basu, S. (2017). Measuring student learning in introductory block-based programming. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. <https://doi.org/10.1145/3017680.3017723>
- Papert, S. A. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic books.
- Wing, J. M. (2011). Research notebook: Computational thinking—What and why. *The link magazine*, 6, 20-23.
- Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, 141, 103607. <https://doi.org/10.1016/j.compedu.2019.103607>

## Not the ‘Usual Suspects’: Caregiver Networking and Imagination in a Family Engagement Program

Kira Baker-Doyle, Sarai Coba-Rodriguez, Andrea Vaughan, Evelyn Pollins  
kirabd@uic.edu, scobaro2@uic.edu, avau6@uic.edu, epolli2@uic.edu  
University of Illinois at Chicago

**Abstract:** Lack of investment in humanizing social infrastructures, especially in historically marginalized communities, has contributed to an epidemic of loneliness and social isolation in the U.S. (Office of the Surgeon General, 2023). This study uses a critical social network theory lens to examine the ways in which a virtual bilingual family literacy program in Chicago, IL fostered social connections among families. Using case study methodology, we analyzed the stories and experiences of 14 caregivers across two program cohorts. Our findings showed that while families made weak social ties in the virtual space, the diversity of the community fostered ‘imagination’ among caregivers, and a sense of value for the community. These findings have implications for critical scholarship on and design of social infrastructures for sustaining speculative thinking and exchange.

### Introduction: Major issues and significance

In 2023, U.S. Surgeon General Vivek Murthy identified social isolation as one of the greatest threats to the health and wellness of individuals and communities in America. While the Surgeon General’s report on this issue cited the pandemic as an amplifier of social isolation, it identified the lack of ‘social infrastructure’ as a key contributor to the rise of social isolation in the country. The report defined social infrastructure as,

...the physical assets of a community (such as libraries and parks), programs (such as volunteer organizations and member associations), and local policies (such as public transportation and housing) that support the development of social connection (OSG, 2023, p. 18).

Furthermore, the report acknowledged that poor and historically marginalized communities have an additional burden overcoming social isolation due to bias, discrimination, and historical disinvestment in social infrastructures such as schools, local businesses, and housing.

In this paper, we ground our work in such a systemic understanding of social isolation and the important role of humanizing social infrastructure in facilitating social connection. Furthermore, we situate our work within the critical, asset-based approach to designing family engagement programming, which seeks to transform programming and educational practices in ways that sustain diverse cultures and languages (Bang & Voussoughi, 2016; Flores & Springer, 2021). This stance seeks to counter the more historically engrained deficit-oriented perspective on family engagement that centers individuals as problematic and focuses on assimilative educational practices (Gutiérrez, 2008).

### Theoretical framework

We use a critical social network theory framework, drawing specifically from Baker-Doyle’s work (2023). A ‘critical’ perspective entails attention to the broader historical and systemic social power structures that influence social network formation, and a focus on systemic transformation of social inequalities. Granovetter (1973) proposed that weakly-linked diverse networks or ties can foster greater social capital, however, there are few (if any) critical social network studies that explore this possibility, particularly with regards to the experiences of marginalized individuals. Drawing from dialectic change theory, which is rooted in critical theories of transformation, Baker-Doyle (2023) supposed that such network constellations may have value within a more mutable understanding of social capital. Under her ‘dialectic social infrastructure design’ proposition, Baker-Doyle theorizes that it is in the dialectic (shifts) between affinity-type networks and more diverse networks that new forms of social capital can arise. In our critical exploration of the ties developed among adults in the RMP program, we apply (and test) this proposition.

### Context and methodology

Our study examines the ways in which a virtual bilingual family literacy program in Chicago, IL called “Read, Make & Play” (referred to here as RMP), facilitated social connection among adults in the program. RMP provided bookbags full of children’s literature, games, and arts and crafts to a cohort of families with young children and facilitated six online, weekly discussion sessions with caregivers and children in a virtual



environment (Zoom) about their experiences using the materials. This study is a culmination of data from two cohorts (Spring 2021 and Summer 2022) with 14 RMP caregiver participants in total. Data collection strategies entailed in-depth interviews and observational data. For data analysis, we drew from qualitative approaches to studying social connection and networks (e.g. Coburn, 2005) and applied a critical lens, focusing particularly on participants stories of belonging and participation, both in the RMP program and in other educational spaces.

### **Findings: Access, participation, and diversity in virtual social infrastructures**

Three main themes arose from our analysis of the data: access, participation, and diversity of ties. The first two themes, access and participation, reflected participants' comments on the affordances and constraints of virtual space. While the virtual space made the program easily accessible to participants, it also limited their ability to interact and thus build strong ties with each other. Caregivers reported sometimes feeling shy to share in the virtual space or finding discouraged by seeing others' cameras turned off. These two themes were present in almost every interview of participants and were somewhat expected based on previous literature on networking in virtual spaces. However, it was the third theme that emerged from the data that was more nuanced and revealing: the value of the diverse connections that participants made in the program.

Program participants indicated that they built weak ties (i.e., not long-lasting, or deeply personal) with other participants, and that most of their connections were with people that they would not normally meet during their daily routine. Several caregivers indicated that the diversity of participants was an asset to the program and their experience. For example, Liza stated, "I liked that it wasn't the usual suspects, the parents that you would see at the park all the time or at the library programs." According to participants, the benefits of connecting with others outside typical social circles were two-fold. First, several noted that they felt affirmed in some of the challenges they faced because they met so many different people that faced the same parenting challenges. Second, they found new ideas around how to engage their children in literacy activities by meeting diverse others. Our findings uncovered the value of diverse connections to caregivers: affirmation, transformation, and imagination.

### **Discussion: Social infrastructures for capacity-building and imagination**

The findings in this study help to expand our understanding of the role that virtual spaces can play in facilitating the creation of diverse, weak ties, for historically marginalized families and the value that they provide to members. While much previous literature on designing supportive social infrastructures for historically marginalized groups has pointed to identity-based affinity groups as an important structure for humanizing capacity-building and wellness (Mosely, 2018), this study reveals the potential value of another social infrastructure, one built of diverse, weak ties, but thoughtfully facilitated through a shared activity or interest. While affinity groups may support capacity-building, these findings suggest that diverse networks, even with remote connections, can foster 'imagination' and affirmation, given careful facilitation.

### **References**

- Baker-Doyle, K.J. (2023) *Critical Network Literacy: Humanizing Professional Development for Educators*. Cambridge, MA: Harvard Education Press.
- Bang, M., & Vossoughi, S. (2016). Participatory Design Research and Educational Justice: Studying Learning and Relations Within Social Change Making. *Cognition and Instruction*, 34(3), 173–193.
- Coburn, C. (2005). Shaping Teacher Sensemaking: School Leaders and the Enactment of Reading Policy. *Educational Policy*, 19(3), 476–509.
- Flores, T. T., & Springer, S. (2021). Our legends and journey stories: Exploring culturally sustaining family engagement in classrooms. *Theory Into Practice*, 60(3), 312–321.
- Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Gutiérrez, R. (2008). Research commentary: A gap-gazing fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, 39(4), 357–364.
- Mosely, M. (2018). The Black teacher project: How racial affinity professional development sustains Black teachers. *The Urban Review*, 50(2), 267–283.
- Office of the Surgeon General (OSG). (2023) *Our epidemic of loneliness and isolation*. U.S. Public Health Service. <https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf>

# CocoNote Supported Project-Based Learning Environment: Perspectives of Construction and Collaboration

Haiyang Xin, Qiannan Niu, Linling Lan, Zhuoxin Xiao, Fan Wu  
tony@cocorobo.cc, niuqiannan@cocorobo.cc, lanlinling@cocorobo.cc, xiaozhuoxin@cocorobo.cc,  
wufan@cocorobo.cc  
CocoRobo Ltd.

**Abstract:** This article introduces CocoNote, a collaborative hypermedia whiteboard designed to address the challenges of implementing Project-Based Learning (PBL) in K-12 education. It facilitates students' transition from knowledge consumers to creators, supporting teachers in PBL course design and offering learning behavior analytics. A study with students from Hong Kong and Mainland China showcases CocoNote's effectiveness in enhancing collaborative problem-solving.

## Introduction

Facing the 21st-century challenge of preparing students for a tech-rich future, traditional education often falls short. Project-Based Learning (PBL) offers a dynamic approach to deepen students' understanding of complex topics, yet its adoption in K-12 is hampered by the transition to self-directed models, the need for skilled teamwork facilitation, and a scarcity of PBL-specific tools for feedback and progress monitoring.

CocoNote, inspired by David Perkins' "Knowledge as Design," offers a collaborative hypermedia whiteboard that transforms students from passive recipients to active creators of knowledge (see Figure 1). By enabling the organization and linkage of various information forms, it fosters collaborative learning and individual insight sharing. CocoNote supports teachers in PBL implementation through preset resources, customizable note combinations, and facilitation of group work and feedback, enhanced by Learning Behavior Analytics and Generative AI capabilities. This comprehensive tool aims to bridge the gap in PBL application, promoting continuous knowledge building and problem-solving.

## Theoretical framework

Project-Based Learning (PBL) is deeply rooted in constructivism, which posits that students learn most effectively when they actively construct knowledge by engaging with ideas in real-world contexts (Krajcik, 2014). To create an effective computer-supported learning environment for PBL, we draw upon activity theory, which provides a framework for us to understand the complex interactions within learning environments (see Figure 2). It emphasizes six key components in the group learning activity of PBL environment (Engeström, 2000): *Subject* (learners, who are the primary agents in the PBL process), *Object* (tasks and problems that drive the learning process, leading to tangible solutions or artifacts), *Community* (collaborative groups formed by teachers and students), *Rules and division of labor* (the norms and role distributions within the PBL groups), *Instrument* (tools and resources provided by the external environment to facilitate learning activities). Above theories guide the design of functionalities that support the dynamic nature of PBL in CocoNote.

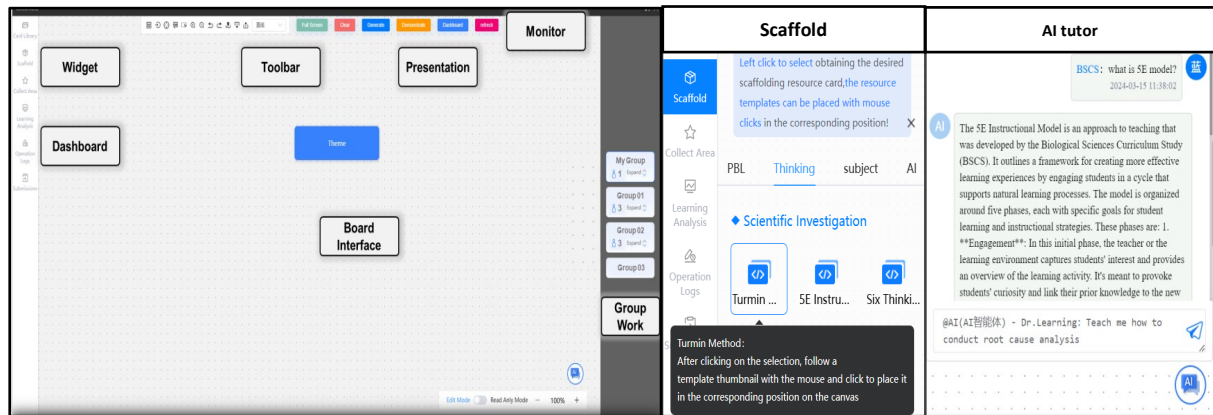
## Practice

In a study, 132 third-grade students from a Hong Kong junior high and over 30 elementary students from Mainland China engaged in Project-Based Learning (PBL) using CocoNote. The Hong Kong groups, consisting of four students each, researched the wellbeing of middle-aged and elderly populations, proposing solutions via CocoNote. They aimed to complete this semester-long project by its conclusion. Simultaneously, the Mainland students, in teams of 3-4, examined how AI could enhance traditional culture through CocoNote, across multiple class sessions. Both sets of projects highlighted the importance of teamwork and innovative problem-solving on a digital platform.

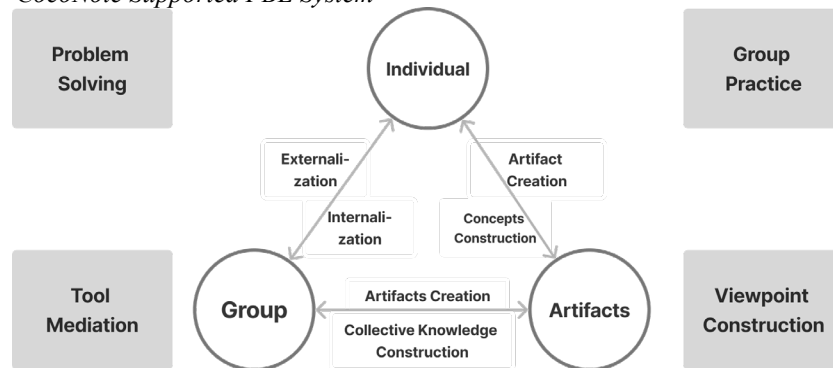
Our examination of CocoNote's application in PBL centered on observing team dynamics and collaborative learning processes, drawing on theoretical concepts of group performance over time. We collected data on student interactions within CocoNote, including discussions on the group whiteboard, navigation paths, and generated project artifacts.

This data provided a foundation for analyzing the collaborative problem-solving process. It was organized by resource, activity session, and task to uncover patterns in collaborative problem solving process (see Table 1). Through coding of chat data for decision-making stages—ranging from problem definition and solution generation to evaluation—insights into the dynamics of group interactions were gained. The objective was to

**Figure 1**  
*The Interface of CocoNote*



**Figure 2**  
*CocoNote Supported PBL System*



demonstrate the superiority of CocoNote-supported learning in enhancing collaborative problem solving, as opposed to traditional learning environments, by detailing the decision-making and implementation processes at various points of problem-solving.

**Table 1**  
*From Raw Data to a Meaningful Sequence: An Illustration*

Raw	Abstracted	Aggregated	Meaningful sequence
Very detailed log	User X, Wiki contribution on page Y, Time T	Action sequence on page Y: User X (leader), User Y, User Y, User Z, User X (leader)	Group leader creates a page, 2 other group members contribute and leader contributes again

## Future directions

As generative AI opens up transformative possibilities for changing the shape of education, CocoNote will explore the use of GenAI and AI Agents for coaching collaboration in PBL-based collaborative learning environments, as well as immersive learning environments that support PBL, such as group learning in augmented reality and virtual reality, to facilitate student knowledge construction and self-directed deep learning.

## References

- Engeström, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, 43(7), 960-974.
- Krajcik, J. S., & Namsou, S. (2014). Project-Based Learning. In Sawyer, R. K. (Eds.), *The Cambridge Handbook of the Learning Sciences*(pp. 275–297). Cambridge University Press.

# Real-time Continuous Perceived Difficulty Annotation during Video Learning Using a More Natural Interface

Xiaobo Liu, Yu Zhang  
liu-xb22@mails.tsinghua.edu.cn, zhangyu2011@tsinghua.edu.cn  
Tsinghua University

**Abstract:** In video-based learning, students' perceived difficulty often varies over time. Capturing these dynamics enhances learning and instructional design. Existing continuous annotation tools often require devices that are not commonly available and may add unnecessary cognitive load. To overcome these challenges, we developed a more natural annotation interface that leverages mouse scroll and a shadow effect. A subsequent user study demonstrated its feasibility in educational practice. Lastly, we outlined its potential applications for learners and instructors.

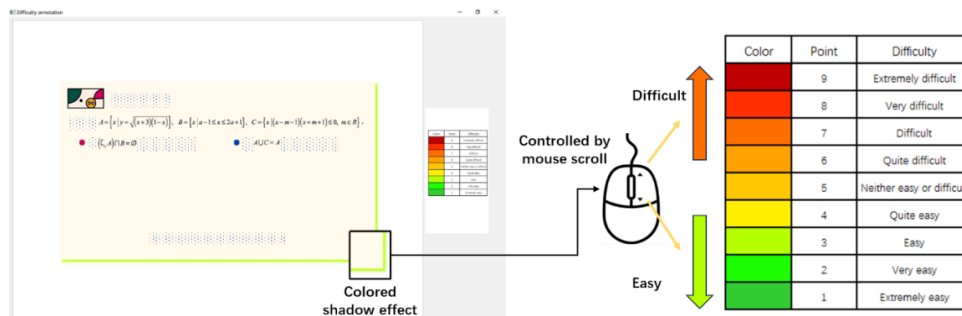
## Introduction

Video has emerged as an essential and popular tool in education, enriching blended learning environments and engaging digital-native learners. Since the advent of video-based learning, its effectiveness has been debated, particularly concerning the insufficient provision of feedback. Although post-video tests and surveys offer valuable insights, they inadequately capture the learning dynamics. Notably, within these dynamics, the perceived difficulty may fluctuate over time, as suggested by the theories on instantaneous cognitive load (Xie & Salvendy, 2000). Analyzing dynamic patterns of perceived difficulty can enhance knowledge consolidation for learners, while recognizing common patterns can guide teachers in instructional design.

Recent studies have utilized real-time continuous annotation tools to record psychological states during video watching. For instance, Girard and C. Wright (2018) introduced an annotation tool employing a joystick to capture emotions in real-time. Similarly, Srivastava et al. (2019) created a tool for instant difficulty rating of video lectures using a physical slider. These innovations have enhanced the collection of dynamic user responses throughout the learning process. However, the limited availability of joysticks and sliders restricts their utility in educational contexts. Moreover, the separate visual interfaces for annotation and video watching in these tools could induce extraneous cognitive load, potentially impacting the learning experience. Therefore, we introduced a more natural interface using mouse scroll and shadow effects to overcome these limitations and enhance the learning experience. A pilot user study was subsequently carried out to examine the feasibility of the tool.

## Methods

**Figure 1**  
*Design and Implementation of the Annotation Tool*



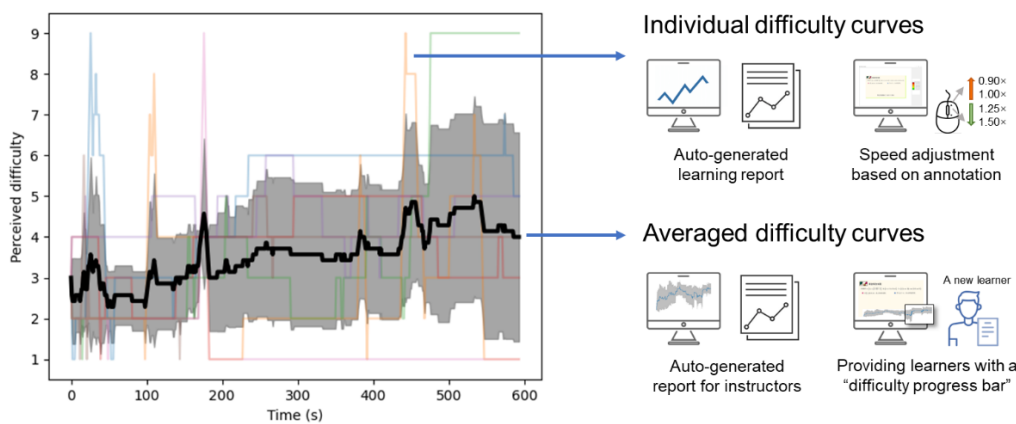
In our system design (Figure 1), the mouse is utilized to annotate the difficulty level through vertical scrolling. The annotations correspond to the perceived difficulty scale (Sweller et al., 2011), which uses a 9-point Likert scale ranging from 1 (extremely easy) to 9 (extremely difficult). To minimize the visual separation between annotation and video viewing, we implement a colored shadow effect to represent the annotated difficulty levels. A shift towards green denotes easier content, suggesting relaxation, whereas a transition to red indicates more challenging material, correlating with stress. This user-friendly interface enables learners to intuitively annotate their perceived difficulty levels.

In the subsequent pilot user study, seven students from the same blended learning course were recruited. They annotated three video clips which were selected from the course material, and then filled out a questionnaire to assess the overall difficulty with the same scale. To analyze the dynamics of the annotations, the mean and quartile values of the curves were calculated, and they were then compared to the overall difficulty ratings.

## Results and discussion

The average value of the curve showed a strong correlation with the overall ratings (Pearson's  $r = 0.815$ ,  $p < 0.001$ ,  $n = 21$ ), suggesting the tool's feasibility in capturing instantaneous perceived difficulty. Notably, this correlation coefficient exceeded those using quartile values (median:  $r = 0.764$ , lower quartile:  $r = 0.751$ , upper quartile:  $r = 0.762$ ). These results are consistent with the theoretical relationship between instantaneous and overall cognitive load (Xie & Salvendy, 2000).

**Figure 2**  
*A Sample Difficulty Curve and Its Potential Applications*



Moving beyond the statistical analysis, we further explored the data through the visualization of individual and averaged difficulty annotation curve for the same video, as illustrated by the example in Figure 2. In the beginning of the video, three distinct difficulty peaks were observed, indicating that these students encountered challenges with different concepts. Analysis of the averaged curve revealed an increasing trend in difficulty, which is consistent with the instructor's expectation.

Drawing on the results of the pilot study, several potential applications are proposed based on the annotation tool, as shown in Figure 2. In a prospective application, individual learning curves could inform an intelligent tutoring system to automatically generate a learning report, identifying challenging concepts for the learner and providing targeted explanations. Additionally, the system might be envisioned to dynamically adjust the video speed in response to annotations of difficulty, thereby creating a suitable pace for learners. Similarly, the averaged curve could offer valuable insight into collective learning challenges, which could potentially tell teachers which concepts are difficult to comprehend, as well as providing a "difficulty progress bar" to assist new learners in navigating challenging parts of a video lesson.

## References

- Girard, J. M., & C. Wright, A. G. (2018). DARMA: Software for dual axis rating and media annotation. *Behavior research methods*, 50, 902-909.
- Srivastava, N., Velloso, E., Lodge, J. M., Erfani, S., & Bailey, J. (2019). Continuous evaluation of video lectures from real-time difficulty self-report. Proceedings of the 2019 CHI conference on Human factors in computing systems,
- Sweller, J., Ayres, P., Kalyuga, S., Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. *Cognitive load theory*, 71-85.
- Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & stress*, 14(1), 74-99.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (62177030, 62107025).

## Scaffolding Students' Adoption of Norms for Peer Critique

Jinzi Zhou, Lauren Albert, Cindy E. Hmelo-Silver, Joshua Danish, Zachary Ryan, Christina Stiso  
zhoujinz@iu.edu, laabler@iu.edu, chmelosi@indiana.edu, jdanish@indiana.edu, zryan@iu.edu,  
cstiso@indiana.edu  
Indiana University

Danielle Murphy, Ravit Golan Duncan, Clark A. Chinn, Qiuyu Lin  
dm880@scarletmail.rutgers.edu, ravit.duncan@gse.rutgers.edu, clark.chinn@gse.rutgers.edu,  
qiuyu.lin@rutgers.edu  
Rutgers University

**Abstract:** This study aims to better understand norms that are potentially relevant for student engagement in peer critique within the modeling context. By analyzing multiple data from a 5<sup>th</sup>/6<sup>th</sup> grade classroom, we identified social and socioscientific norms within small groups. Students demonstrated productive within-group engagement in peer critique governed by norms, while encountering some difficulty with adopting certain socioscientific norms.

### Introduction and theoretical framework

Peer critique can serve to authentically support modeling practice (Ford, 2008). Models are abstracted representations of complex natural systems that can predict and explain scientific phenomena. Model critique involves evaluating models and assessing model-evidence fit (Duncan et al., 2018). Engaging students with model critique can facilitate collective construction of scientific knowledge (Ford, 2008). However, it remains relatively infrequent and challenging for students (González-Howard & McNeill, 2020). Current instructional practice insufficiently emphasizes the importance of critiquing models as a practice for science advancement. Consequently, students are often provided with limited opportunities to engage in critique, lacking familiarity with norms for this practice.

Norms regulate activities within a learning community and affect learning opportunities for students (Cobb et al., 2001). They can be categorized into social norms and socioscientific norms. Social norms refer to general expectations that regulate classroom practices; socioscientific norms highlight the aspects of discourse that are specific to science practices, in this case, peer critique of models. Drawing on sociocultural perspectives, we view peer critique as the activity mediated by shared goals, tools, norms, division of labor, and classroom community (Engeström, 1987). We focus on social and socioscientific norms, which shape and are shaped by shared goals of the activity. We situate this research in the context of using a modeling tool, the *Model and Evidence Mapping Environment* (MEME) for a model-based ecology unit. The study is guided by this research question: *What norms emerged when students provided and addressed peer critique within small groups?*

### Methods

This study involved 15 students in a 5<sup>th</sup>/6<sup>th</sup> grade classroom at an independent school in the Midwest US. Students, working in small groups, participated in seven 90-minute sessions spanning 4 weeks. In Sessions 2-5, they conducted two cycles of peer critique using MEME, each including a 15-20-min model critique and a subsequent 15-min model revision. The class discussed and developed *criteria for good models* and guidelines for *providing constructive critique*, serving as expected norms for students to adopt. Data sources included screen recordings of group activities, students' MEME models before and after, and peer comments. Two focal groups were chosen because they appeared to represent group interactions. Screen recordings were transcribed and segmented into episodes, each covering either one critique of a model, a discussion of a comment received, or a revision on the model. We inductively identified and coded norms that were present and referenced (3 times or more) in each episode; cases of norm violation demonstrated students' awareness of established norms. The first two authors jointly analyzed all the data, resolving all differences through discussion.

### Findings and discussion

We identified 32 episodes during peer critique activity. We present the norms coded as emerged in Table 1. Overall, our intervention, including MEME features (e.g., commenting box), led both groups to embrace the norms to similar and varying degrees. With respect to social norms, both groups demonstrated productive engagement in participating in the peer critique activity. Specifically, they took turns controlling the computer, constantly sought agreement in discussions, and provided clarifications to each other's questions. Their successful

collaboration cultivated a friendly and positive microculture within their groups as they progressed through the critique process. Group 2 seemed to adopt the routine of analyzing the model, selecting criteria, writing comments, and the norm of providing constructive feedback, which were evident in their completion of 15 model reviews accompanied by written comments with increasing specificity. In contrast, Group 1 displayed a less salient adoption of those two norms, as shown with fewer clear comments; however, they maintained lengthy and productive critique discussions during the reviewing process.

**Table 1**  
*Norms Emerged in Peer Critique Activity*

<b>Social norms</b> ( <i>general classroom expectations</i> )	<b>Socioscientific norms</b> ( <i>specific to peer critique</i> )
1) Take turns controlling the computer	1) Maintain shared attention to MEME models
2) Follow the routine: analyze the model- select the criterion – write comments	2) Use criteria for what counts as a good model to critique: <i>supported by evidence, understandable, consistent, show all steps of processes</i>
3) Provide constructive critique: “two plus one wish”	3) Analyze each component of model structure before giving critique
4) Provide clarifications to each other’s questions	4) Select the most relevant criterion for each comment
5) Seek agreement	5) Check evidence to verify model-evidence fit
	6) Understand what the critiques are
	7) Address peer comments on models
	8) Add evidence and details in descriptions to make models better

Students adopted some of the socioscientific norms well while encountering challenges in adhering to others. Both groups constantly maintained shared attention to MEME models. Through consistent use of criteria, they demonstrated a critical approach when identifying problematic models, including issues such as inclusion of irrelevant entities, incorrect links of evidence, or incorrect arrow directions between entities. Importantly, both groups showed a strong commitment to the accuracy of evidence interpretations by checking evidence included in the model before providing comments. This socioscientific norm emerged from the students, whereas other norms were co-established by the facilitators and students. This could be attributed to students’ grasp of evidence and appreciation of the criteria in earlier model construction practice, and MEME’s feature of providing easy access to evidence. Notably, Group 1 tended to prioritize the socioscientific norm 3) over other social and socioscientific norms. By contrast, Group 2 did not attend to this norm in several model critiques. Although both groups addressed some peer comments, they had difficulty in understanding the critiques that lacked specificity, ignoring several of them. While the facilitators explicitly requested that students add more evidence and details to their models, adhering to this norm was challenging.

Making invisible norms visible is essential to help students’ evidentiary practices (Duncan et al., 2018). Through exploring norms that mediated student engagement in peer critique, we can provide insights for designing scaffolds for students’ appropriation of norms. Particularly, explicit discussion about criteria of a good model, forms of constructive comments, and assessing evidence support this practice. However, the variations in how students embraced certain norms highlight the need to further explore how to better support students in appreciating and appropriating the norms embedded within design scaffolds.

## References

- Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards’ conceptualization of evidence. *Journal of Research in Science Teaching*, 55(7), 907-937.
- Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in classroom mathematical practices. *The journal of the Learning Sciences*, 10(1-2), 113-163.
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki, Finland: Orienta-Konsultit Oy.
- Ford, M. (2008). ‘Grasp of practice’ as a reasoning resource for inquiry and nature of science understanding. *Science & Education*, 17(2), 147-177.
- González-Howard, M., & McNeill, K. L. (2020). Acting with epistemic agency: Characterizing student critique during argumentation discussions. *Science Education*, 104(6), 953-982.

## Acknowledgement

This work was supported by the National Science Foundation (Grant No. DRL-1761019 and DRL-1760909).

# Inadequately Gifted: A Thematic Analysis of Memes as Embodied by gifted Experiences on Reddit

Andria Gentry, University of North Texas, andria.gentry@unt.edu

**Abstract:** This paper investigates how members of the r/aftergifted sub-Reddit form a community surrounding feelings of inadequacy and lived experiences within the gifted identity. Three uniquely formatted memes are analyzed, shedding light on common expressions within the community and demonstrating the potential of meme analysis to inform iterative cycles of research grounded in participants' perspectives and language. The goal is to provide voice to the growing r/aftergifted community and highlight the need to hear these perceptions as means of understanding the potential long-term experiences along with the "gifted" label.

## Introduction

Despite giftedness being recognized widely, there is no universal definition. Research calls attention to the importance of student voice in understanding the relationship between emotionality and giftedness (Albright & Montgomery, 2023). However, scholarship foregrounding perceptions within the Gifted community holds emphasis on perceptions within educational settings, or within educational achievement (Coleman et al., 2015), frequently voiced from the perspective of parents or teachers. Memes, often depicted as images, text, or videos, act as means of communicating expressions of humor, shared experiences, or overall cultural artifacts (Rogers & Giorgi, 2023). By acknowledging the work of members of the r/aftergifted community, this paper seeks to understand how individuals are desettling expectations of giftedness (Bang et al., 2012), describe their lived experience, and create opportunities for understanding themselves through the sharing of memes.

## Research methods

The data gathered within this paper comes from the publicly accessible r/aftergifted community on Reddit. This community was formed as a forum for discussing feelings of inadequacy that members may feel in relation to an identity within the gifted label. My process of analysis followed Thematic Analysis guidelines defined by Braun and Clarke (2006). I gathered memes, read and re-read each meme, taking notes on what each meme expressed. I then coded each meme to generate concise codes. I then examined and gathered codes to generate themes. Finally, I constructed narratives alongside memes to contextualization.

## Findings

### Burnt out child prodigy starter pack

One common meme format was a 'Starter Pack' (Figure 1A), where the idiocultural experience of "burnt out child prodigies" is described. Representations of what is associated with this subculture are featured through different media types. Phrases such as 'filling the void' nod to the act of replacing something that may be missing from lived everyday experience of members of the group. An image of a child holding their head in their hands is coupled with "immediately gives up if not perfect at something first try". This reference to 'giving up' on tasks when not immediately achieving high standards was frequented across r/aftergifted, as were depictions of medicine bottles, typically paired with language of mental health disorders (e.g., anxiety, depression). In comments section, members frequently shared connections to the memes or advice to others.

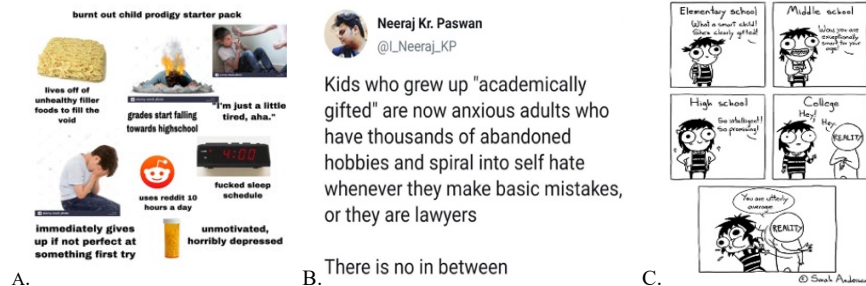
### Thousands of abandoned hobbies

Another common meme structure consisted of taking a screenshot of a humorous or relevant post from other social media platforms and sharing it on r/aftergifted. As depicted through "Thousands of Abandoned Hobbies" (Figure 1B), an over-generalization of what gifted populations experience in adulthood is made through humor. In this instance, gifted adults are being described as either being "anxious adults who have thousands of abandoned hobbies and spiral into self-hate whenever they make basic mistakes", or as "lawyers", with there being no middle ground between the two. In the comments, self-deprecating humor was expressed. Similarly to the previous meme, a theme of abandoning hobbies was discussed. One community member, while not identifying as a lawyer, did identify with the meme through commenting, "Some of us are teachers... with a bunch of abandoned hobbies, who are trying desperately to teach the next generation how to be OK with not getting everything right the first time. Because we wish someone had done that for us." While expressing a connection with the meme, there were also



times people expressed experiences of feeling an expectation to achieve things correctly the first time expressed, which appeared as a common theme throughout memes and comments.

**Figure 1**  
*Meme Starter Pack and Two Illustrations of Themes*



## What it's like in schools

“What It's Like in Schools” by the original artist Sarah Andersen (Figure 1C) depicts the experiences of a gifted character throughout schooling. Within K-12 schooling, the character was told they were phrases such as a “smart child” and “promising”. In college, the character was depicted being ‘hit by Reality’ in both a literal and metaphorical sense, while being told that they were “utterly average”. Throughout the Elementary, Middle, and High school boxes, external expectations were being expressed to the character unprompted. In College, the character was the first to speak to the character of Reality. Comments from community members once again shared expressions of a connection. Members emphasized a possible theme of comments from others shaping members’ expectations for themselves. While comments such as “You’re so smart for your age!” were depicted frequently throughout adolescence and childhood, these comments from others seem to disappear after high school leaving members feeling alienated from the label.

## Discussion and implications

Meme-analysis demonstrates the need for further research on lived experiences within those who have been identified as gifted after they are no longer involved in gifted programming, by showing the growing community of those who identify as ‘aftergifted’. Themes of mental health, feelings of not meeting others’ expectations, and perfectionism are stated throughout. Further research is needed to understand how members of r/aftergifted connected these experiences with those of mental health and feelings of ‘masking’ one’s identity in relation to the gifted identity. I would highlight the importance of hearing perspectives of those who are no longer in gifted programming’s experience across their lives to better understand the potential harm this label may cause when intellect is foregrounded rather than whole persons learning and growth. I hope to cultivate space to have their voices heard by drawing on this in-process analysis to develop protocols for understanding their experience.

## References

- Albright, E. A., & Montgomery, D. (2023). Perceptions of the emotional self for adolescents who are gifted. *Roeper Review*, 45(1), 6–20. <https://doi.org/10.1080/02783193.2022.2145399>
- Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2012). Desetting expectations in science education. *Human Development*, 55(5-6), 302–318. <https://doi.org/10.1159/000345322>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Coleman, L. J., Micko, K. J., & Cross, T. L. (2015). Twenty-five years of research on the lived experience of being gifted in school. *Journal for the Education of the Gifted*, 38(4), 358–376. <https://doi.org/10.1177/0162353215607322>
- Fine, G. A. (1979). Small groups and culture creation: The idioculture of Little League Baseball Teams. *American Sociological Review*, 44(5), 733. <https://doi.org/10.2307/2094525>
- Rogers, R., & Giorgi, G. (2023). What is a meme, technically speaking? *Information, Communication & Society*, 1–19. <https://doi.org/10.1080/1369118x.2023.2174790>

## Acknowledgments

I would like to thank my mentor and professor Dr. D. Teo Keifert for their constant support in developing conceptual frameworks of this paper. I would also like to thank my classmates for their continued support.

# Teachers' Discourse on WhatsApp In a Virtual Professional Learning Community

Iris Saban, Linor L. Hadar  
irisgreenhouse@gmail.com, lhadar@edu.haifa.ac.il  
University of Haifa, Israel

**Abstract:** In this qualitative research, we explored teachers' discourse within a WhatsApp-based teacher community. We analyzed 513 WhatsApp posts, categorizing them into five discourse types: general pedagogical discourse, sharing advice or experiences, responses to requests, personal expressions and greetings, and procedural discourse. The analysis reveals that the discourse within the community enhances teachers' collaborative partnership and fosters their active engagement.

## Introduction

Research underscores the pivotal role that teacher communities play in advancing teachers' professional learning. These collaborative models provide essential support to teachers, thereby engendering meaningful professional learning processes (e.g., Sudeck et al., 2005; Lieberman & Pointer Mace, 2010). Studies suggest a considerable potential for enhancing teachers' professional learning when teacher communities incorporate collaborative technological tools or when such learning transpires in virtual or blended formats (Damico, 2017; Lieberman and Pointer Mace, 2010; Ghamrawi, 2022)

Virtual communities operate within virtual environments, which are effective platforms for facilitating optimal teacher learning due to their enhanced collaborative features. Teachers actively exchange knowledge and teaching experiences within virtual spaces (Lantz-Andersson et al., 2018). Therefore, the integration of technology within a teachers' communities is considered to hold substantial potential in promoting teachers' professional learning, both at the individual level and in enhancing collaboration skills. Virtual teacher communities encompass diverse methodologies, each with unique approaches and practices (e.g., Riding, 2001).

This study is focused on a specific model of a virtual teacher community. We analyzed discourse characteristics inherent to this virtual community, which uses WhatsApp as the primary communication platform. This presentation provides a preliminary analysis of the discourse features that manifest exclusively in conversations among the community's teachers, disregarding the discourse initiated by the group facilitators.

## The virtual teachers' community model

In 2017, The Center for Educational Technology in Israel launched a virtual teachers' community to help teachers across Israel learn and integrate "Jewish-Israeli culture studies" into the curriculum. Inspired by the success of social network communities, such as those on Facebook, this initiative aims to promote learning, personal growth, and collaboration among teachers. Approximately 200 middle school teachers voluntarily participate in the community. While the community includes a core of dedicated participants, there is annual turnover, with new teachers joining and others departing. WhatsApp serves as the primary mode of communication in the community. Most community activities revolve around discussions and collaborations conducted within the WhatsApp group. The community is facilitated by two coordinators who possess expertise in both content and pedagogy. The coordinators are responsible for the community's activities, including providing resources, establishing discourse on specific themes, or creating asynchronous learning modules. Community-initiated dialogues complement coordinated activities, such as sharing and developing lesson plans collaboratively.

## Methods

### Goal and research question

This study explores the characteristics of teachers' discourse within a WhatsApp-based community, focusing on discourse features that manifest in conversations among the community's teachers, disregarding discourse initiated by the group facilitators. Lefstein et al. (2020) underscore the significance of teacher discourse, particularly within informal exchanges. WhatsApp conversations provide valuable insights into teachers' professional interactions.

### Study participants, data sources, and analysis

The study analyzed 513 WhatsApp posts originating from 167 middle school teachers from various schools in Israel. The teachers' posts were anonymized. The data were derived from the teachers' community WhatsApp

conversations, spanning from September to December 2020. Data analysis employed a qualitative thematic inductive methodology, focused on identifying patterns of repetitive word or term usage, exploring the contexts in which teachers employed them and discerning their underlying purposes thereof.

## Results and conclusion

Data analysis yielded five categories: (1) Teachers' sharing of advice and ideas grounded in their experiences. For example: sharing of lesson plans, learning materials, and insights from teaching and learning experiences, Sample quote: "Hi, I am sharing an experiential assignment that I created to encourage students' joy of learning, connecting to the concept of responsibility (chapter 4 in the textbook) ...". (2) Pedagogical discourse, for example: discussing instructional methods or strategies for responding to students' statements. Sample quote: "Would you show the song/presentation to secular students?..." (3) Responses to requests for specific assistance. This involves sending content in response to queries and engaging in discussion related to requests. Sample quote: "I suggest giving creative options such as a Kahoot, engaging presentations, or even a family newspaper. The goal is to provide guidance on enjoyable activities for students that won't add to their already heavy workload" (4) Procedural discourse. for example: requests to add a teacher into the community, seeking links, expressing interest in participating in online meetings. Sample quote: "A week ago we talked about a video of Israel as an opening activity, can someone share the link?" (5) Personal expressions and greetings. for example: compliments, acknowledgments, wishes, appreciation, encouragement to fellow teachers. Sample quote: "It's amazing!!!! Thank you. Happy and healthy new year to all of you ". Table 1 presents the distribution of categories.

Among the 167 teachers engaged in the community, 129 (77%) generated at least one post, with some teachers contributing up to 52 posts within the observed time frame. Beyond the exchange of content or pedagogical discourse pertaining to classroom activities, the teachers' personal emotional commitment is discernible through the utilization of personal expressions and the acknowledgment of the significance associated with their community participation. These preliminary findings are intriguing, and we eagerly anticipate a comprehensive examination of this dataset in the ongoing trajectory of our study.

**Table 1**

*Discourse characteristics:*

	Teachers' sharing advice	Pedagogical discourse	Responses to requests	Personal expressions / Greetings	Procedural discourse	Sum
N (%)	147 (29%)	61 (12%)	40 (8%)	211 (41%)	54 (10%)	513

## References

- Damico, N. (2017). 'Creepers' and Supporters: A Qualitative Analysis of the Practices of Beginning Educators in a Virtual Teacher Community Space. *Journal of Language and Literacy Education*, 13(2), 121. [http://jolle.coe.uga.edu/wpcontent/uploads/2017/11/Damico\\_JoLLE2017.pdf](http://jolle.coe.uga.edu/wpcontent/uploads/2017/11/Damico_JoLLE2017.pdf)
- Ghamrawi, N. (2022). Teachers' virtual communities of practice: A strong response in times of crisis or just another Fad? *Education and Information Technologies*, 27(5), 5889–5915. <https://doi.org/10.1007/s10639-021-10857-w>
- Lantz-Andersson, A., Lundin, M., & Selwyn, N. (2018). Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teaching and teacher education*, 75, 302-315.
- Lefstein, A., Vedder-Weiss, D., & Segal, A. (2020). Relocating research on teacher learning: Toward pedagogically productive talk. *Educational researcher*, 49(5), 360-368.
- Lieberman, A., & Pointer Mace, D. (2010). Making practice public: Teacher learning in the 21st century. *Journal of Teacher Education*, 61(1-2), 77-88. <https://doi.org/10.1177/0022487109347319>
- Riding, P. (2001). Online teacher communities and continuing professional development. *An international journal of teachers' professional development*, 5 (3), 283-296.
- Sudeck, Maria & Doolittle, Virginia & Rattigan, Peter. (2009). Professional development schools: Learning communities for leaders and teachers as change agents. In C. A. Mullen (Ed.), *The Handbook of Leadership and Professional Learning Communities* (pp. 51–58). Palgrave Macmillan US.

## Acknowledgments

We express our gratitude to The Center of Educational Technology for providing all the data at our disposal, and to our colleague who collaborated on the research.

## A Structural Equation Modeling Investigation of the Theory of Immersive Collaborative Learning (TICOL)

Valdemar Stenberdt, University of Copenhagen, valdemaras@psy.ku.dk  
Noëlle Bender, University of Copenhagen, noelle.bender21@icloud.com  
Adéla Plechatá, University of Copenhagen, adela.plechata@psy.ku.dk  
Guido Makransky, University of Copenhagen, gm@psy.ku.dk

**Abstract:** In this paper we tested the predictions of the Theory of Immersive Collaborative Learning (TICOL) through structural equation modeling (SEM) in a sample of 126 high school students learning about climate friendly, healthy cooking in an Extended Reality-supported Collaborative Learning (XRCL) environment. Students were randomly assigned a group and condition (learning in a park or apartment environment). SEM indicated that socioemotional, but not cognitive, interaction was important for learning in this context.

### Introduction

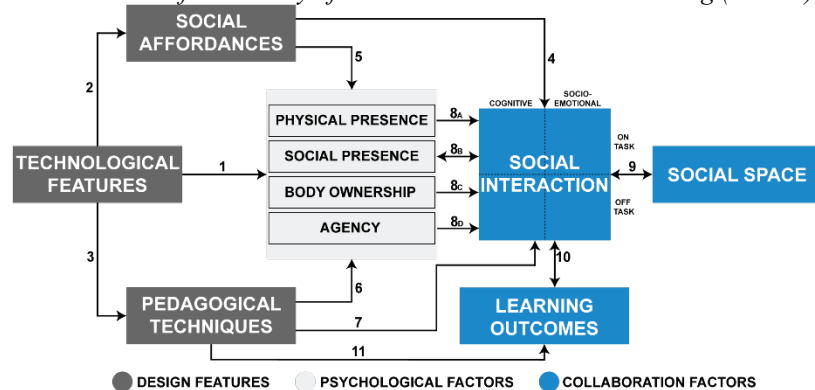
The proliferation of new IT platforms like the metaverse has spurred a host of new social applications that are accessed through head-mounted displays (HMD). This is expected to revolutionize the way we interact online (Mystakidis, 2022), and to ultimately influence how we collaborate, learn, and work. While video-based IT systems allow for distributed collaboration, they do not alter the perception of reality to the extent that they can influence the fundamental rules of human communication. In comparison, Extended Reality (XR) accessed through HMDs allows users to actively collaborate in physically and socially unconstrained environments that feel real (Slater, 2009). That is, they offer realistic experiences that are not restricted to the physical or social limitations of the physical world, allowing learners to instantaneously change who they are, where they are, who they are with, and what they can do (Makransky & Petersen, 2023).

### The state of the art

Collaborative learning is any situation in which people work together on the same task with the goal of learning something (Dillenbourg, 1999). XR is an umbrella term encapsulating technologies that enable an extension of reality (Pimentel et al., 2022). Using an HMD provides a vivid, multi-sensory experience which is highly immersive (Slater, 2009). XR-supported Collaborative Learning (XRCL) thus involves collaboration taking place through HMDs. The field of XRCL research is still in its infancy with only few studies conducted, a subset of which investigate learning (Han et al., 2022). Therefore, we turn to the fields of individual XR learning and Computer-Supported Collaborative Learning (CSCL). Meta-analyses have found a small effect size advantage for XR lessons compared to non-immersive lessons (Wu et al., 2020). Similarly, meta-analyses of CSCL research have found that collaboration has a significant positive effect on knowledge gain, skill acquisition, and student perceptions (Jeong et al., 2019). However, such outcomes are moderated by the educational level of learners, domains of learning, technology, design, and pedagogy (Jeong et al., 2019; Makransky, 2021). The key takeaway is that merely immersing learners in interactive settings or assembling them into groups with assigned tasks doesn't automatically ensure learning (Makransky & Petersen, 2023). Theoretically driven research evidence is needed to investigate when, how, and why the fundamental factors that make XRCL unique influence the quality of collaborative learning. The Theory of Immersive Collaborative Learning (TICOL; Makransky & Petersen, 2023) describes central processes and variables which are hypothesized to play a role in XRCL (see figure 1), described briefly below.

Technological features include hardware and software design. Social affordances are the social-contextual facilitators relevant for the learners' socio-emotional interactions. Pedagogical techniques refer to instructional design techniques and educational scaffolds. TICOL distinguishes between social interaction that occurs in the cognitive or socio-emotional dimension as well as on-task or off-task contexts. Moreover, TICOL describes how these types of social interactions are crucial for developing a positive social space which creates a feedback loop to social interaction and ultimately impacts learning outcomes. TICOL identifies four fundamental psychological factors which are central to XRCLs ability to support this: Physical presence refers to experiencing virtual physical objects as actual physical objects. Social presence refers to experiencing virtual social actors as actual social actors. Body ownership is the illusion that a virtual body belongs to oneself. Finally, agency is the sense of being the one who is causing or generating an action.

**Figure 1**  
*An Illustration of the Theory of Immersive Collaborative Learning (TICOL).*



## Method

The sample consisted of 126 first-year Danish high school students, average age = 17.5 (SD = 1.1), 62 female, 60 male, 1 other, and 3 prefer not to say. The data stems from a larger study described in Plechatá et al. (in preparation). The XR lesson revolved around climate friendly and healthy cooking. Students were randomized into a dyad and one of two conditions: collaborating in an apartment or a park. Students created two dishes and had access to either nutritional (fat, carb, and protein content) or environmental (estimated carbon emissions) information. After each dish, students received collective feedback and roles swapped. Additionally, in the park condition the environment changed according to student solutions. We measured physical presence (cronbach's alpha ( $\alpha$ ) = 0.74), social presence ( $\alpha$  = 0.74), social interaction ( $\alpha$  = 0.72), social space ( $\alpha$  = 0.84), and estimated CO<sub>2</sub> emissions of dishes as the learning outcome. To test the predictions of the TICOL, Structural Equation Modeling (SEM) was carried out in R using lavaan. The study was approved by the institutional review board.

## Results

The initial model fit was less than ideal (CFI = 0.853, RMSEA = 0.078, SRMR = 0.119). It was improved through three steps: Improving the measurement and structural models and removing insignificant paths. The final model fit was acceptable (CFI = 0.902, RMSEA = 0.064, SRMR = 0.079). The main finding was that, in the current sample, socioemotional interaction ( $\beta$  = -0.304,  $p$  = 0.031) predicted emissions, while cognitive did not ( $p$  = .630).

## References

- Dillenbourg, P. (1999). What do you mean by “collaborative learning”? In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches* (pp. 1–19). Elsevier.
- Han, E., Miller, M. R., Ram, N., Nowak, K. L., & Bailenson, J. N. (2022, May 14). Understanding Group Behavior in Virtual Reality: A Large-Scale, Longitudinal Study in the Metaverse. *72nd Annual International Communication Association Conference, Paris, France*.
- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review, 28*.
- Makransky, G. (2021). The Immersion Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (pp. 296–303). Cambridge University Press.
- Makransky, G., & Petersen, G. B. (2023). The Theory of Immersive Collaborative Learning (TICOL). *Educational Psychology Review, 35*(4), 103.
- Mystakidis, S. (2022). Metaverse. *Encyclopedia, 2*(1), 486–497.
- Pimentel, D., Fauville, G., Frazier, K., McGivney, E., Rosas, S., & Woolsey, E. (2022). *An Introduction to Learning in the Metaverse*. Meridian Treehouse.
- Plechatá, A., Morton, T., & Makransky, G. (2024). *Relevant Context Boosts Impact of Social Interaction on Identification with Environmentalists, Collective Action and Sustainable Eating: Insights from a Lab-in-the-Field Experiment* [Manuscript in preparation]. University of Copenhagen.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Phil. Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3549–3557.
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology, 51*(6), 1991–2005.

## Well-Being in CSCL Environments: Investigating Gender Variations

Khadija El Aadmi-Laamech, Patricia Santos, Davinia Hernández-Leo  
khadija.elaadmi@upf.edu, patricia.santos@upf.edu, davinia.hernandez-leo@upf.edu  
Department of Information and Communication Technologies, Pompeu Fabra University

**Abstract:** This paper assesses well-being perceptions among educators in STEMarium, an online collaborative community of teachers for STEM Learning Designs, through the lens of the Self-Determination Theory (SDT) with a focus on gender disparities. We analyze the responses of 44 Catalan educators. This study is relevant in the exploration of well-being aspects in computer-supported collaborative learning environments. It can have implications such as (1) better understanding the interactions and relationships among members of the community, and (2) considering gender differences as factors that may lead to distinct implications.

### Introduction

Collaborative environments enhance educators' practices by supporting online communities, enabling knowledge sharing, strategy exchange, mutual support, and collective inquiry, thereby enhancing teaching practices and promoting continuous professional development and innovation in education (Gutiérrez-Páez et al., 2023). Amidst this digital transformation, understanding the psychological underpinnings that contribute to user well-being is paramount (Nageswaran et al., 2022). Central to this understanding is the Self-Determination Theory (SDT), which posits that well-being is largely determined by the fulfillment of three basic psychological needs (BPNs): autonomy, competence, and relatedness (Ryan and Deci, 2017).

Despite the growing reliance on collaborative environments, a notable research gap persists regarding user well-being perceptions in these digital spaces –specifically, how users emotionally and psychologically engage with and are impacted by these systems (Nageswaran et al., 2022). Additionally, in this study we deepen the well-being analysis by adding a subjective layer of well-being perception (Batz and Tay, 2018): how gender perceptions of well-being influence (or not) the fulfillment of the basic psychological needs in collaborative environments. Gender differences have been a central theme in both well-being studies (García-Castilla et al., 2021) as well as collaborative learning research (Chan et al., 2013). We collect and base our results on the STEMarium platform, an ILDE-based (Integrated Learning Design Environment) online teacher community aimed to support learning design (LD) and teacher collaboration (Hernández-Leo et al., 2018; Gutiérrez-Páez et al., 2023). STEMarium is focused on sharing learning designs in the STEM field (Science, Technology, Engineering, and Mathematics). We formulate two research questions based on the initial arguments: RQ1) “How do educators of the STEMarium platform perceive their well-being regarding their use of the platform?”, and RQ2) “Are there gender differences regarding the well-being perceptions when using the STEMarium platform?”.

### Methodology

To carry out this study, we make use of the SDT-based instruments developed by Peters et al. (2018). Specifically, we make use of two questionnaires that evaluate two spheres of technology experience: interface and task. These are 1) Technology Need Satisfaction - Interface (TENS-Interface) and 2) TENS-Task. For both questionnaires we analyze the data by applying descriptive analysis: first, we apply a simple central tendency analysis for a first look at the data, segregating it by gender later on. Afterwards, and to detect any potential gender differences in the perception of the BPNs when using the STEMarium platform, we analyze through a correlation analysis the relationships between male and female responses for each questionnaire item, as well as item intercorrelation (i.e. all possible pairs of items). Significant correlations are then interpreted and discussed. The total number of participants is that of  $n=44$ , gender distribution is  $n=26$  females and  $n=18$  males.

### Results

*TENS-Interface.* (RQ1) The analysis reveals a moderately high perception of competence, autonomy, and relatedness among STEMarium platform users, with competence and autonomy being particularly strong. This suggests that users find the platform empowering and feel a sense of control over their interactions. (RQ2) The segregated data suggest that men who find STEMarium's interface confusing or hard to use are less likely to feel the technology is useful or meets their needs. However, women's perceptions of interface usability don't strongly impact their feelings about the technology's utility or intrusiveness. On the other hand, women are more likely than men to feel technology is controlling if they also find it intrusive. As for the social aspect of the platform, women seem to feel stronger connections between technology's ability to foster relationships and its overall

support for meaningful connections than men. *TENS-Task*. (RQ1) The aggregated data from the STEMarium platform users reveal that participants exhibit a moderate to high sense of competence and a high level of autonomy when engaging with tasks. This suggests that the task of creating and sharing resources successfully instills a sense of skill and volition. The scores for relatedness, though still moderate, are slightly lower than the former two, indicating a need to strengthen the sense of community and support while performing the task. (RQ2) Competence: Both men and women report feeling competent, with men showing a higher average score than women. Women's competence scores, while still above the midpoint, suggest room for enhancing their perceived efficacy on the task of creating and sharing resources. Women also show interconnectedness between feeling competent and autonomous, suggesting that these constructs may reinforce each other more for women than for men. Autonomy: High scores from both genders indicate a strong sense of autonomy. This reflects a collaborative online environment that provides space for independent and self-directed engagement. Relatedness: The levels of relatedness for both genders highlight an area for potential development to foster a stronger community feel on the platform when creating and sharing resources, with relatedness being a central need for collaborative environments.

## Implications and conclusions

*TENS-Interface*. While both genders exhibit positive perceptions across the BPNs, the gender differences in the correlations and consistencies of responses suggest that male and female users may experience and interact with the STEMarium platform differently: males seem to emphasize the functionality and efficacy aspects, whereas females appear more attuned to the platform's social and relational dimensions. This could suggest that the incentives to be offered to encourage the use of the platform would benefit from understanding these behaviors. *TENS-Task*. The significant differences between genders in how certain emotions and task perceptions are correlated highlight some important findings: 1) Women's sense of closeness to others is significantly related to how meaningful they find the connections supported by the task (e.g. coauthoring in the creation of resources), a relationship less evident for men. 2) Guilt appears to undermine women's confidence in their ability to perform the task, whereas for men, confidence remains unaffected by such feelings. 3) Women's guilt-driven motivation impacts their perception of the task's social support, contrasting with the negligible effect on men.

Understanding these gender-specific patterns is important for designing more personalized user experiences like incentives and automatic recommendations (Gutiérrez-Páez et al., 2023). For example, interventions to reduce feelings of guilt may be particularly beneficial for women, while strategies to enhance men's sense of community could lead to a more balanced engagement with the platform.

## Limitations and future work

The present limitations of this study include the limited sample size; consequently, subsequent research should aim to expand on this aspect. It should also be noted that the study does not account for other demographic variables that may influence well-being perceptions. Further research also calls for conducting the same analysis on other platforms of the same environment as STEMarium (i.e. ILDE), for potential generalizability of results.

## References

- Batz, C., & Tay, L. (2018). Gender differences in subjective well-being. *Handbook of well-being*, 1-15.
- Chan, R., Huang, J., Hui, D., Li, S., & Yu, P. (2013). Gender differences in collaborative learning over online social networks: Epistemological beliefs and behaviors. *Knowledge Management & E-Learning: An International Journal*, 5, 234-250.
- García-Castilla, F., Martínez-Sánchez, I., Campos, G., & Resino, D. (2020). Impact of Gender and Relationship Status on Young People's Autonomy and Psychological Wellbeing. *Frontiers in Psychology*, 11.
- Gutiérrez-Páez, N. F., Santos, P., Hernández-Leo, D., & Carrió, M. (2023). A study of motivations, behavior, and contributions quality in online communities of teachers: A data analytics approach. *Computers & Education*, 201, 104829.
- Hernández-Leo, D., Asensio-Pérez, J. I., Derntl, M., Pozzi, F., Chacón, J., Prieto, L. P., & Persico, D. (2018). An integrated environment for learning design. *Frontiers in ICT*, 5, 9.
- Nageswaran, P., Leedham-Green, K., Nageswaran, H., & Baptista, A. (2022). Digital wellbeing: Are educational institutions paying enough attention?. *Medical Education*, 57, 216 - 218.
- Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in psychology*, 797.
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

# Multimodal Observations of Pair Programming Processes in a Shared Visual Workspace

Siem Buseyne, KU Leuven & University of Lille, siem.buseyne@kuleuven.be  
 Fien Depaepe, KU Leuven, fien.depaepe@kuleuven.be  
 Annelies Raes, KU Leuven, annelies.raes@kuleuven.be

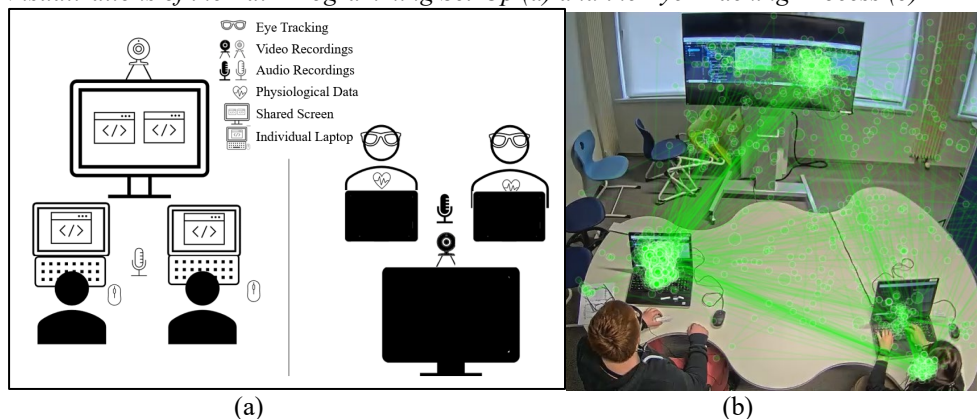
**Abstract:** Previous research has identified promising avenues for investigating collaborative processes of pair programming, a type of computer-supported collaborative learning. In our current research, we aim to advance our comprehension of these dynamics, specifically within a shared visual workspace. To achieve this, we utilize multimodal learning analytics, emphasizing the integration of verbal data and eye-tracking data, to provide more profound insights into this collaborative programming environment.

## Introduction

Collaborative problem solving (CPS) competencies are increasingly recognized as crucial for today's workforce and society. One CPS approach, known as pair programming, can enhance both soft and technical skills. Pair programming is defined by Ally et al. (2005) as a collaborative process where two programmers work together, tackling various tasks, including analysis, design, implementation, and testing. This practice typically involves two primary roles: the driver, responsible for writing the code, and the navigator, who reviews the code, offers feedback, and provides guidance to the driver. Pair programming is traditionally executed on a single computer or workstation, where participants share the screen, keyboard, mouse, and other devices while programming (Ally et al., 2005; da Silva Estácio & Prikladnicki, 2015). Over the years, various adaptations of this method have emerged, including distributed pair programming and the use of shared workspaces, such as shared large screens. Research has highlighted the benefits of these shared workspaces for fostering collaborative engagement (Chen et al., 2021). However, limited research has explored pair programming within educational contexts using shared workspaces.

Much of the existing research on pair programming has concentrated on factors influencing its outcomes. For instance, studies have compared pair programming with individual programming (Demir & Seferoglu, 2021). To a lesser extent, research has delved into the processes of pair programming and how various factors, such as group composition, task complexity, and technology use, influence these processes (Walle & Hannay, 2009). Furthermore, most research on pair programming processes has relied on unimodal data sources, predominantly observations and content analysis of verbal communication to extract specific patterns and behaviors. However, emerging research suggests the need to incorporate additional indicators and a combination of data sources through multimodal learning analytics. Jermann and Nüssli (2012), for example, highlighted the necessity to establish measurable and automated assessments of behavioral indicators signifying a qualitative interaction during pair programming.

**Figure 1**  
*Visualizations of the Pair Programming Set-Up (a) and the Eye-Tracking Process (b)*





One innovative avenue for investigating pair-programming processes is through eye tracking, next to audio data. While most research in this area has relied on screen-based eye trackers, these devices have limitations, as they primarily focus on screen use. In collaborative contexts, researchers are interested in various aspects beyond the screen, including interaction with other team members. Therefore, the adoption of mobile eye tracking devices for analyzing pair programming processes holds promise, as it enables the study of multiple participants' fixations on different areas of interest and their joint attention (Sharma et al., 2021), which is the focus of this study.

## Methodology

Participants comprised 20 students aged 16-18 from Flanders. To be eligible, participants needed prior experience with programming languages (i.e., HTML, CSS, JavaScript). We excluded individuals with advanced programming skills to maintain a challenging task. During data collection, students had to code the "Guess Who" game following pair programming and agile principles. The coding process was divided into six iterations, each with a specific task. Each iteration had four phases: thinking, programming, testing, and problem-solving. After each iteration, students swapped roles, ensuring that both participants took the role of navigator at some point in the task. As visualized in Figure 1, pairs were video- and audio recorded and a selection of three groups wore mobile eye trackers during the experiment. Students' conversations were transcribed and analyzed using a CPS coding scheme (Sun et al., 2020). Via the eye-tracking data, measures of joint attention were, among other processed. Furthermore, teams were asked to fill out a questionnaire after the collaborative process for assessing the quality of CPS processes. Specifically, for doing so, a questionnaire was designed based on the rating scheme for assessing the quality of CSCL processes (Meier et al., 2007).

## Outlook

Findings of our research will be presented during the conference. We would also like to discuss further research directions for enhancing pair programming processes.

## References

- Ally, M., Darroch, F., & Toleman, M. (2005). A Framework for Understanding the Factors Influencing Pair Programming Success. In H. Baumeister, M. Marchesi, & M. Holcombe (Eds.), *Extreme Programming and agile Processes in Software Engineering*. Springer. [https://doi.org/10.1007/11499053\\_10](https://doi.org/10.1007/11499053_10)
- Chen, L., Liang, H. N., Wang, J., Qu, Y., & Yue, Y. (2021). On the use of large interactive displays to support collaborative engagement and visual exploratory tasks. *Sensors*, *21*(24). <https://doi.org/10.3390/s21248403>
- da Silva Estácio, B. J., & Prikladnicki, R. (2015). Distributed Pair Programming: A Systematic Literature Review. *Information and Software Technology*, *63*. <https://doi.org/10.1016/j.infsof.2015.02.011>
- Demir, Ö., & Seferoglu, S. S. (2021). A Comparison of Solo and Pair Programming in Terms of Flow Experience, Coding Quality, and Coding Achievement. *Journal of Educational Computing Research*, *58*(8), 1448–1466. <https://doi.org/10.1177/0735633120949788>
- Jermann, P., & Nüssli, M.-A. (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 1125–1134. <https://doi.org/10.1145/2145204.2145371>
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, *2*(1), 63–86. <https://doi.org/10.1007/s11412-006-9005-x>
- Sharma, K., Olsen, J. K., Verma, H., Caballero, D., & Jermann, P. (2021). *Challenging Joint Visual Attention as a Proxy for Collaborative Performance*. 91–98.
- Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education*, *143*. <https://doi.org/10.1016/j.compedu.2019.103672>
- Walle, T., & Hannay, J. E. (2009). Personality and the Nature of Collaboration in Pair Programming. *Third International Symposium on Empirical Software Engineering and Measurement*.

## Acknowledgements

This study was with the financial support of imec and Flanders Innovation & Entrepreneurship. We extend our sincere gratitude to Robin De Croon and Olly Clerx for their support in this research project.

# Task Complexity and Split Attention in Individual and Collaborative Learning

John Guzmán, Unidad Educativa Comunitaria Intercultural Bilingüe Rumiñahui, guzmanescal@gmail.com  
Jimmy Zambrano R., Universidad del Pacífico, jimmy.zambrano@upacifico.edu.ec

**Abstract:** We aimed to determine the effects of split attention and task complexity on the individual and collaborative learning of high school students. An experiment with 192 students revealed that when students work with low-complexity problems and integrated information, individual students learn more than students working in groups. However, when students learn high-complexity and split information, group performance is better than individual performance. The study pretends to give suggestions for educational practice.

## Introduction

Limitations of novice working memory can pose difficulties in processing mathematical tasks (Retnowati, et al., 2016). When there are two or more sources of information that need to be processed simultaneously to understand the task, students may experience a lack of understanding or poor learning. This leads to a phenomenon known as split attention (Ayres, et al., 2014). This occurs due to the physical separation of information sources, requiring more attentional resources that diminish learning. Another effect of cognitive load related to group learning is the collective working memory (Kirschner, et al., 2011). This occurs when students learning in a group integrate their mental resources to process a complex task. When the task is simple, it is better to learn individually. Much is still unknown about the interaction of these two effects in mathematics learning tasks with different levels of complexity (i.e., the level of interactivity of the element). Consequently, the objective of this study was to explore how the interactivity of the element and split attention affect individual and collaborative learning tasks. Split-attention is an extraneous load that has been studied by various authors with the aim of exploring the impacts it can have on school learning. Despite the extensive study of split attention effects, its impact on group learning has not been explored. Collaborative learning as an instructional method that involves two or more learners collaborating to achieve a shared learning goal. Transactional activities are elements of information external to the learning tasks and induces extraneous cognitive load. This load can be beneficial when interacting with other factors such as high task complexity (Kirschner, et al., 2011) or having prior collaborative experience based on relevant tasks (Zambrano R. et al., 2023).

Our research questions were to what extent does material inducing integrated attention benefit individual students more than group students in low complexity tasks? How does material inducing separated attention affect individual students compared to group students in low complexity tasks? What differences are observed in performance between individual and group students when using material with integrated information in tasks with high element interactivity? What is the impact of material inducing separated attention on the performance of individual students compared to group students in tasks with high element interactivity?

## Method

### Participants

One hundred ninety-two students from a public high school in Guayaquil, Ecuador, participated in the study as part of their mathematics classes (97 male students; 95 female students). They were randomly assigned to the learning conditions. Participation authorized by the school authorities, and all students were novices in the learning tasks.

### Design, procedure and material

A factorial design of 2 (*element interactivity*: simple vs. complex) x 2 (*material type*: separated attention vs. integrated attention) x 2 (*social condition*: individual vs. group) was employed. The dependent variable was performance in mathematics. The materials were within the domain of mathematics. During the learning stage, students received a booklet containing the developed topic of complex numbers with four solved exercises. Performance for participants working with high element interactivity was scored out of 13 points. In the immediate test, participants received a booklet with four tasks to solve one day after the learning phase. Both in the pretest and immediate test, participants received the booklet with learning activities containing the procedure for addition, subtraction, multiplication, and division operations of complex numbers with the same level of complexity similar

to the learning tasks, along with their respective assessments. Participants completed these tasks independently, and they were instructed to solve them in the provided booklet.

## Results

A design 2 (*element interactivity*: high vs. low) x 2 (*social condition*: individual vs. group) x 2 (material type: separated attention vs. integrated attention) analysis of variance (ANOVA) was employed.

In the learning phase, regarding performance with low element interactivity tasks, the ANOVA revealed that material was statistically significant in favor of the integrated information,  $M = 6.01$ ,  $SD = .47$ ,  $F(1, 92) = 13.55$ ,  $MSE = .50$ ,  $p < .001$ ,  $\eta_p^2 = .13$ , compared to separated information,  $M = 5.45$ ,  $SD = .90$ . Social condition was also significant,  $F(1, 92) = 5.05$ ,  $MSE = .50$ ,  $p = .03$ ,  $\eta_p^2 = .05$ , suggesting that the performance of groups,  $M = 5.84$ ,  $SD = .79$ , was better than individual performance,  $M = 5.50$ ,  $SD = .67$ . The interaction of main effects was not significant. Regarding performance in tasks with high element interactivity, the material and social condition were not significant. However, the interaction of main effects was significant,  $F(1, 92) = 28.96$ ,  $MSE = 3.43$ ,  $p < .01$ ,  $\eta_p^2 = 0.24$ . Bonferroni post-hoc tests suggest that the material with integrated information benefits the individual condition more than the groups,  $p < .01$ ,  $\eta_p^2 = 0.18$ , and that separated information benefits the group condition more than the individuals,  $p < .01$ ,  $\eta_p^2 = 0.09$ .

In the immediate post-test, only the main effect of element interactivity was significant,  $F(1, 184) = 29.77$ ,  $p < .01$ ,  $\eta_p^2 = 0.14$ , suggesting that low element interactivity,  $M = 9.92$ ,  $SD = .31$ , resulted in better performance compared to high element interactivity,  $M = 7.5$ ,  $SD = .31$ . The remaining main effects were not statistically significant. Regarding the significant interactions of main effects, Bonferroni tests for the interaction between element interactivity and social condition indicated that in low element interactivity tasks, individual learning,  $p < .01$ ,  $\eta_p^2 = .20$ ,  $M = 12.05$ ,  $SD = .51$ , was more effective than group learning,  $M = 7.80$ ,  $SD = .36$ . However, in high element interactivity tasks, group learning,  $p < .01$ ,  $\eta_p^2 = .08$ ,  $M = 8.78$ ,  $SD = .36$ , is better than individual learning,  $M = 6.22$ ,  $SD = .51$ . Bonferroni tests for the interaction between material and social condition indicated that material with integrated information benefits individual students more,  $p < .01$ ,  $\eta_p^2 = .08$ ,  $M = 9.83$ ,  $SD = .51$ , than groups,  $M = 7.41$ ,  $SD = .36$ . However, material with separated information benefits both learning conditions equally,  $p = 2.4$ . Bonferroni tests regarding the significant interaction between element interactivity, social condition, and material suggest that tasks with low element interactivity and integrated material,  $p < .01$ ,  $\eta_p^2 = .11$ , or separated material,  $p < .01$ ,  $\eta_p^2 = .11$ , benefit individual students more than groups (Table 3). Tasks with high element interactivity and integrated material benefited individual students and groups equally,  $p = .45$ . However, separated material proved to be better for group students,  $p < .01$ ,  $\eta_p^2 = .19$ .

## Discussion

Our study provides data suggesting that the advantage of individual and group learning is mediated by the characteristics of the learning task and study material. In tasks of low complexity, such as solved examples, it is not relevant to learn individually or in groups with materials that induce divided attention. However, when the task imposes a high cognitive load, for example, incomplete problem-solving tasks, and the material does not impose divided attention, it is better to learn individually. If the material induces divided attention, it is better to learn in small groups. These findings can help guide instructional decisions for teachers. Teachers can decide whether to use an individual or collaborative learning format based on the complexity of the task and the level of information integration.

## References

- Ayres, P., Sweller, J., & Mayer, R. (2014). The split-attention principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning* (pp. 206-226). Cambridge University Press. <https://doi.org/10.1017/cbo9781139547369.011>
- Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, 25(4), 615-624. <https://doi.org/10.1002/acp.1730>
- Retnowati, E., Ayres, P., & Sweller, J. (2016). Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *Journal of Educational Psychology*, 109(5), 666-679. <https://doi.org/10.1037/edu0000167>
- Zambrano R., J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2023). Effect of task-based group experience on collaborative learning: Exploring the transaction activities. *British Journal of Educational Psychology*, 879-902. <https://doi.org/10.1111/bjep.12603>

# Investigating the Knowledge Building Process of Collaborative Learning between Student and Virtual Tutor Supported by ChatGPT: A Discourse Analysis

Xiaoyan Chu, Zhejiang University, University of Manchester, xiaoyanchu@zju.edu.cn  
 Jiaqi Xu, Zhejiang University, University College London, jiaqi.xu@zju.edu.cn  
 Xuesong Zhai, Zhejiang University, xszhai@zju.edu.cn

**Abstract:** The advent of generative artificial intelligence (AI) such as ChatGPT has brought forth new possibilities in education, leading to further advancements in the design and application of virtual tutors. This study examined the impact of collaborative learning with ChatGPT virtual tutors on academic performance among 94 graduate students. Utilizing the Tech-SEDA framework, discourse analysis was conducted on 344 dialogue data entries. Results indicate that while learners engage in knowledge building with generative AI involving real-world tasks, the predominant mode of collaboration remains focused on information acquisition.

## Introduction

In recent years, there has been a rapid evolution of Artificial Intelligence (AI) in educational contexts (Zhai et al., 2021). Cutting-edge generative AI demonstrates superior capabilities in generating resources and providing learning guidance, positioning itself as a potential virtual tutor for students (Lo, 2023). Virtual intelligent tutors now assume an increasingly indispensable role in online collaborative learning, facilitating enhanced human-computer interaction (Haq et al., 2020). Kuhail et al. (2023) elucidated the supportive role of educational chatbots in collaborative learning through a systematic review of empirical studies. Online collaborative learning environments offer learners opportunities to actively engage and interact in knowledge building, thereby fostering the development of meaningful knowledge structures through collaborative efforts (Yücel & Usluel, 2016). Moreover, the functionalities of generative artificial intelligence align with the core tenets of knowledge building theory, including "Idea Improvements" and "Knowledge-Creating Dialogue" (Scardamalia & Bereiter, 2022).

Despite these advantages, there remains a lack of clarity regarding how learners engage in knowledge building while collaborating with a ChatGPT-supported virtual tutor. This study undertakes an empirical investigation into collaborative learning with a ChatGPT virtual tutor within a postgraduate program, aiming to elucidate the impact of this instructional approach on learners' outcomes and the dynamics of the knowledge building process. Therefore, this study addresses two primary research questions: 1) What impact does collaborative knowledge building with ChatGPT have on learners' learning outcomes? 2) What are the characteristics of collaborative knowledge building process between learners and ChatGPT?

## Methodology

This study involved 94 postgraduate students from a Chinese university. A self-developed learning platform using Unity 3D and powered by GPT-3.5 served as the foundation for this investigation. Dialogue data generated by students was logged in the backend system for thorough analysis. The course duration extended over 32 hours, during which student learning outcomes were evaluated through both pre- and post-course assessments. To ensure consistent difficulty levels, test questions were sourced from a standardized question bank. Adhering to Hennessy's (2023) technology-mediated educational dialogue coding framework, two independent coders analyzed the dialogue data, achieving an inter-rater reliability score exceeding 0.86. Any discrepancies were resolved through discussion.

## Preliminary results and discussion

The results of a paired-samples *t*-test conducted on pre- and post-tests demonstrated a significant difference in both retention knowledge ( $MD=0.131, t=4.694, p<0.01$ ) and transfer knowledge ( $MD=0.174, t=4.322, p<0.01$ ). Furthermore, post-test scores exhibited a statistically significant increase compared to pre-test scores. These findings indicate that collaborative knowledge building with ChatGPT virtual tutors may positively influence learners' academic performance to a considerable extent.

**Table 1**  
*Categories and the Number of Each Coding Item*

Category	Subcategory	Number
----------	-------------	--------

Build	IB - Invitation to build on ideas	208
	B - Building on ideas	0
Challenge	CH - Challenge	11
Reason	IR - Invitation for reasoning	42
	R - Reasoning	0
Invitation	IC - Invitation for co-ordination	2
Co-ordination	SC - Simple co-ordination	0
	RC - Reasoned co-ordination	0
Inquiry	II - Inquiry invitation	19
Back	RB - Reference back	4
Wider context	RW - Reference to wider context	4
Focusing	F - Focusing	53
Reflect	RD - Reflect on dialogue or activity	1

As illustrated in Table 1, learners predominantly engage in collaborative knowledge building with ChatGPT tutors at a foundational level. This involves inviting the tutor to contribute to idea development, engage in reasoning, and explore further inquiries. One plausible explanation for this trend is the learners' perception of generative AI, like ChatGPT, primarily as a tool for gathering information rather than as a genuine learning companion. Learners frequently extend invitations to the ChatGPT virtual tutor to generate new ideas and participate in reasoning and interpretation. ChatGPT's partnership with learners is consistent with the knowledge-building concepts of "Knowledge-Creating Dialogue" and "Idea Improvements" (Scardamalia & Bereiter, 2022). In comparison with the findings of Kuhail et al. (2023), which indicate that current interactions between learners and chatbots are primarily driven by the chatbots, this study suggests that when collaborating with generative AI, greater learner agency or additional pedagogical interventions may be more necessary.

### Limitations and future research

This study provides an initial exploration into the impact of ChatGPT virtual tutors on learners' academic performance and the facilitation of knowledge building within collaborative learning environments. However, the absence of a group control method in this study limits the extent of conclusions that can be drawn. Nonetheless, it offers valuable insights into the collaborative knowledge-building process between learners and generative artificial intelligence. The integration of generative AI in education presents opportunities for further investigation, including the implementation of prompt engineering to empower intelligent agents in executing more complex learning assistance tasks, thereby fostering deeper human-computer learning interactions. Additionally, to enhance the richness of the study's data sources, the inclusion of supplementary behavioral data, beyond discourse data alone, is recommended for a comprehensive analysis of the learners' evolving learning processes.

### References

- Haq, I. U., Anwar, A., Basharat, I., & Sultan, K. (2020). Intelligent tutoring supported collaborative learning (itscl): a hybrid framework. *International Journal of Advanced Computer Science and Applications*, 11(8): 523-535.
- Hennessy, S. (2023). Analysing educational dialogue around shared artefacts in technology-mediated contexts: A new coding framework. *EdArXiv Preprints*. <https://osf.io/preprints/edarxiv/9e32k>.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018.
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410.
- Scardamalia, M., & Bereiter, C. (2022). Knowledge Building and Knowledge Creation. In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (Cambridge Handbooks in Psychology, pp. 385-405). Cambridge: Cambridge University Press.
- Yücel, Ü. A., & Usluel, Y. K. (2016). Knowledge building and the quantity, content and quality of the interaction and participation of students in an online collaborative learning environment. *Computers & Education*, 97, 31-48.
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021, 1-18.

### Acknowledgments

This research was funded by the CSC project 202306320092.

# Exploring the Link Between Personality Traits and Verbal Interactions in Computer-Supported Collaborative Problem Solving

Siem Buseyne, KU Leuven & Université de Lille, siem.buseyne@kuleuven.be

Amelie Vrijdags, Hudson Belgium, amelie.vrijdags@hudsonsolutions.com

Sameh Said-Metwaly, KU Leuven & Damanhour University, sameh.metwaly@kuleuven.be

Thierry Danquigny, Université de Lille, thierry.danquigny@univ-lille.fr

Jean Heutte, Université de Lille, jean.heutte@univ-lille.fr

Fien Depaepe, KU Leuven, fien.depaepe@kuleuven.be

Annelies Raes, KU Leuven, annelies.raes@kuleuven.be

**Abstract:** This study aims to investigate the interplay between personality traits and verbal behavior within the context of computer-supported collaborative problem solving (CPS). To address this, audio data were collected from nine professional teams during CPS tasks as part of a training. Each team comprised four members, originating from various private and public organizations in Flanders, Belgium. Personality assessments were conducted using the Business Attitudes Questionnaire. Using the audio data, measures of verbal interactions were processed using (a) content analysis based on a coding scheme for computer-supported CPS and (b) linguistic inquiry and word count. The results of this research provide first insights into the influence of personality traits on team's verbal interactions in CPS processes.

## Introduction

The past few years have seen a surge in interest in collaborative problem solving (CPS), a collaborative process in which two or more individuals work together to address a problem. This collaborative effort involves the sharing of knowledge, skills, and resources with the goal of collectively reaching a solution. Fiore et al. (2010) discern three main CPS processes: (a) establishing and maintaining shared understanding; (b) taking appropriate action to solve the problem; and (c) establishing and maintaining team organization. CPS is considered a key area in the computer-supported collaborative learning (CSCL) research field, which has gained increasing attention due to rapid advancements in educational technology. These innovations have transformed how collaborative processes can be studied and analyzed. In research on CPS, this includes the analysis of verbal and nonverbal communicative aspects. Particularly, advancements in artificial intelligence and natural language processing have enabled the adoption of advanced (semi-)automated processing, such as automatic speech recognition and facial emotion recognition. Recently, there has been a growing focus on how personality traits of team members influence CSCL (Kucukozer-Cavdar & Taskaya-Temizel, 2016), particularly how personality affects how well teams perform. For example, in line with previous research (Curşeu et al., 2019), extraversion, conscientiousness, and agreeableness could demonstrate significant relevance for teamwork due to their associations with both task-related and positive interpersonal aspects. However, there is a lack of research on this topic within the specific context of CPS. This study aims to address this gap by examining the connection between team members' personalities and how they communicate verbally in CPS settings.

## Method

This research took place in the context of a nonformal CPS training for adult teams (Buseyne, Vrijdags, et al., 2023). The study was approved by the Data Protection Officer and the Ethical Committee of KU Leuven, Belgium. Participants were 21 men and 15 women, drawn from nine pre-existing teams comprising a mix of private and public organizations in the Flemish region. They had diverse backgrounds and occupied various job roles.

Prior to the training, participants were given the task of completing the Business Attitudes Questionnaire (BAQ) developed by Vrijdags et al. (2014). The BAQ gauges four facets of personality for each of the Big Five domains (i.e., emotional stability, extraversion, openness, altruism, and conscientiousness) and additionally evaluates five compound personality traits categorized under the Professionalism label (i.e., Ambitious, Critical, Results-oriented, Strategic, and Autonomous). Following the same protocol, multiple training sessions were conducted in an educational lab setting. During the second part of this training, which included a 30-minute CPS task, team-level discussions were audio- and video-recorded. The audio recordings were verbatim transcribed. Next, to evaluate the affective, social, and cognitive dimensions of participants' interactions during each phase of the task, Linguistic Inquiry and Word Count (LIWC) was employed (Pennebaker et al., 2015). Subsequently, participants' utterances were annotated using a coding scheme for computer-supported CPS (see Buseyne,

Rajagopal, et al., 2023), which is based on Sun et al. (2020). Interrater reliability, calculated on part of the data for two independent raters, was sufficiently high for the aggregated sub-category level ( $\kappa = .79$ ).

The data analyses were done using R (Version 4.3.1). To examine the link between personality traits and indicators of verbal interactions, multilevel linear regression analyses were performed using the nlme package, based on maximum likelihood. Model selection for the analyses were performed via stepwise elimination based on a likelihood-ratio test using the buildmer package (version 2.11). The model considered the repeated measurements of each dependent variable per individual by incorporating the individual as level 2. Group and task variables were included as covariates in all models.

## Selected results

The results of the regression analyses after backward elimination reveal various relationships between personality traits and the categories and sub-categories of CPS utterances. For example, a significant positive association was found between Ambitious and "Establishing, Constructing, and Maintaining Shared Knowledge and Understanding" ( $b = 6.27, SE = 2.40, p < .05$ ). For the link between personality traits and the selected LIWC categories, results showed, among others, that persons scoring higher on Result-Oriented had a higher absolute word-count ( $b = 63.50, SE = 14.86, p < .001$ ) and more Strategic team members exhibited a higher use of cognitive process words ( $b = 2.10, SE = 0.67, p < .01$ ). These results are partially in line with previous research (e.g., Tasa et al., 2011).

## Conclusion

Several aspects may hinder the generalizability of our findings. Specifically, it is important to consider that this study was done with a limited sample size, in a specific CPS context, and with adults from different backgrounds. Therefore, additional research is needed to assess these relationships across different contexts. Next, our current study examined the individual effects of personality traits on CPS interactions. Future research should investigate potential interactions among different traits to gain a better understanding.

In conclusion, our study provides an innovative approach for assessing the link between personality traits and verbal interactions within CPS, through the integration of several methods (e.g., LIWC). The findings offer valuable insights for organizations and teams seeking to optimize CPS processes and harness the strengths of team members with diverse personality traits. By getting insights into how personality traits influence different aspects of CPS interactions, stakeholders (e.g., trainers) could further optimize team and CPS processes.

## References

- Buseyne, S., Rajagopal, K., Danquigny, T., Depaepe, F., Heutte, J., & Raes, A. (2023). Assessing verbal interaction of adult learners in computer-supported collaborative problem solving. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13391>
- Buseyne, S., Vrijdags, A., & Raes, A. (2023). Productive Failure as a Method for Learning about Effective Collaborative Problem Solving. *International Journal of Designs for Learning*, 14(1), 46–61. <https://doi.org/10.14434/ijdl.v14i1.35221>
- Curşeu, P. L., Ilies, R., Vîrgă, D., Maricuţoiu, L., & Sava, F. A. (2019). Personality characteristics that are valued in teams: Not always “more is better”? *International Journal of Psychology*, 54(5), 638–649. <https://doi.org/10.1002/ijop.12511>
- Fiore, S. M., Smith-Jentsch, K. A., Salas, E., Warner, N., & Letsky, M. (2010). Towards an understanding of macrocognition in teams: Developing and defining complex collaborative processes and products. *Theoretical Issues in Ergonomics Science*, 11(4), 250–271. <https://doi.org/10.1080/14639221003729128>
- Kucukozer-Cavdar, S., & Taskaya-Temizel, T. (2016). Analyzing the Effects of the Personality Traits on the Success of Online Collaborative Groups. *Procedia - Social and Behavioral Sciences*, 228, 383–389. <https://doi.org/10.1016/j.sbspro.2016.07.057>
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015* (2015). Pennebaker Conglomerates. [www.LIWC.net](http://www.LIWC.net)
- Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D’Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education*, 143. <https://doi.org/10.1016/j.compedu.2019.103672>
- Tasa, K., Sears, G. J., & Schat, A. C. H. (2011). Personality and teamwork behavior in context: The cross-level moderating role of collective efficacy. *Journal of Organizational Behavior*, 32(1), 65–85.
- Vrijdags, A., Bogaert, J., Trbovic, N., & Van Keer, E. (2014). *Business Attitudes Questionnaire (Psychometric technical manual)*.

## Collaborative Mayhem: Supporting Elementary Students Social-emotional Learning in Engineering Design

Corey Schimpf, University at Buffalo, schimpf2@buffalo.edu  
Askari Medhi, University at Buffalo, smehdi@buffalo.edu  
Linda Smith, Explore Interactive, lsmith@exploresupport.com  
Amanda Thompson, Explore Interactive, athompson@exploresupport.com

**Abstract:** Engineering design is increasingly used to support science and engineering learning. However, the open-ended nature of design projects and team collaboration can be difficult for elementary students to navigate. In this work we present a conjecture map for how a new augmented reality app may be able to support students social-emotional learning. The app features affective state reporting and team feedback. We present hypothesized outcomes and how we plan to measure them.

### Introduction

The push to integrate engineering education in grades K–12, including in K-5 elementary schools, has increased dramatically since the introduction of the Next Generation Science Standards. Elementary school teachers have expressed several challenges, including insufficient resources, students' lack of readiness for collaborative work, and academic time constraints (Haverly et al., 2022; Radloff & Capobianco, 2019). Collaborative teamwork, in particular, represents a steep challenge as this work can evoke strong emotions when working on open-ended problems with multiple people who may have different ideas than oneself. In this space, we introduce MindLabs (ML), an augmented reality platform to support students to engage in engineering design challenges around science topical areas, e.g., forces through simple machines. Newly released features include student's ability to share affective states and team member feedback. We present a conjecture map (Sandoval, 2014) for how we believe this environment can support students' social-emotional learning (SEL) and chart our initial plan for measuring how ML impacts students. Our conjecture is: *Engineering Design requires collaboration, which is emotionally taxing. Providing ways for students to share discomfort and approval gives them voice and supports SEL development.*

### Background

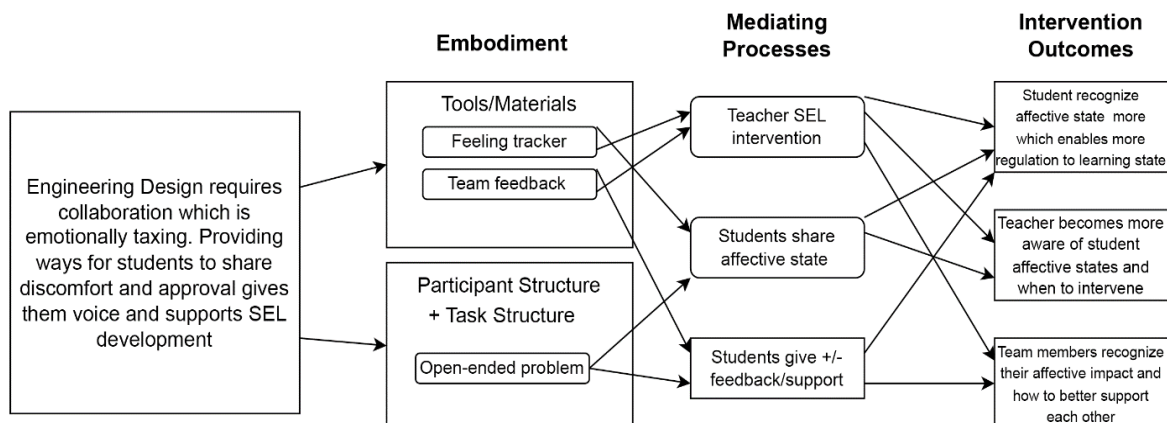
Although emotions and emotional regulation have been studied in many areas of education (Järvenoja et al., 2020), the intersection of emotions and design remains understudied. Main areas in this literature include user emotions in design and design team emotions during collaborations. For example, in user emotions Yoon et al. (2020) highlight the importance of designing for user emotions to support user wellbeing. For an example of design team emotions, Sozo and Ogliari (2019) found emotional climate affects team performance, with emotions directly affecting project success and team satisfaction. Social-emotional learning (SEL) involves the development of self-awareness, self-regulation, and interpersonal competencies, which are seen as essential for achieving success in educational, occupational, and personal contexts for children and adults (CASEL, 2022). CASEL notes “SEL can help all young people and adults thrive personally and academically, develop and maintain positive relationships, become lifelong learners, and contribute to a more caring, just world” (CASEL, 2022).

### Results

Figure 1 displays our conjecture map; we highlight its major features. The embodiment, tools and materials include a feelings tracker for students to report their affective state from sixteen states grouped into four categories (e.g., happy, sad, worried, and frustrated feelings) and a system for giving team members feedback on their collaboration. The mediating processes we believe these embodiments support are teachers SEL-related interventions with teams, students sharing their affective state, and student feedback to teammates through their design process. We believe that through students reporting their affective states, they will develop greater recognition of how affect emerges in learning and start to notice what contextual factors are related to changes in their affective state, e.g., teammate behavior. This will be further reinforced through the teacher addressing student emotions with the teams. Moreover, by receiving timely updates on students affective reporting, teachers will be able to more fully recognize students' affective dynamics, particularly for emotions that are not outwardly expressed.



**Figure 1**  
*Student Team SEL Conjecture Map*



We briefly describe our plan to measure the mediating processes and intervention outcomes to understand how this conjecture may unfold in practice. First, both students' affective states reporting and sharing feedback with their teammates, will be captured by the platform. Data will be stored in a CSV format for analysis. Teacher interventions will be captured through class observations and interviews. These will allow us to naturalistically identify how teachers intervene with teams and will be bolstered by best practices from the literature. The degree to which the outcomes for the teacher are met will be measured through teacher interviews, focusing on their evaluation of the tools and of their own intervention strategies. Lastly, changes in individual or team member recognition of affective states and moves to re-engage the team will be measured by classroom observations and a cross-case analysis of higher and lower functioning teams. Going forward we will be collecting data soon to explore how this conjecture unfolds in the classroom and how we may better support elementary students SEL in collaborative design.

## References

- CASEL'S SEL Framework. (2022). CASEL. Retrieved October 18, 2022, from <https://casel.s3.us-east-2.amazonaws.com/CASEL-SEL-Framework-11.2020.pdf>
- Haverly, C., Lyle, A., Spillane, J. P., Davis, E. A., & Peurach, D. J. (2022). Leading instructional improvement in elementary science: State science coordinators' sense-making about the Next Generation Science Standards. *Journal of Research in Science Teaching*, 1–32.
- Järvenoja, H., Malmberg, J., Törmänen, T., Mänty, K., Haataja, E., Ahola, S., & Järvelä, S. (2020). A Collaborative Learning Design for Promoting and Analyzing Adaptive Motivation and Emotion Regulation in the Science Classroom. *Frontiers in Education*, 5, 111.
- Radloff, J., & Capobianco, B. M. (2019). Investigating elementary teachers' tensions and mitigating strategies related to integrating engineering design-based science instruction. *Research in Science Education*.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences*, 23(1), 18-36.
- Sozo, V., & Ogliari, A. J. (2019). Stimulating design team creativity based on emotional values: A study on idea generation in the early stages of new product development processes. *International Journal of Industrial Ergonomics*.
- Yoon, Jay & Pohlmeier, Anna & Desmet, Pieter & Kim, Chajoong. (2020). Designing for Positive Emotions: Issues and Emerging Research Directions. *The Design Journal*, 24.

# Effect of Providing Knowledge Construction as A Preparatory Activity on Learner's Critical Thinking Achievement in Online Discussion

Lintang Matahari Hasani, Hiroshima University, lintang.matahari97@gmail.com

Kasiyah Junus, Universitas Indonesia, kasiyah@cs.ui.ac.id

Lia Sadita, National Research and Innovation Institute (BRIN), lia.sadita@brin.go.id

Tsukasa Hirashima, Yusuke Hayashi

tsukasa@lel.hiroshima-u.ac.jp, hayashi@lel.hiroshima-u.ac.jp

Hiroshima University

**Abstract:** Fostering critical thinking in online discussion is challenging. Knowledge construction activities have been suggested as having the potential to make learners well-prepared for engaging in a productive discussion. This study investigates the effectiveness of two knowledge construction activities (concept map reconstruction & summary writing) for online discussion preparation. The analysis reveals that students who reconstructed a concept map showed a significantly higher rate of occurrence of critical thinking descriptors with various levels of achievement.

## Introduction

Critical thinking is one important aspect in a successful online discussion that requires learners to co-construct knowledge through sustained transaction of ideas and discovery that lead to a deep understanding. Prior studies found that learners remained at lower-level thinking during online discussions (Bradley et al., 2008; Rourke & Kanuka, 2009). There is also a common tendency of insufficient questioning of the others' ideas or arguments (Garrison & Cleveland-Innes, 2005). To address such issues, various facilitations had been proposed (e.g., peer facilitation). However, past studies showed that some learners relatively remained at lower-level thinking despite being provided with peer facilitation strategies (Chen et al., 2019). A related prior study showed that providing knowledge construction activity in the form of reconstructing a teacher-created concept map (kit-build concept mapping/KBCM) before an online discussion leads to higher cognitive presence compared to writing a summary (Hasani et al., 2023). KBCM is an activity of organizing knowledge by reconstructing a teacher-created concept map from given components (Yamasaki et al., 2010). However, despite the evidence regarding the potential of KBCM in fostering cognitive presence, it is still unknown whether such a treatment could lead to a higher rate of critical thinking level of achievement in an online discussion, e.g., the correctness of the statements, the validity of logical inferences, etc. This study addresses the following research question: To what extent the learners' level of critical thinking achievement was observed in online discussion after being prepared with different knowledge construction activities (KBCM and summary)?

## Methods

This study utilized the data gathered from a related prior study that investigated the effectiveness of the proposed treatments for fostering cognitive presence (Hasani et al., 2023). A quasi-experimental design involving two linear algebra classes (experimental: 48 students; control: 13 students) in an Indonesian university was implemented. The learners participated in the following activities: (a) lecture sessions, (b) preparatory activity (KBCM or summary writing as the treatment) before discussion, (c) discussion training, and (d) online discussion activity. The discussion activity was designed based on the community of inquiry framework (Garrison et al., 1999). A coder who has experience in OCL coded the discussion transcript. A coding scheme developed based on a rubric by Bernstein and Isaac (2018) was used, which includes the following critical thinking descriptors: novel perception, refuting bias, providing alternatives, logical result of supportive evidences, clear expression of strategies and solutions, as well as identifying differences with examples. Three levels of achievement (need improvement, competent, and exemplary) are used to assess learners' ability to perform each descriptor in the discussion. The results were quantified to obtain the aggregate number of all mentioned descriptors occurring in the discussion for each learner.

## Results and discussion

Both groups started from the same level of understanding, which details could be found in the related prior study by Hasani et al. (2023). A total of 237 messages and 19 messages were retrieved from the experimental and control

groups, respectively. The result of Mann-Whitney tests on the aggregate number of the critical thinking descriptor occurrence per learner after  $p$ -value adjustment using Holm's method is shown in Figure 1.

**Figure 1**  
*Number of Code Occurrence/learner Showing Different Level of Critical Thinking*

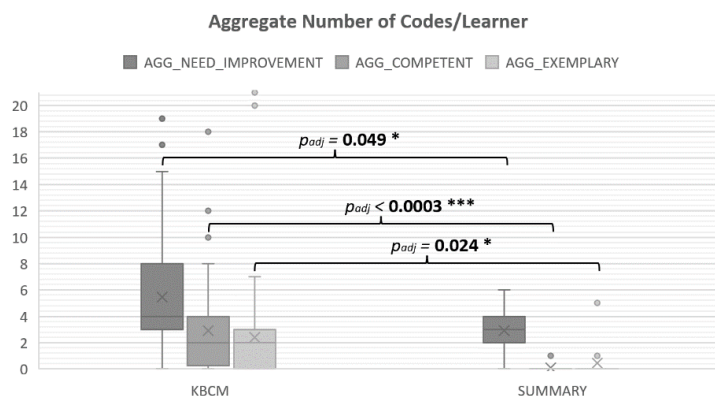


Figure 1 shows a higher number of critical thinking descriptors per person that were categorized as competent and exemplary in KBCM group. This shows that the learners in the experimental group have a higher tendency to provide the correct explanation in a systematic and logically valid way. Learners who used KBCM as a preparatory activity have access to the concept map during discussion tend to discuss more, thus showing a higher rate of messages with varying levels of critical thinking for the mentioned descriptors. Significantly higher number of “need improvement” codes is not a detriment to the learning goal because it enables the instructors and peers to correct mistakes and misconceptions during the discussion. The activity of reconstructing concept maps requires the learners to reflect on their understanding and revisit relevant materials. Such different attitudes could lead to different kinds of discussion, in which the experimental group discussed more productively, e.g., questioning more about the others’ ideas, confirming understanding, etc. Reconstructing concept maps could also increase learners’ awareness of common terms or concepts that are referred to during the discussion. Thus, the learners are enabled to perform more activities listed as descriptors of critical thinking in discussion. To conclude, in comparison to the typical summary writing activity widely used as mathematics writing exercise, KBCM has the potential to be a better method for fostering behaviors that indicate the existence of critical thinking in online discussions. However, this study has the following limitations: one expert-coder inclusion and small sample size due to real-classroom settings experiment. Thus, a follow-up study is needed.

## References

- Bernstein, A. G., & Isaac, C. (2018). Critical Thinking Criteria for Evaluating Online Discussion. *International Journal for the Scholarship of Teaching and Learning*, 12(2). <https://doi.org/10.20429/ijstol.2018.120211>
- Bradley, M. E., Thom, L. R., Hayes, J., & Hay, C. (2008). Ask and you will receive: how question type influences quantity and quality of online discussions. *British Journal of Educational Technology*, 39(5), 888–900.
- Chen, Y., Lei, J., & Cheng, J. (2019). What if online students take on the responsibility: Students’ cognitive presence and peer facilitation techniques. *Online Learning Journal*, 23(1), 37–61.
- Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, 2(2–3), 87–105.
- Hasani, L. M., Junus, K., Sadita, L., Hirashima, T., & Hayashi, Y. (2023). A preliminary study on knowledge reconstruction activity for fostering cognitive presence in online discussion. In J. L. Shih (Ed.), *Proceedings of the 31 st International Conference on Computers in Education* (pp. 475–484). APSCE.
- Rourke, L., & Kanuka, H. (2009). Learning in communities of inquiry: A review of the literature. *International Journal of E-Learning & Distance Education*, 23(1), 19–48.
- Yamasaki, K., Fukuda, H., Hirashima, T., & Funaoi, H. (2010). *Kit-build concept map and its preliminary evaluation*. Asia-Pacific Society for Computers in Education.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 23H01003.

## Supporting Joint Attention for Hybrid Groupwork With Multi-Image and Cursor Sharing in University Classes

Hiroki Oura, Tokyo University of Science, [houra@rs.tus.ac.jp](mailto:houra@rs.tus.ac.jp)  
Ryo Yoshikawa, Aichi Shukutoku University, [yskw@asu.aasa.ac.jp](mailto:yskw@asu.aasa.ac.jp)  
Fumihiko Mitsunaga, Yamato University, [mitsunaga.fumihiko@yamato-u.ac.jp](mailto:mitsunaga.fumihiko@yamato-u.ac.jp)  
Takaki Kondo, Tokyo University of Science, [1722704@ed.tus.ac.jp](mailto:1722704@ed.tus.ac.jp)  
Wakako Fushikida, Tokyo Metropolitan University, [fushikida-wakako@tmu.ac.jp](mailto:fushikida-wakako@tmu.ac.jp)  
Hiroshi Kato, The Open University of Japan, [hkato@ouj.ac.jp](mailto:hkato@ouj.ac.jp)

**Abstract:** The practice of hybrid teaching/learning has become widespread primarily in higher education since the onset of COVID-19. Despite its demands and benefits, there are practical issues for both instructors and students to adjust their teaching/learning strategies as well as their technological support. The present study focuses on joint attention in hybrid groupwork with printed materials and discusses requirements for *hybrid interactional space*, followed by proposing a technology design to support joint attention with multi-image and cursor sharing.

### Introduction

Since the onset of COVID-19, the practice of hybrid teaching and learning with video conferencing services such as Zoom have become widespread primarily in higher education. It is beneficial to ensure learning opportunities for students who struggle to commute for various reasons, and it also is beneficial for universities as they can expect more enrollments. Regardless of these demands and benefits, however, there are practical issues in that both instructors and learners need to adjust their teaching and learning strategies as well as technological support to enable desirable learning process and outcome between classroom and online spaces (Raes et al., 2020).

One of such issues is the difficulty in establishing and maintaining *joint attention* (Tomasello, 1995) in hybrid groupwork. An exemplary situation in the form of hybrid groupwork is when two or more students engage in a group task in a classroom with the other one who joins them from online. In such a situation, it can happen that group conversation tends to progress mainly between the two or more students in the classroom; they can quickly share their resources on their desk and see their nonverbal cues each other such as facial expressions and pointing gestures, whereas the other online student struggles to recognize those cues and keep up with the conversation. CSCL researchers has stressed the significant role of joint attention in collaborative learning (e.g., Barron, 2003) and proposed ways to support joint attention (e.g., Schneider & Pea, 2013), whereas these issues in hybrid teaching and learning environments have not been much discussed yet. Thus, this line of research will be arguably valuable for the CSCL community. Accordingly, the present study discusses requirements to support joint attention in hybrid groupwork and demonstrates one example of technology design to support it.

### Supporting joint attention in hybrid synchronous groupwork

Dillenbourg and Traum (2006) discussed levels of “grounding” as the process of sharing learning resources in which mutual understanding rests among group members. Suthers (2006) pointed out the significant role of “deixis” with deictic referents such as pointing at a referenced object by finger and cursor as well as language (spoken/text words) to mediate intersubjective meaning making. More recently, Schneider and Pea (2013) proposed ways to support visual joint attention by sharing their “gazes” with eye trackers. Together, essential requirements for establishing group joint attention are to enable students to share a) learning resources, b) nonverbal cues (e.g., deictic referents), and c) language among group members synchronously.

When the requirements are applied to hybrid groupwork, practical issues can be identified, which are mainly due to the differences between on-site (i.e., classroom) and online students. For example, students can share their electronic resources such as PDFs using “screen sharing” feature in video conferencing services. However, some students use or prefer printed materials *even if the same electronic materials are available*, since their screen space on laptops, tablets, and/or smartphones is limited. In such situations, students often struggle to establish joint attention on referenced objects. For instance, some students attempt to show printed material by pointing to an object with their finger(s) over their camera; some take a picture of printed material with their smartphone and send it to each other via a social networking service such as LINE; and/or some communicate about a specific part of the material using spoken words, such as by saying, “See the left side on page five in the handout.” These indicate that students attempt to find ways to share their resources and referents, whereas their choices vary and yet unknown is which option more likely leads to better collaborative process and outcome.

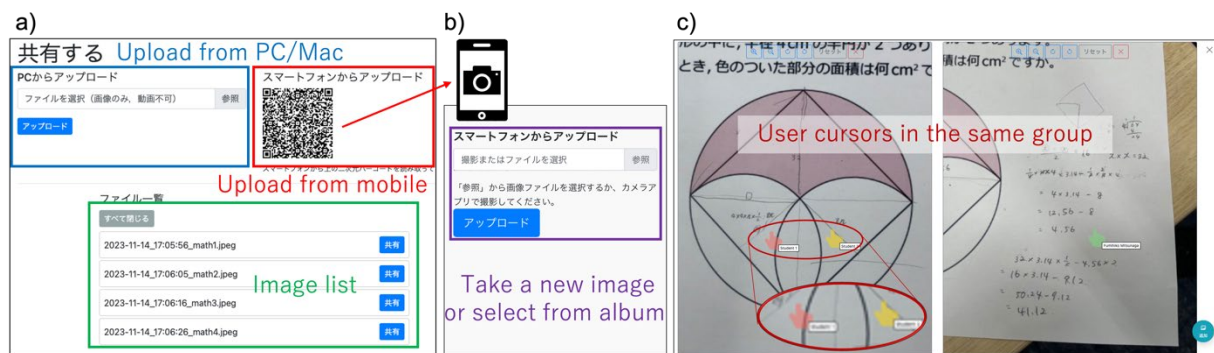
In this regard, we argue that there is theoretical space for research as well as technology design to support students to create *hybrid interactional space* by instantly sharing their multiple forms of learning resources and deictic referents to enhance their joint attention across spaces in hybrid groupwork. Accordingly, we propose an approach to help share printed materials and deictic referents by *multi-image and cursor sharing*.

## Enhancing joint attention through multi-image and cursor sharing

This approach is designed to facilitate the sharing of printed material images captured by a mobile camera, alongside the use of cursors as deictic referents among group members, within our hybrid class system with video conferencing features. Initially, students can share an image by clicking the “Share” button. This action opens a) a user interface for image uploading and displays a list of uploaded images, as illustrated on the left in Figure 1. They have the option to upload images either from a PC or Mac, or directly from a mobile camera by scanning a QR code, b) which opens their mobile web browser to either upload an existing image or capture a new one, as depicted in the middle of Figure 1. The uploaded image appears in the image list and can be shared with group members by selecting the item(s), which displays c) a shared image viewer, as shown on the right in Figure 1. Subsequently, each group member can independently move their cursor within the image area, with their corresponding finger-shaped cursors being synchronized across all screens. Additionally, members can manipulate the image by zooming in or out, rotating, and dragging and dropping, all of which are synchronized processes. They can also add up to four images for juxtaposition, facilitating case comparisons. This technology design is expected to aid students in developing common ground with printed materials complementary to screen sharing through deictic actions and spoken words, thereby enhancing joint attention during hybrid groupwork.

**Figure 1**

System Views of a) Image List and Uploader, b) Mobile Uploader, and c) Shared Image Viewer and User Cursors



## Future work

We are planning to conduct an experiment to examine how the proposed technology support can change students' collaborative learning process and outcome in hybrid groupwork.

## References

- Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, 12(3), 307–359.
- Dillenbourg, P., & Traum, D. (2006). Sharing Solutions: Persistence and Grounding in Multimodal Collaborative Problem Solving. *Journal of the Learning Sciences*, 15(1), 121–151.
- Raes, A., Detienne, L., Windey, I., & Depaepe, F. (2020). A systematic literature review on synchronous hybrid learning: Gaps identified. *Learning Environments Research*, 23(3), 269–290.
- Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375–397.
- Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 315–337.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development*, pp.103–130. Lawrence Erlbaum Associates, Inc.

## Acknowledgments

This work was supported by JSPS Grants-in-Aids for Scientific Research (#JP 21H00897).

# Examining Intrinsic and Extrinsic Game Mechanics in Jigsaw CSCL Scripts

René Lobo-Quintero, Davinia Hernández-Leo  
renealejandro.lobo@upf.edu, davinia.hernandez-leo@upf.edu  
Universitat Pompeu Fabra, Barcelona

**Abstract:** Collaborative Learning Flow Patterns (CLFP) suggest structures for productive Computer-Supported Collaborative Learning (CSCL). It is known that its benefits derive from collaboration scripting mechanics that trigger effective social interactions. Less has been studied from the lens of game mechanics, even if their implementation often suggests playful effects, such as engagement or enjoyment. This study examines the presence of game mechanics in the Jigsaw, as the best-known CLFP which has been implemented in multiple CSCL scripts. Both the CLFP and three cases are analyzed to identify the game mechanics that are intrinsic to the pattern and the extrinsic mechanics that are added in its diverse particularizations as scripts.

## Introduction

Collaborative Learning Flow Patterns (CLFPs) are topic-independent structures of potentially effective scripted sequences of learning activities that can be adapted to multiple educational scenarios (Hernández-Leo et al., 2006). These patterns can help teachers design and incorporate scripted collaborative learning scenarios into their teaching practice. In this paper, we will focus on the Jigsaw CLFP, which is a cooperative/collaborative learning strategy that enhances the process of listening and commitment to the team as well as interdependence and teamwork. In this collaborative pattern, each member of the team has to excel in a well-defined subpart of the educational material, undertaking the role of expert. The experts form a different group to discuss the nuances of the subject and later return to their teams to teach their colleagues.

This paper introduces the intersection of game mechanics with the Jigsaw collaborative learning flow, focusing on how these mechanics can be harnessed to address challenges and enhance the learning experience. Through the lens of the Learning Mechanics-Game Mechanics (LM-GM) framework (Arnab et al., 2014), The LM-GM model maps game mechanics and pedagogical elements abstracted from literature on game studies and learning theories. The model helps to relate a set of standardized learning mechanics to another set of standard game mechanics. It empowers designers to explore the interactions between these mechanics and ensures that a system is solidly rooted in both pedagogical and entertainment perspectives. In this study, we will explore how game design principles are applied inside the Jigsaw CLFP and in collaborative educational settings, studying the game mechanics presented in documented case studies and establishing which game mechanics can be considered intrinsic to the pattern. The main research questions that guide this work are: Which game mechanics are present in the Jigsaw CLFP? Which game mechanics are intrinsic to the Jigsaw CLFP?

## Analysis of the GM in the selected cases

In order to identify the gameful characteristics presented in the Jigsaw CLFP and to distinguish which are the intrinsic game mechanics of the pattern, an analysis of learning-game mechanics used in diverse applications of the Jigsaw CLFP was conducted. Three papers that report case studies that apply Jigsaw were selected considering the following criteria: 1) They report explicit implementations of the Jigsaw pattern; 2) They provide sufficient details of the learning design of the script; 3) The study associated with the script is consolidated (i.e., vs. work in progress); 4) The three papers are written by different authors.

In case 1, the authors implemented a Jigsaw in a face-to-face class using Google Classroom and Google Slides as collaborative tools (Chang & Benson, 2020). During the Jigsaw activity, students cooperate and collaborate, giving their points of view and answering the given question by the instructor (communal discovery GM). Due to the scripted nature of the activity, students didn't have all the information at the beginning (cascading information, GM). They had to help each other in order to build the case (design GM). Each phase of the Jigsaw had a limited time. Students had to take the time limitation into account in order to prepare the presentation of the case. At the post-class step, students received feedback from their classmates and teacher (Assessment GM).

In case 2, the authors present the implementation of a 13-week jigsaw activity (Splichal et al., 2018). Students engaged in 90-minute face-to-face sessions, followed by writing their reflections on the Knowledge Forum tool each week. The task had a high level of complexity. During the 13 weeks, students participated in the jigsaw activity following the instructor's guidance. Students decided their jigsaw group name and assumed one of

the four possible roles (role play and ownership GM), co-constructing an understanding of each stakeholder's perspective. (cooperation, collaboration, communal discovery GM), integrated ideas in the four article sets and worked on proposing solutions that were presented in a poster fair (design, assessment GM), and finally collaboratively wrote a group reflection essay (planning, appointment GM).

In case 3, the authors present an online jigsaw activity using a tool called EthicsApp (Alvarez et al., 2021). Students participated in an ethical discernment activity where they had to adopt an assigned role and prioritize actions that could be taken into account to solve the studied situation (role play, realism GM). In the expert and mixed groups phase, students prioritize the lines of action and provide justification for the prioritization, considering the discussion that just took place (cooperation, collaboration, communal discovery, design/edit GM). In the final step, students presented their prioritization to the rest of the class (ownership GM). The discussion emphasized the importance of considering the stakeholders of interest in the decision-making process, as well as the changes in prioritization found through the three previous activity phases (assessment GM).

The analysis performed shows that the three case studies had several of the game mechanics described in the LM-GM framework. The identified intrinsic game mechanics presented in the three case studies and in the definition of the Jigsaw CLFP are cooperation, collaboration, cascading information, questions and answers, communal discovery, and design/editing. These intrinsic game mechanics correspond with the notion of the jigsaw pattern. Additionally, looking at each one of the cases, the feedback and assessment mechanics depend on the technological tool used and the activities prepared after the collaborative activity is finished. Additionally, the appointment and time pressure depend on the complexity of the script. Finally, depending on the focus of the task, mechanics such as role play, ownership, and planning can be supported.

## Discussion and conclusions

This paper examined the game mechanics embedded within the Jigsaw CLFP through an analysis of three case studies applying the pattern in CSCL. Our study revealed a diverse array of game mechanics that are instrumental in shaping the Jigsaw CLFP's educational efficacy. From the face-to-face interactions supported by cloud platforms in Case 1 to the extended project-based collaboration in Case 2 and the ethical role-playing facilitated by a social platform in Case 3, we observed that game mechanics such as communal discovery, cascading information, questions and answers, and design are recurrent and can be considered intrinsic to the pattern. These mechanics foster a learning environment where cooperation and collaboration are not just encouraged but necessary for the completion of tasks. The constraints imposed by time limitations and the subsequent feedback processes can further augment the learning experience, ensuring that students remain interested and engaged.

## References

- Alvarez, C., et al. (2021). A Social Platform for Fostering Ethical Education through Role-Playing. IntechOpen. DOI: 10.5772/intechopen.96602
- Arnab, S., et al. (2014). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2), 391–411.
- Chang, W.-L., & Benson, V. (2020). Jigsaw teaching method for collaboration on cloud platforms. *Innovations in Education and Teaching International*, 59:1, 24-36.
- Hernández-Leo, D., et al. (2006). COLLAGE: A collaborative Learning Design editor based on patterns. *Educational Technology and Society*, 9(1), 58–71.
- Splichal, J. M., Oshima, J., & Oshima, R. (2018). Regulation of collaboration in project-based learning mediated by CSCL scripting reflection. *Computers & Education*, 125, 132-145.

## Acknowledgments

This work has been partially funded by MICIUAEI/10.13039/501100011033 (PID2020-112584RB-C33, CEX2021-001195-M) and the Volkswagen Stiftung (Courage, ref. 95 566). DHL (Serra Hünter) also acknowledges the support by ICREA under the ICREA Academia programme and the Department of Research and Universities of the Government of Catalonia (SGR00930).

# Designing for Large-Scale Social Annotation Experience: Real-Time Interactions on the Margins of Educational Material

Jumana Almahmoud, MIT, jumanam@mit.edu  
 Michele Igo, UC Davis, mmigo@ucdavis.edu  
 Marc Facciotti, UC Davis, mtfacciotti@ucdavis.edu  
 Kamali Sripathi, UC Davis, ksripathi@ucdavis.edu  
 David Karger, MIT, karger@mit.edu

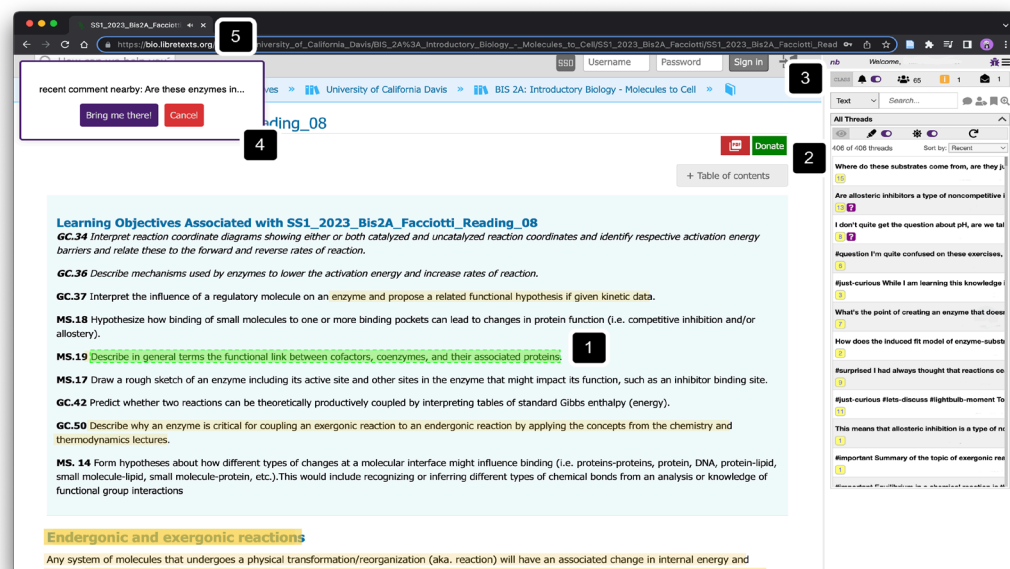
**Abstract:** This study explores the use of synchronous social annotation (SA) tools to enhance real-time discussions in educational texts. Visual cues were introduced to facilitate contextual synchronous engagement in a large undergraduate class of 1009 students. Results revealed significant improvements in peer-to-peer interactions, discussion quality, and social presence, alongside quicker peer response times.

## Introduction and motivation

This study examines the integration of synchronous social annotation (SA) tools in educational settings, leveraging both SA and synchronous chat platforms' benefits for enhancing discussion and learning. SA tools traditionally facilitate asynchronous, focused discussions on specific text parts (Gao, 2013; Kalir, 2020), in contrast with synchronous platforms that promote real-time interaction (Ferschke et al., 2015). Bridging these modes, grounded in social constructivist theories (Bruner, 1997), aims to optimize engagement and learning, especially in large classes where the volume of discussion presents both challenges and opportunities (Chen et al., 2021). Our initiative explores a synchronous SA platform's design, enhancing real-time margin discussions in large-scale settings, addressing the tension between document and temporal comment locality. We enhanced the open-source SA tool NB (Haystack, 2019/2021) for a 1009-student Introductory Biology course, observing improved peer interactions, discussion quality, and social presence. Preliminary interviews highlighted the asynchronous limitations and the informal yet contextually disjointed nature of platforms like Discord. This informed our design to include synchronous features, focusing on enhancing student engagement and comprehension by facilitating quicker, contextually anchored peer responses (Figure 1). Our research questions delve into designing quality synchronous SA experiences, their impact on discussion quality and learning, and their ability to promptly assist confused students.

**Figure 1**

1- Spotlight Real-Time Annotations. 2- Default to Recent. 3- Options for the Sync Experience. 4- Pop-Up Notifications. 5- Sound Notifications.



Note: 3-option to turn off real-time interactions, # of users online: students & instructors



## Methods

Using a mixed-methods approach, we evaluated the effect of synchronous features in social annotation tools. In this study, we compared two classes using the SA tool NB. One experimental group class consisted of 1,000 students with access to synchronous features. In contrast, the other control group class consisted of 216 students who did not have access to these features. We analyzed both groups for engagement and interaction patterns using predefined terms. We conducted in-depth interviews with 13 participants to gather their experiences and perceptions of NB's synchronous features. The interviews aimed to uncover participants' reading habits, experiences with synchronous interactions, and sense of social presence within the platform. We also used quantitative log data to complement the qualitative data and measure engagement and interaction patterns. To assess the impact of synchronous features on discussion quality, we randomly selected 24 threads and conducted a study. Our assessment focused on communication clarity, correctness, resolution, and comment build-up. We evaluated the impact of synchronous features on user engagement, discussion quality, and the overall educational experience on the SA platform using qualitative insights from interviews and quantitative analysis of interaction logs and discussion quality.

## Results and discussion

Our study showed that synchronous features in a social annotation platform improved student engagement and discussion quality. Real-time notifications and prioritizing recent annotations reduced the response time to 1.47 hours from 19.76 hours in the control group, indicating the effectiveness of synchronous features in facilitating prompt peer responses. Synchronous features also enhanced interaction immediacy, with 42.5% of annotations receiving replies within 60 minutes compared to only 5.82% in the control group. Interviews revealed that students valued real-time interactions, with 51% engaging with spotlight comments and 37% with sidebar comments, suggesting a preference for discussions closely related to their current reading passage. This proximity encouraged richer student interactions, fostering a stronger sense of community and collaboration. Moreover, students reported an increased sense of community, appreciating the ability to interact synchronously without switching between multiple platforms, contrasting with the more informal and disjointed nature of platforms like Discord.

The study also highlighted a significant improvement in the quality of discussions, with synchronous threads receiving higher ratings from instructors for quality, resolution, and build-up compared to asynchronous ones. Despite a slightly higher occurrence of incorrect statements in synchronous discussions, these were often resolved, leading to more constructive and layered student exchanges. The presence of instructors online was beneficial, encouraging immediate feedback and creating a supportive learning environment. However, the data also indicated an inverse relationship between instructor presence and student engagement, with more student replies occurring in the absence of instructors. This suggests that while instructor guidance is valuable, autonomous peer-to-peer interactions also play a critical role in the learning process, aligning with social constructivist theories. In conclusion, integrating synchronous features in SA platforms significantly enhances the learning experience, fostering quicker peer interactions, improving discussion quality, and strengthening the community among students. While the presence of instructors is instrumental in guiding learning, enabling students to interact autonomously is equally essential for fostering engagement and deepening understanding.

## References

- Bruner, J. (1997). *The Culture of Education*: Harvard University Press.
- Chen, B., Håklev, S., & Rosé, C. P. (2021). Collaborative Learning at Scale. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 163–181). Springer International Publishing. [https://doi.org/10.1007/978-3-030-65291-3\\_9](https://doi.org/10.1007/978-3-030-65291-3_9)
- Ferschke, O., Yang, D., Tomar, G., & Rosé, C. P. (2015). Positive Impact of Collaborative Chat Participation in an edX MOOC. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (pp. 115–124). Springer International Publishing. [https://doi.org/10.1007/978-3-319-19773-9\\_12](https://doi.org/10.1007/978-3-319-19773-9_12)
- Gao, F. (2013). A case study of using a social annotation tool to support collaboratively learning. *The Internet and Higher Education*, 17, 76.
- Haystack. (2021). NB [JavaScript]. The Haystack Research Group. <https://github.com/haystack/nb> (Original work published 2019)
- Kalir, J. H. (2020). Social annotation enabling collaboration for open learning. *Distance Education*, 41(2), 245–260. <https://doi.org/10.1080/01587919.2020.1757413>

# Linking Disagreement Reasons and Integrative Conclusions: An Epistemic Script to Promote Integration of Multiple Conflicting Information Based on Disagreement Analysis

Toshio Mochizuki, Senshu University, tmochi@mochi-lab.net  
Eowyn P. O'Dwyer, Rutgers University, eowyn.odwyer@rutgers.edu  
Clark A. Chinn, Rutgers University, clark.chinn@gse.rutgers.edu  
Myat Min Swe, Htay Min Khaung, Seiji Sekine  
myatmins@spiceworks.co.jp, htay.kaung@spiceworks.co.jp, sekine@spiceworks.co.jp  
Spiceworks Myanmar

**Abstract:** This study addresses the challenge that lay students often consider few possible reasons for disagreement when formulating integrative conclusions using multiple conflicting documents. We introduced an epistemic script, encouraging students to link disagreement reasons and conclusions to improve their reasoning about disagreements. A comparative experiment revealed that the experimental groups with the script showed a stronger tendency to integrate their identified disagreement reasons into their conclusions more effectively.

## Introduction

In the contemporary world, citizens need to integrate multiple pieces of conflicting information during daily problem solving (Kienhues et al., 2016). However, laypeople frequently struggle to interpret and evaluate diverse viewpoints and integrate conflicting information from multiple sources (Bråten et al., 2011). One source of this difficulty is confirmation bias or motivated reasoning, which is the tendency for individuals to engage with and more favorably evaluate evidence that aligns with their preferences and prior beliefs (Barzilai et al., 2020). Successful integration of multiple conflicting information requires identifying disagreements (Thomm et al., 2017) and evaluating, understanding, and resolving them. Nonetheless, laypeople tend to provide few reasons when formulating possible explanations for disagreements among conflicting sources of information (Thomm et al., 2017).

Previous research revealed that CSCL-based epistemic scaffolds (Sandoval & Reiser, 2004) and epistemic scripts (Weinberger et al., 2005) effectively facilitated productive epistemic discourse for scrutinizing evidence and identifying disagreements among the documents (Mochizuki et al., 2023). However, the study also revealed that the students often analyzed disagreements productively but then failed to use this analysis when reaching conclusions. This study addresses this challenge by providing an additional epistemic script which encourages students to link their analyses of disagreements to their conclusions, thereby generating more integrative conclusions that considers more of the available information.

## Linking disagreement reasons and conclusions for comprehensive reasoning

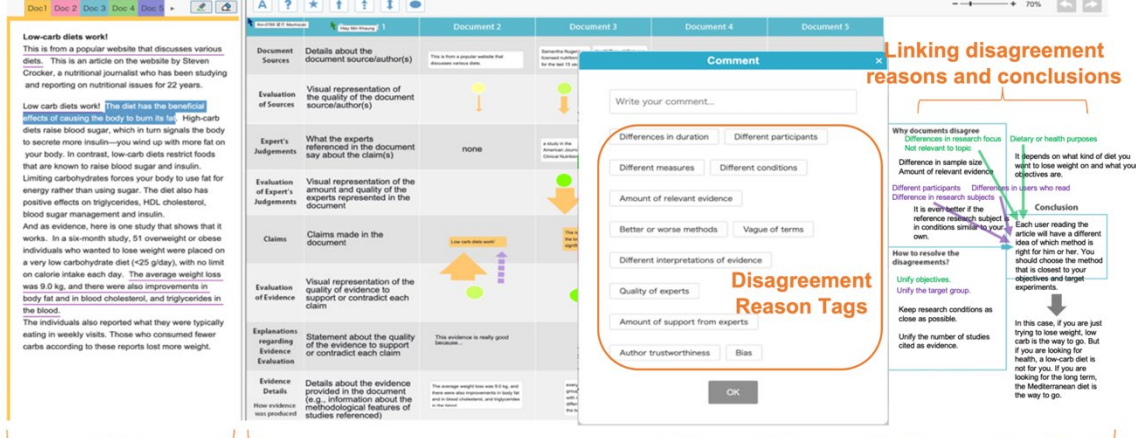
The Web-based CSCL environment is designed to engage learners in systematically analyzing sets of conflicting documents. On a graphic organizer called the “analysis tableau,” learners collectively synthesize information, identify disagreements among the documents, and scrutinize the underlying reasons for these disagreements, with the ultimate aim of formulating reasonable, evidence-based conclusions. The tableau is grounded in the Grasp of Evidence Framework (Duncan et al., 2018) and incorporates epistemic scripts and epistemic scaffolds.

This study aims to explore the impact of an additional epistemic script to prompt students to connect their identified disagreement reasons with their conclusion ideas during the process of formulating conclusions (Figure 1). This approach leverages the power of linking elements within a graphic organizer, facilitating awareness and fostering reasoning about the relationships among these elements. Moreover, it encourages students to become aware of any information to which they had not well attended.

## Method

Sixteen triads participated in the Zoom-based study. Each triad was provided with five different credible documents within the system; these discussed various dietary approaches, their potential benefits and drawbacks, and referenced relevant scientific studies. Each triad was tasked with collaboratively reaching reasonable conclusions on the best dietary approaches based on their analysis of the documents within the tableau. Each triad was randomly assigned to either the experimental or control group. In the experimental condition, triads received explicit prompts to link reasons for disagreements with their emerging conclusions, whereas the control group did not receive such guidance. All group conversations were recorded and transcribed. The research included an

**Figure 1**  
*Implemented Epistemic Scripts and Scaffolds including Linking Disagreement Reasons and Conclusions*



**Multiple Documents**

**Analysis Tableau**

**Disagreement Reason Tags**

**Linking disagreement reasons and conclusions**

\*Orange annotations are explanations for the interface

analysis of student utterances related to students' reasons for disagreements and the formulation of conclusions based on these reasons, including how students referenced these reasons while articulating their conclusions.

## Results and discussion

The experimental triads used an average of 4.2 disagreement reasons ( $SD = 1.93$ ) when developing conclusions out of an average of possible 6.0 reasons ( $SD = 2.21$ ) that the groups had generated during their analyses. In contrast, the control triads used an average of 2.0 disagreement reasons ( $SD = 1.10$ ) from an average of potential 5.0 reasons ( $SD = 2.00$ ) that were initially identified before their formulation of their conclusions. Thus, the experimental groups, on average, integrated the reasons for disagreement that they identified into their conclusions at a rate of 70.3% ( $SD = 20.9$ ), whereas the control groups demonstrated the lower rate of 51.0% ( $SD = 39.2$ ). The experimental groups tended to provide more comprehensive explanations to underpin their conclusions. Their analyses drew from diverse facets associated with different disagreement reasons. For example, in some instances, the experimental groups elucidated more than two distinct aspects of disagreement reasons, encompassing factors like varying durations (e.g., low-carb diets are better in the short term, whereas the Mediterranean diet is better for the long term) and divergent focal points of studies (e.g., different populations studied) cited in the documents. In contrast, the control groups typically furnished more simplistic conclusions, primarily referencing a single dimension of the disagreements, such as only duration. Further analysis will reveal how using this epistemic script enriched students' discourse in a more comprehensive way.

## References

- Barzilai, S., Thomm, E., & Shlomi-Elouz, T. (2020). Dealing with disagreement: The roles of topic familiarity and disagreement explanation in reasoning about conflicting expert claims and sources. *Learning and Instruction*, 69, 101367.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, 46 (1), 48–70.
- Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching*, 55 (7), 907–937.
- Kienhues, D., Ferguson, L. E., & Stahl, E. (2016). Diverging information and epistemic change. In J. A. Greene, W. A., Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 318–330). Routledge.
- Mochizuki, T., O'Dwyer, P. E., Chinn, C. A., Myat Min Swe, Htay Min Khaung, & Sekine, S. (2023). Epistemic scaffolds to promote disagreement identification and resolution for multiple conflicting documents. In C. Damşa, M. Borge, E. Koh, & M. Worsley (Eds.), *15th International Conference on Computer-Supported Collaborative Learning (CSCL)*, International Society of the Learning Sciences, pp. 313–317
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3), 345–372.
- Thomm, E., Barzilai, S., & Bromme, R. (2017). Why do experts disagree? The role of conflict topics and epistemic perspectives in conflict explanations. *Learning and Instruction*, 52, 15–26.
- Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33, 1–30.

## Tracing the Synchronized Physiological Arousal during GA, Supported Collaborative Learning Process

Xiaoran Li, Beijing Advanced Innovation Center for Language Resources, Beijing Language and Culture University, ella.li68@yahoo.com

**Abstract:** Due to the complexity and multi-dimension of collaborative learning, researchers started to investigate the process with multimodal data. The current study aims to investigate the synchronized physiological arousal (SPA) in group awareness (GA), supported collaborative learning by using heart rate and interview data. The findings indicated that the overall proportion of SPA was higher in the GA-supported groups, and this difference was more pronounced in the middle and late phases of collaborative writing sessions.

### Background

Collaborative learning is a multi-dimensional process. Traditionally, researchers collected logs, audio recordings data to investigate the process. In recent years, the physiological measure has been gradually taken into consideration. Among them, the synchronized physiological arousal (SPA) is an important indicator, which refers to two or more group members are in a state of physiological arousal, which is a desirable goal in learning as it accounts for learners' cognitive and/or affective activation (Palumbo et al., 2017). Therefore, the SPA is a particularly meaningful situation and is informative of collaboration (Malmberg et al., 2019). Although a few studies have explored the SPA during collaborative learning (Nguyen et al., 2022), few studies have explored the characteristics of SPA in collaborative learning supported by technical tools. Therefore, this study attempts to fill in the gaps by exploring the SPA characteristics in GA-supported collaborative learning. Through a self-developed automatic analysis platform, this study identifies the SPA by heart rate data. Combining with interview data, this study also tried to provide further explanations for these SPA characteristics. Two research questions guide this research: RQ1: How the SPA characteristics in GA-supported collaborative learning differ from the non-GA-supported collaborative learning? RQ2: What further explanation can the interview data provide for these characteristics?

### Method

#### Participants and Collaborative writing activity

The research was conducted in the English Academic Reading and Writing course in a normal university in Beijing, which was selected by 32 first-year postgraduate students (6 male students). All students volunteered to participate in the study and were formed into 10 groups. A survey on students' English academic reading and writing experience was firstly distributed and the results showed that there is no polarization in student' English academic reading and writing level. The 10 groups were randomly assigned to 5 GA-supported groups and 5 non-GA-supported groups. The English Academic Reading and Writing course lasted one semester. The task of collaborative writing was to complete a 4000-6000-word English essay on a self-determined topic during the semester. Each group was given about 8 weeks to select a topic, search for literature, and write the essay. The average collaborative writing time for each group was about one hour per week.

#### C-SRS Group awareness tool

For the GA-supported groups, this study provided the C-SRS group awareness tool (Challenge-social regulatory strategies group awareness tool). The purpose of the C-SRS group awareness tool is to improve groups' awareness of diverse challenges and the ability to regulate identified challenges by leading students to reflect on their collaboration and providing explicit instructions. The GA-supported group utilized the C-SRS tools once an hour to reflect and regulate their collaborative writing.

#### Data collection and data analysis

*Heart rate data:* To automatically identify the SPA, this study developed a heart rate analysis platform based on the previous psychophysiological research (Mason et al., 2018), which indicated increases in HR reactivity may reflect arousal and have been used as popular measures for detecting psychological alertness in facing stimulus (Critchley et al., 2013). Two steps are included to identify SPA. For the first step, this study utilized the moving window approach, with a window width of 1 min and a moving step of 1 s, to calculate the heart rate data during

the collaborative writing session to find the segment with the smallest standard deviation and longest duration as the baseline heart rate. For the second step, the heart rate analysis platform divided the heart rate data into 30-second episodes. When an individual's average heart rate within the 30-second episode was higher than the baseline heart rate, a physiological arousal episode is identified. SPA is recognized when two or more group members appear in a physiological arousal state

*Interview data:* At the end of the semester, all students were interviewed with questions such as "Were you satisfied with your group's collaborative writing?" "Have you encountered any challenges? How did you team cope with the challenges." "Whether the challenges your group encounter affect the quality of your collaboration?" "Do you think the use of learning tools can help your collaborative writing and how? (GA-supported groups only)". All interview data were transcribed, and thematic analysis was used to analyze interview data in an attempt to find themes that could explain the SPA characteristics.

## Results

A total of 22 collaborative writing sessions with 137 hours and 56 minutes (12 hours and 20 minutes with 1480 episodes from GA-supported groups) were collected.

*For RQ1:* In GA-supported groups, nearly half of the episodes were SPA (640/1480). The overall proportion of SPA in GA-supported sessions (45.27%) was significantly higher than that of non-GA-supported sessions (23.54%). Each session was equally divided into three phases. In phase 1 (the first third of collaborative writing session), the proportion of SPA in GA-supported groups was slightly higher than that of non-GA-supported groups (55.2% vs 42.8%). While in the last two phases (the last two thirds of collaborative writing session), the proportion of SPA in GA-supported groups was much higher than that of non-GA-supported groups (46.9% vs 20.4%; 33.9% vs 8.1%). In sum, the proportion of SPA was higher in the GA support group than in the non-GA group, and this difference was more pronounced in the middle and late phases of collaborative writing sessions.

*For RQ2:* One main theme identified in non-GA-supported groups is "Collaborative learning gradually becoming ineffective due to some unsolvable challenges". On the contrary, the theme identified in GA-supported groups is "the C-SRS tool providing explicit instructions for students to cope with the challenges and therefore maintaining higher engagement in collaboration". Students indicated that C-SRS tools helped them reflect on their collaboration and provided diverse social regulation strategies to inspire them dealing with the challenges. As the challenges were gradually solved, the engagement in collaborative writing session is consistently high.

## Conclusion

This exploratory study tried to investigate the SPA characteristics in GA-supported collaborative learning process not only to extend the understanding of the collaborative learning from the physiological perspective but also to unpack how the GA tool may affect students' physiological characteristics. The preliminary findings showed higher SPA in GA-supported groups because C-SRS may help them to reflect on and deal with the challenges during the task so as to maintain the constant high engagement. This study can inspire researchers conducting multimodal data study on collaborative learning process. Additionally, the self-developed heart rate analysis platform could be further used in future studies.

## References

- Critchley, H. D., Eccles, J., & Garfinkel, S. N. (2013). Interaction between cognition, emotion, and the autonomic nervous system. In R. M. Buijs, & D. F. Swaab (Eds.), *Handbook of clinical neurology* (Vol. 117) (pp. 59–77). Elsevier.
- Mason, L., Scrimin, S., Zaccoletti, S., Tornatora, C. M., & Goetz, T. (2018). Webpage reading: psychophysiological correlates of emotional arousal and regulation predict multiple-text comprehension. *Computers in Human Behavior*, *87*, 317–326.
- Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X. H., & Siipo, A. (2019). Going beyond what is visible: what multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior*, *96*, 235–245.
- Nguyen, A., Järvelä, S., Rosé, C., Järvenoja, H., & Malmberg, J. (2022). Examining socially shared regulation and shared physiological arousal events with multimodal learning analytics. *British Journal of Educational Technology*, *54*, 293–312.
- Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., et al. (2017). Interpersonal autonomic physiology: a systematic review of the literature. *Pers. Soc. Psychol. Rev.* *21*, 99–141.

## Are Your Top Social Annotation Friends Also Your Offline Group Members?

Gaoxia Zhu, Nanyang Technological University, gaoxia.zhu@nie.edu.sg  
Tianlong Zhong, Nanyang Technological University, TIANLONG001@e.ntu.edu.sg  
Hanxiang Du, Western Washington University, duh2@wwu.edu  
Juan Zheng, Lehigh University, juz322@lehigh.edu  
Xiaoshan Huang, McGill University, xiaoshan.huang@mail.mcgill.ca  
Shan Li, Lehigh University, shla22@lehigh.edu  
Chenyu Hou, Nanyang Technological University, chenyu004@e.ntu.edu.sg

**Abstract:** Previous research has highlighted homophily (i.e., the tendency to interact with similar others) in online social networks, but its presence in social annotation activities, particularly influenced by group affiliations, remains underexplored. The current research addresses this gap by analyzing the online interactions of 93 undergraduates in a learning media course on Perusall. Findings reveal students' preference for interactions within the same offline groups during social annotations, contributing to a deeper understanding of homophily in hybrid settings. These insights are vital for enhancing student engagement in hybrid learning environments and promoting more effective interactions.

### Objectives and significance

Homophily is the tendency for individuals to connect with others who share similar traits, such as gender, race, and educational backgrounds (Aiello et al., 2012). Research has shown that homophily exists not just in physical interactions but also in online environments (Aiello et al., 2012). However, how offline characteristics affect online networks remains a topic for ongoing investigation. In higher education, students increasingly engage in social annotation by highlighting and commenting on course materials, responding to peers, and negotiating meanings and building knowledge (Zhu et al., 2023). Tools like Perusall facilitate this process by offering flexibility and supporting hybrid learning, free from time and space constraints. Previous studies indicate that social annotation enhances higher-order skills such as collaboration and critical thinking (Chen, 2019), boosts engagement and motivation, and provides additional benefits like automated grading (Li & Li, 2023). As social annotation often involves assigning readings as pre-class preparation and consolidating with in-depth in-class discussions, understanding the influences of offline characteristics on students' online interactions is crucial. This study will probe into offline group affiliation-based homophily in social annotations to provide educators with insights to foster more effective student interactions.

### Methods

#### Participants and learning environment

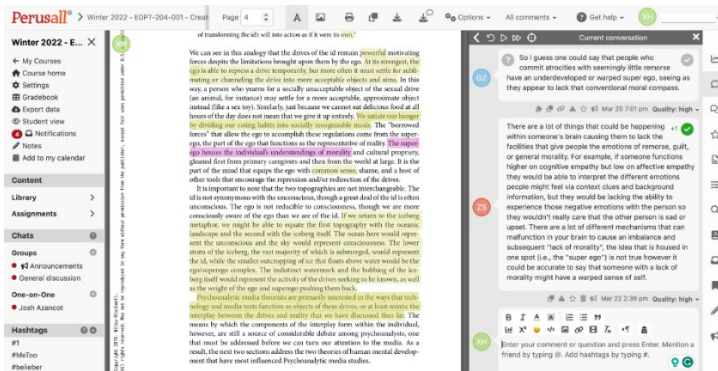
The research included 97 undergraduates (47 female,  $M_{age}=20.67$ ) enrolled in a learning media course at a North American university during the winter semester (weeks 17 to 13) of 2022. The participant demographic was primarily White (59 participants), with 5 Chinese, 3 South Asian, 2 Southeast Asian, 1 Black, 1 Filipino, 1 Japanese, 1 Korean, and 1 Latin American student. Additionally, 3 participants identified themselves as other racial backgrounds, and the others did not disclose their race. The students were organized into 16 groups, typically with six members each. They were tasked with annotating ten readings for seven weeks using Perusall, a digital social annotation tool. Four students withdrew from the course, resulting in a final sample size 93.

Perusall serves as an interactive platform for collectively annotating and discussing designated readings. It allows students to participate in the annotation process by adding comments or posing questions about the readings, as illustrated in Figure 1. In this study, students were required to actively annotate ten readings by creating their own annotations and engaging with others' annotations through comments.

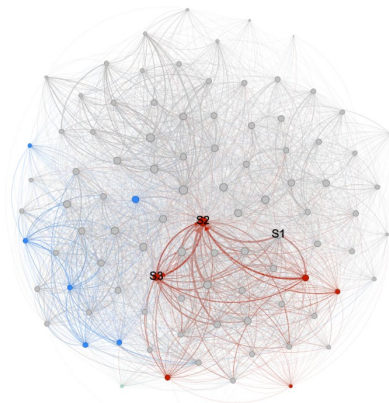
#### Data collection and analysis

Participants' asynchronous social annotations of assigned readings were collected from Perusall's log files. Over seven weeks, participants made a total of 7,482 social annotations. For example, as Figure 1 displays, GZ and ZS subsequently annotated the passage "The super-ego houses the individual's understandings of morality". This sequence suggests that ZS read GZ's annotation before commenting. We constructed a social network graph based on such connections (see Figure 2). NetworkX, a Python library, was employed to plot a social network graph representing the interactions within the 7,482 social annotations.

**Figure 1**  
Example of Annotations in the Persuall platform



**Figure 2**  
The Social Annotation Network Graph



We identified each student's social best friend –the neighbor with the highest degree in the network graph. Since most groups had six students, we also identified each student's top five social annotation friends and examined the co-membership with the student's offline group affiliations. Identifying social best friends is a prevalent method to analyze student interaction patterns in online and offline learning contexts. This study examined whether students tended to interact with their offline group members. As a baseline measure, we computed the likelihood of a student's best friend or top five friends in the same offline group by chance. The probability that two randomly selected students belong to the same group can be calculated as follows: Each of groups  $G_1, \dots, G_{14}$  has 6 students, Group  $G_{15}$  has 5, and Group  $G_{16}$  has 4. Student name is  $S$ , and their best friend is  $F$ .  $P[S \text{ and } F \text{ are in the same group}] = (\sum_{i=1}^{14} P[S \text{ and } F \text{ are in group } G_i]) + P[S \text{ and } F \text{ are in group } G_{15}] + P[S \text{ and } F \text{ are in group } G_{16}] = 14*6/96*5/92+5/93*4/92+4/93*3/92=5.28\%$ . Similarly, the chance that at least one of a student's top five friends is from the same group is calculated, which is about 24.22%.

## Results and discussions

This study showed that 20.43% of students had their best social annotation friend from their offline group, and 67.74% had at least one offline groupmate among their top five friends. These probabilities suggest a greater likelihood of students interacting with their offline group peers in online social annotations. These results indicate the homophily phenomenon in social annotations based on offline group affiliations. Our findings indicate a trend where students prefer engaging with peers from their offline groups during online social annotation, reflecting observations that offline socialization boosts online community engagement and a sense of belonging (Ke et al., 2011). This is likely due to the comfort and familiarity developed through face-to-face interactions, which lowers social barriers and promotes participation. Such interactions, often grounded in shared academic aims and experiences, as Nussbaum et al. (2009) noted, facilitate the extension of classroom discussions and collaboration towards common objectives in online settings. This emphasizes the influence of offline group dynamics on online collaboration, enriching our comprehension of hybrid learning environments.

## References

- Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, 6(2), 1-33.
- Chen, B. (2019). Designing for networked collaborative discourse: An UnLMS approach. *TechTrends*, 63(2), 194-201.
- Ke, F., Chávez, A. F., Causarano, P.-N. L., & Causarano, A. (2011). Identity presence and knowledge building: Joint emergence in online learning environments? *International Journal of Computer-Supported Collaborative Learning*, 6, 349-370.
- Li, M., & Li, J. (2023). Using Persuall to motivate students' curriculum-based academic reading. *Journal of Computers in Education*, 10(2), 377-401.
- Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face collaborative scaffolding. *Computers & Education*, 52(1), 147-153.
- Zhu, X., Shui, H., & Chen, B. (2023). Beyond reading together: Facilitating knowledge construction through participation roles and social annotation in college classrooms. *The Internet and Higher Education*, 59, 100919.

# The Multimodal Analysis of Learners' Mathematical Proof Activity in a CSCL Environment

Masataka Kaneko, Toho University, [masataka.kaneko@phar.toho-u.ac.jp](mailto:masataka.kaneko@phar.toho-u.ac.jp)  
 Hironori Egi, The University of Electro-Communications, [hiro.egi@uec.ac.jp](mailto:hiro.egi@uec.ac.jp)  
 Takuya Kitamoto, Yamaguchi University, [kitamoto@yamaguchi-u.ac.jp](mailto:kitamoto@yamaguchi-u.ac.jp)  
 Takeo Noda, Toho University, [noda@c.sci.toho-u.ac.jp](mailto:noda@c.sci.toho-u.ac.jp)

**Abstract:** This study investigates learners' mathematical proof activity in a CSCL environment. While the interrelation between their utterances and gestures should be elucidated to fully understand the progress of their thinking, automatic detection of gestures is indispensable to make the research scalable. As a preliminary step, this study used the depth data derived from sensors to detect learners' writing behavior. The result indicated that the same framework might apply to gestures with appropriate arrangement of the relevant tools.

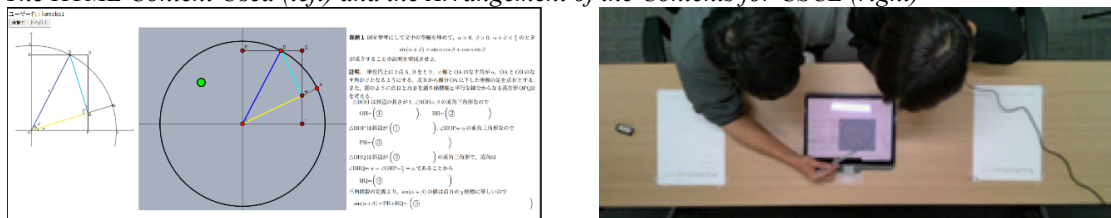
## Introduction and background

This study investigates mathematical proof activity in a CSCL environment. The topic is the generalization of the proof for the addition theorem for trigonometric functions. It can be assumed that learners experience some commognitive conflict while generalizing their prior knowledge to newly emerging situations while they are engaged in discursive activities to handle those conflicts (Sfard, 2008). Moreover, the theory of mathematical cognition indicates that those activities are accompanied by gestures as a conceptual metaphor (Lakoff & Nunez, 2001). In line with these theoretical perspectives, some previous learning analytics studies have demonstrated that incorporating the multimodal features related to gesture and posture can improve our understanding of learners' dialogue acts (Ezen-Can, et al., 2015). Regarding learners specifying the similarities and the differences in each proof component between the rudimentary case and the general cases, the progress of learners' thinking is assumed to be manifested in the temporal transition of their behaviors. Our prior study indicated that the gestures observed in the proof activity were mainly composed of pointing gestures (Alibali & Nathan, 2012) indexing some geometric objects drawn on a PC screen or some mathematical expressions in the proof script. Moreover, the qualitative study of learners' verbal and nonverbal communications indicated that the modality change of their behavior can be accelerated through advances in their thinking. Although these findings can lead to a better explication of learners' proof activity, it is not easy to make this research scalable without detecting their gestures automatically. This paper shows our trial of using sensing technology to automatically detect learners' hand movements and classify their behaviors.

## Method and results

Participants were first-year students in a Japanese university majoring in physics. They have learned the proof for the theorem in the rudimentary case where the relevant angles are acute. They were asked to validate that the theorem still holds for more general cases where the angles may be obtuse. They used the HTML content (the left of Figure 1) including both the script of the proof in the rudimentary case and the simulation screen generated by the dynamic geometry software CindyJS (<https://cindyjs.org>). An iPad in which this HTML content is implemented and the recording paper were arranged on a table as shown to the right of Figure 1.

**Figure 1**  
*The HTML Content Used (left) and the Arrangement of the Contents for CSCL (right)*



The participants' behavior was videotaped and the recorded image was imported into the behavioral analysis system Sportscode (<https://www.hudl.com/products/sportscode>). In addition, the Intel RealSense depth camera (<https://www.intelrealsense.com/stereo-depth>) was used to measure the depth from the sensor to the specific regions including two papers and an iPad and depth data were logged with time stamps. The depth camera was

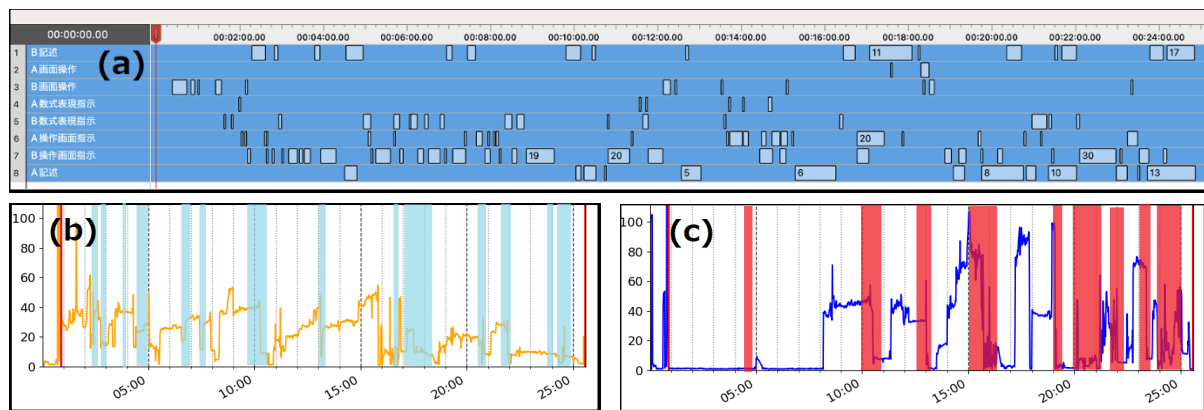


placed over the learners' heads so that it would not obstruct their concentration. The videotaped behaviors of each participant were examined qualitatively and classified into the categories "description on the paper", "pointing to geometric objects", "pointing to mathematical expression", and "manipulation of the simulation screen". The result of this classification was coded on a single timeline in Sportscode. The difference in depth from the initial value was calculated and visualized on another timeline.

Figure 2 (a) shows the Sportscode timeline in which the "description on the paper" behavior of each participant was coded on the top and bottom layers respectively. Figure 2 (b) and (c) show the fluctuations of the depth from the sensor to the two paper regions. Since movements over these paper regions can cause changes in the depth data, the time intervals coded on the top and bottom layers in (a) were overlaid onto the timelines (b) and (c) with light blue and red frames respectively.

**Figure 2**

*The Sportscode Timeline (a) and the Fluctuations of the Depth from the Sensor to Paper Regions (b and c)*



While large values in (b) and (c) indicate participants' head movements over the paper region, little non-zero values indicate their hand movements. It can be seen that many overlaying regions include the latter cases. This result shows that, by arranging the position of papers farther from the participants, the depth data can be used to automatically detect their "description on the paper" behavior. Qualitative analysis of the participants' utterances shows that they became aware of the interrelation between the positional change of geometric objects and the corresponding change in the signature of mathematical expressions at about 14:30 after beginning their discussion of the problem. Figure 2 shows that the modality change in participants' behavior was triggered by this progression of their thinking.

## Conclusion and future works

While participants' indexing of geometric objects by pointing to them occurs in the central part of the iPad region, the indexing of mathematical expressions occurs on the right side of the screen. It was observed that some of the participants' hand movements occurred on both regions simultaneously. The result of this study indicates that, by arranging the positions of these contents appropriately, the automatic detection of learners' pointing will become possible using the depth data derived from the sensor. More comparative studies are needed across a wider variety of learners and topics to validate our assumption about the relationship between the modality change of learners' behavior and the progress of their thinking.

## References

- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *The Journal of Learning Sciences*, 21, 247–286.
- Ezen-Can, A., Grafsgaard, J. F., Lester J. C., & Boyer, K. E. (2015). Classifying student dialogue acts with multimodal learning analytics. *Proceedings of the 5<sup>th</sup> International Conference on Learning Analytics and Knowledge*. 280–289.
- Lakoff, G., & Nunez, R. (2001). *Where mathematics comes from*. Basic Books.
- Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press.

## Acknowledgments

This work was supported by JSPS KAKENHI (21K02752, 22K02951).

## Epistemic Agency for STEM Learning with Design-Based Scaffolds

Jialiang LIU, The University of Hong Kong, ljliang@connect.hku.hk

Gaowei Chen, The University of Hong Kong, gwchen@hku.hk

Carol K. K. Chan, The University of Hong Kong, ckkchan@hku.hk

Jianhua Zhao, Southern University of Science and Technology, zhaojh@sustech.edu.cn

**Abstract:** This study was to develop a design-based scaffold that enhance students' epistemic agency by utilizing documentation and reflection to foster their STEM inquiry. Quantitative findings suggest that students from the experimental class ( $n = 25$ ) engaged in higher-level epistemic moves than those from the comparison class ( $n = 25$ ). Qualitative data analysis illustrated the collaborative trajectory with the scaffold in STEM learning.

### Epistemic agency for STEM learning with design-based scaffold

Students were expected to assume epistemic agency to deal with the uncertainty and diverse ideas in STEM learning. However, a significant portion of STEM education in higher education still adheres to traditional instructional models that restrict students' agency over their own learning (Apkarian et al., 2021). Thus, we are interested in exploring the ways that design-based scaffolds can be leveraged to create more opportunities for epistemic agency in higher education STEM.

The design-based scaffold included two parts, "problem-scoping" scaffold and "iteration" scaffold. The "problem-scoping" scaffold were developed to help students in understanding the boundaries of problems and identifying problems constrains and goals. The scaffolds included the following prompts: "What is the core problem?", "What information is needed to explain the problem?", "What are the constraints?", and "What ideas do you have for problem solving". The "iteration" scaffold was provided to students for iterative refinement including "What works have been developed so far (text, images, and videos)?", "Does it meet the constraints of the core problems?", "What are the problems with the current work?", "What measures can be taken to improve the work?", and "What are your additions and refinements in response to others' ideas?".

Two questions were asked: (1) Did students in the experimental class performed high-level epistemic moves? (2) How did student engage in collaborative trajectory with the design-based scaffolds?

### Methods

The study involved a quasi-experimental design that conducted in two undergraduate classes "STEM Education" at a university in South China. In phase I, students were encouraged to propose diverse ideas regarding "Pandemic," "Rocket," "Bridge," and "Culture" and were divided into four groups based on their interests. In phase II, students first formulated and scoped authentic problems with "problem-scoping" scaffold and engaged in artifacts production with "iteration" scaffold. Students document the process of artifacts production and reflected on the documentation to identify error and deficiencies, check the meeting of constraints, goal, and attainment, and develop strategies for improvement, add new ideas to the design, and transform the products into better ones. In phase III, students collectively reflected on artifacts and STEM understanding that they gained in STEM learning.

For the comparison class, students engaged in STEM learning following same procedure without scaffolds.

### Finding

#### RQ1: Difference on Epistemic Moves

The MANOVA analysis (Figure 1) demonstrated that students in experimental class engaged higher-level epistemic moves (connection  $F(1, 48) = 6.441, p < .05$  and synthesizing,  $F(1, 48) = 24.067, p < .01$ ) compared to comparison class (simple claims  $F(1, 48) = 11.159, p < .05$ ).

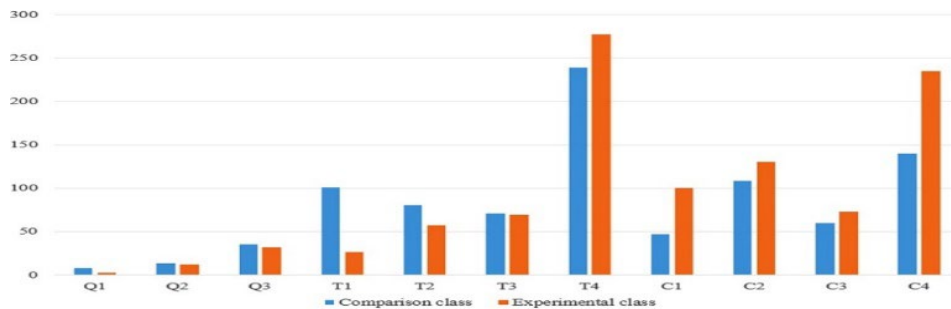
#### RQ2: Epistemic Agency for Collaborative Knowledge-Building Practices with Design-Based Scaffolds

Six themes with a specific case, "Bridges," are introduced to characterize the processes that collaborative STEM inquiry in knowledge-building environment.

*Theme 1: Proposing diverse ideas and forming groups. S2: What are the types of bridges? What materials are needed for construction?*

#### Figure 1

*Comparison of Epistemic Moves between the Two Classes*

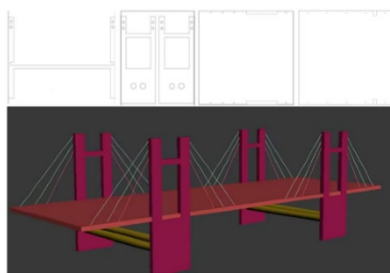


Questioning: Q1: Fact-seeking, Q2: Explanation-seeking, Q3: Sustained inquiry; Theorizing: T1: Simple claim, T2: Proposing an explanation, T3: Supporting an explanation, T4: Improving an existing explanation; Community: C1: Connection, C2: Lending support, C3: Problems detecting, C4: Synthesizing notes)

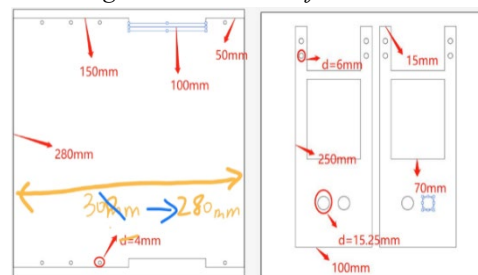
Theme 2: Scoping the core problems with problem scoping scaffolds. S2: [What are the core problems?] How is the bridge load-bearing? S1: [Constraints?] What are the factors that affect the load-bearing capacity, such as size and material, that need to be considered? How much tonnage can be supported?]

Theme 3: Ideation and Sketching. S4: To achieve our goals in the experiment, there was a reasonable division of labor with a designated note-taker, a modeling analyst, and a draughtsman. Everyone actively contributed their personal opinions, raised any remaining doubts, and other members provided active clarification. Any areas of uncertainty were subsequently documented and resolved (Figure 2).

**Figure 2**  
Initial Sketches of Bridge Structures



**Figure 3**  
Visualizing Documentation of Error Detecting



Theme 4: Modeling and transforming design into products. S6: [“What are the problems with the current work”] Due to certain deviations in the dimensions of the cut-out parts, there are certain gaps between the parts, the slots and holes. S3: [Measures for improvement] Based on the previous records of sketches, the following modifications can be made, see the picture (Figure 3).

Theme 5: Reflecting and improving for final artifacts. S3: What are the problems with the current work] - revise the design and explore what is appropriate in the mortise and tenon construction? S4: [Measures for improvement] Choose a triangular shape for a more stable mortise and tenon construction (Figure 4).

Theme 6: Collective reflection on the trajectory of collaborative design.

**Figure 4**  
Artifact from the “Bridges” Group



## References

Apkarian, N., Henderson, C., Stains, M., Raker, J., Johnson, E., & Dancy, M. (2021). What really impacts the use of active learning in undergraduate STEM education? Results from a national survey of chemistry, mathematics, and physics instructors. *PLoS one*, 16(2).

## Co-Designing Synchronous Asymmetric Multiplayer Learning Game Scenarios

Gaëlle Guigon, Sorbonne Université & IMT Nord Europe, France, [gaelle.guigon@imt-nord-europe.fr](mailto:gaelle.guigon@imt-nord-europe.fr)  
Mathieu Vermeulen, IMT Nord Europe, France, [mathieu.vermeulen@imt-nord-europe.fr](mailto:mathieu.vermeulen@imt-nord-europe.fr)  
Mathieu Muratet, Sorbonne Université, France, [mathieu.muratet@lip6.fr](mailto:mathieu.muratet@lip6.fr)  
Thibault Carron, Sorbonne Université, France, [thibault.carron@lip6.fr](mailto:thibault.carron@lip6.fr)

**Abstract:** This article presents a qualitative experiment with instructional designers. It aimed at designing asymmetrical multiplayer role-playing scenarios using a digital visual modeling tool with two designers. We want to know if co-design with the tool RLG Maker can assist in developing scenarios and what the obstacles are. We assume that collaboration during the design phase encourages the construction and enrichment of the scenario. We have analyzed the speech acts of four groups of instructional designers and have confirmed that the tool, used collaboratively, encourages the scripting of these games. Some aspects need to be reworked, in particular navigation in the tool, to improve usability.

### Introduction

Serious games are known as an interesting way of reproducing real-life situations (Stroe et al., 2019). There are a number of barriers to their design, including the difficulty of designing a scenario that is suited to the educational objectives being pursued (Marne & Labat, 2014). One promising option is multiplayer serious games, especially those with multiple objectives distributed between the different participants, each of whom playing a different role. This is what we call an asymmetric game. Thus, our research focuses on the design of asymmetric role-playing scenarios for serious games by instructional designers and teachers. Asymmetric role-playing games for learning, that we call RLG (Role Learning Games), seem to be rare and there are few resources (especially models or methods) for designing them (Guigon et al., 2021). The work presented in this paper follows previous research which consisted in developing a model to design this type of game and reifying this model in the form of a tangible kit (Guigon et al., 2023). In this paper we focus on the following general problem: how can we design a digital visual modeling tool to facilitate for two designers the creation of asymmetrical learning game scenarios? To tackle this problem, we have developed a digital version from the tangible kit, called RLG Maker. This is a visual modeling tool which helps non-computer scientists design asymmetric role-playing scenario. We conducted an experiment with part of our target audience: instructional designers. These people are used to working with teachers, particularly in higher education, and are generally familiar with digital tools. The aim of this experiment is to evaluate if co-designing with RLG Maker can assist in developing scenarios and determine what the obstacles are. We suppose that collaboration on design with this type of tool would make it easier for non-experts to script this type of game.

### Method

This experimentation brought together eight members of the educational teams from a group of *Grandes écoles* in France. The session lasted a total of 90 minutes, with a 20-minute briefing, 45 minutes of tool testing in pairs followed by a 25-minute debriefing conducted in the form of a focus group. Each pair worked on a computer and had access to the online tool (<https://rlgmaker.imt-nord-europe.fr/>), as well as a paper guide to using RLG Maker. This tool guides designers step by step through the process of creating an asymmetric role-playing scenario for teaching purposes. In order to conduct a thematic analysis, each team had an audio-recording device, and the room was video-recorded. At the end of the session, participants filled out the System Usability Scale (SUS) (Brooke, 1996) questionnaire to assess the usability of RLG Maker. Furthermore, to analyze the work done on the tool, the files generated during the session were collected (in json format). With these files, it is possible to open again the final scenario file of each group and also to determine the sequence of actions taken within the tool. Finally, we coded the transcripts to answer our research question and focused them on two themes: the verbatims in favor of collaborative learning, and those about utility and usability.

### Findings and discussion

The thematic analysis according to (Attride-Stirling, 2001) highlighted the following main results.

#### Focus 1: Collaborative learning

The thematic analysis of the discussions reveals that there is a significant amount of interaction between the teams, and these interactions are diverse. The members of each group were given the opportunity to speak on an equal footing, each was concerned about the other's opinion and wanted to move forward together by helping each other. The roles were well distributed and everyone found their place: either to integrate the content into the tool, or to give the information to be integrated or to guide the other in the use of the tool since they generally had the paper guide available. This confirms our intuition that collaboration is useful in this type of work.

## Focus 2: The utility and usability of the tool

The verbatims referring to this question show that in the vast majority of cases, users understand how the tool works. To illustrate this, on many occasions a tester gave advice on how to use the tool to their partner. The main problem was ergonomic: the tool is designed to be used mainly with a mouse, so browsing was difficult with a pad. The average of the eight responses to the SUS questionnaire had a usability score of 60.3/100. This is a fair score, but there is still room for improvement, particularly from an ergonomic point of view. Seven out of eight people answered at the end of the experiment that the tool was useful, usable and acceptable, while only one answered that it was useful, acceptable but not usable because of the navigation problem (confirmed by audio recording).

## Conclusion and future work

This paper presents an experiment with a visual modeling tool, called RLG Maker, for scripting asymmetrical role-playing learning games with eight instructional designers. Our problem was: how can we design a digital visual modeling tool to facilitate the creation of asymmetrical learning game scenarios for two designers? To tackle this problem, we have worked on the question: how can co-design with RLG Maker can assist in developing scenarios and what the obstacles are. This question was addressed through an experiment that focused on two specific lines: collaboration between co-designers and the utility and usability of the tool. These two focuses of the experiment revealed that the tool seemed capable of producing an asymmetrical learning game scenario as intended. Nevertheless, some usability problems hampered the design, such as browsing difficulties using the laptop pad. Improvements are already being implemented and will be tested during a future test session, this time involving teachers.

## References

- Attride-Stirling, J. (2001). Thematic Networks: An Analytic Tool for Qualitative Research. *Qualitative Research - QUAL RES*, 1, 385–405. <https://doi.org/10.1177/146879410100100307>
- Brooke, J. (1996). SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.
- Guigon, G., Muratet, M., Vermeulen, M., & Carron, T. (2023, July). Facilitate the design of Role Learning Games: The RLG Kit. *International Conference on Advanced Learning Technologies and Technology-Enhanced Learning*. <https://hal.science/hal-04138615>
- Guigon, G., Vermeulen, M., Muratet, M., & Carron, T. (2021). Towards an Integration of the Multi-role Dimension in the Design of Learning Games: A Review of the Literature. *Games and Learning Alliance* (pp. 258–264). Springer International Publishing. [https://doi.org/10.1007/978-3-030-92182-8\\_26](https://doi.org/10.1007/978-3-030-92182-8_26)
- Marne, B., & Labat, J.-M. (2014). Model and Authoring Tool to Help Teachers Adapt Serious Games to their Educational Contexts. *International Journal of Learning Technology*, 9(2), 161. <https://doi.org/10.1504/IJLT.2014.064491>
- Stroe, I. P., Ciupe, A., Meza, S. N., & Orza, B. (2019). FireEscape: A Gamified Coordinative Approach to Multiplayer Fire-Safety Training. *2019 IEEE Global Engineering Education Conference (EDUCON)*, 1316–1323. <https://doi.org/10.1109/EDUCON.2019.8725148>

# Training an Avatar-Based Conversational AI Chatbot and Investigating the Implementation of Communicative Action Principles

Erum Afzal, Justus-Liebig-University Gießen, erum.afzal@erziehung.uni-giessen.de  
Julia Fecke, Justus-Liebig-University Gießen, julia.fecke@zfl.uni-giessen.de  
Lenuta Alboae, Alexandru Ioan Cuza University of Iasi, adria@info.uaic.ro  
Edith Braun, Justus-Liebig-University Gießen, edith.braun@erziehung.uni-giessen.de

**Abstract:** Conversational AI shows promising outcomes in developing essential communicative skills for prospective teachers. However, existing models often lack training in educational and social science theories, like Habermas' communicative action design theory. To bridge this gap, the aim of our research is to train an AI avatar-based system using Knowledge Graph (KG) technology for developing authentic communicative action skillset among early-stage professionals. The first insight into the effectiveness of Conversational AI in developing these skills has been demonstrated in pilot projects through role-play simulations on Avatar-based learning platforms. This synergistic approach aims to generate more authentic and human-like conversations.

## Introduction and theoretical background

With strong communication abilities, educators can engage students effectively, establish supportive classroom environments, and convey knowledge with clarity and precision. Simulation-based training has emerged as a well-suited method for developing communicative skills among prospective teachers. By immersing students of teacher education in virtual learning environments, such as 3D avatars and agent-based interactions, learners can practice and refine their conversational abilities in a safe and controlled setting (Fecke, Afzal & Braun, 2023). Leveraging chatbots and AI-driven systems, Conversational AI provides personalized learning experiences, and assistance with queries, thereby holding significant implications for teacher education. Despite the potential of Conversational AI, existing models often lack a solid theoretical foundation in educational and social science theories. This gap poses a significant challenge in harnessing the full potential of AI in education, particularly in the authentic application of conversational AI in pedagogical contexts. Integrating established theoretical frameworks with Knowledge Graph, such as Habermas' communicative action design theory, could address this limitation. Habermas' theory emphasizes the distinction between strategic and understanding-oriented communicative action designs, highlighting the importance of mutual understanding in teacher-student interactions. In light of these challenges and opportunities, this research aims to bridge the gap by integrating a knowledge graph developed on Habermas' communicative action design theory with the Large Language Model (LLM). In addition, the utilization of a knowledge graph offers a structured representation of domain-specific knowledge, enabling the AI system to access and utilize relevant information effectively (Peng et al., 2023). This synergistic approach addresses the constraints of Conversational AI, making it more adaptive and responsive to the needs of educational professionals across various social science disciplines.

## Research questions

Reflection on the theoretical background following Research Question is framed.

How can an AI System, utilizing Knowledge Graph (KG) technology, be effectively trained to develop the communicative action skillset among prospective teachers, given the current deficiencies in existing models regarding educational and social science theories.

## Developing knowledge graph embedded systems

The development and implementation of a knowledge graph architecture in Protégé based on Habermas's communicative action framework offer a robust framework for structuring and organizing information relevant to training programs. The knowledge graph architecture is designed to represent the ontology accurately based on Habermas's communicative action framework for training programs. For this, we are preparing a dataset of video recordings in which students performed role-play simulations with a professional trained actor based on Habermas communicative action. In total we have 10 different role-play scenarios (Braun et al., 2018) with typical situations from the everyday working life of a teacher. From the simulation scripts source and Target entities are extracted as shown in table along with the relation of each participatory Entity.

**Table 1**  
*Table Representing Relationship between Source and Target Entities*

Source	Relationship	Target
Actor	plays	Role
Characteristic	has	CommunicationType
CommunicationType	has	Characteristic
Guide	hasAssociatedRolePlay	RolePlay
Guide	has	Instructions
Guide	hasBackgroundInformation of	RolePlay
Guide	has	Task
Guide	has	Objective
Guide	has	RelationalGoal
Role	has	RolePlay
RolePlay	has	CommunicationType
RolePlay	hasObjective	Objective

The ontology encompasses classes such as *RolePlay*, *CommunicationType*, *Role*, *Actor*, *Guide*, and *Characteristic*, along with their respective instances and properties. For instance, *RolePlay* instances include various simulation role plays categorized into Understanding-oriented and Strategic communication types, each associated with specific objectives, scenarios, and guides for actors and students. The relationships as shown in *Table 1* between these classes, such as *belongsToCommunicationType* linking *RolePlay* to *CommunicationType*, and *hasRole* connecting *RolePlay* to *Role*, ensure the accurate representation of the ontology's structure. Overall, the ontology structure facilitates better organization, understanding, and querying of information related to the role play simulations in teaching training programs. Integrating Habermas' Communicative Action Design Theory into Conversational AI by leveraging the large language model (LLM) and knowledge graphs in the system architecture aims to create a powerful and contextually relevant educational tool for prospective teacher.

## Conclusion

In summary, this research serves as a pivotal step in bridging the gap between social science theories, like communicative action design, and the application of conversational AI. Our study highlights a significant gap in existing AI models, lacking a foundation in educational and social science theories. By integrating Habermas' communicative action design theory, our approach, using a knowledge graph aims to revolutionize AI-driven conversations, making them more authentic and human-like. Future work involves thorough validation in educational settings, personalized options, adaptive feedback mechanisms, and enhanced privacy measures.

## References

- Braun, E. (2021). Performance-based assessment of students' communication skills. *International Journal of Chinese Education*, 10(1). <https://doi.org/10.1177/22125868211006202>
- Braun, E., Athanassiou, G., Pollerhof, K. & Schwabe, U. (2018). Wie lassen sich kommunikative Kompetenzen messen? – Konzeption einer kompetenzorientierten Prüfung kommunikativer Fähigkeiten von Studierenden. *Beiträge zur Hochschulforschung*, 40(3), 34-55.
- Fecke, J., Afzal, E., & Braun, E. (2023). A Conceptual System Design for Teacher Education: Role-Play Simulations to Train Communicative Action with AI Agents. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023*, pp. 2197-2200. International Society of the Learning Sciences.
- Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11), 13071-13102.

## Character Based Models and Computational and Embodied Action Tweaking for Sensemaking

Joshua A. Danish, Nitasha Mathayas, Mengxi Zhou, Selena Steinberg, Morgan Vickery  
jdanish@iu.edu, mathaya2@illinois.edu, mz13@iu.edu, ssteinb@iu.edu, moravick@iu.edu  
Indiana University

**Abstract:** Embodied computational models where students can use their bodies to help explore how a phenomena works are quite powerful. Yet there remain challenges in how to help learners attend to important aggregate outcomes and underlying mechanisms. This poster documents our design efforts to explore how different activities where learners can “tweak” their models with script or action help orient them to aspects of their models in productive ways.

### Introduction and background

Modeling, the process through which individuals construct, refine, interrogate and engage in inquiry with a representation of a phenomenon is central to doing and learning science (Schwarz et al., 2009). While many different representational forms can be used in modeling, such as drawings, graphs, mathematics and code (Pierson et al., 2022), we focus on embodied computational models (Danish et al., 2020). Embodied computational models leverage a mix of theoretical advances which highlight the value of bodily experiences and movement in supporting learning, and of advances in technology that allow the tracking of such movement to integrate movement into computer models and simulations (Danish et al., 2022; Lindgren & Johnson-Glenberg, 2013). We see these types of models as unique spaces for supporting learning and building theory about embodiment.

We have seen consistently in our work, and in the work of others, that embodied models require a great deal of coordination among students, which is both productive and potentially challenging (Danish et al., 2020). This work highlights the need to focus on how we support reflection in embodied modeling contexts, helping learners to sense-make through and about the models they are co-constructing with their peers and the system. One solution we have been exploring to support sensemaking while creating embodied computational models is the integration of scripting or simplified coding into the learning environment. Our goal in this effort has not been to shift gears to programming. While we value programming and computational thinking and efforts to integrate them more broadly in formal education (Grover & Pea, 2018), we also recognize the time and expertise they may demand, and the immense time pressures that teachers already face. Therefore, our approach is inspired by Guzdial et al’s (2023) notion of “teaspoons of code” that are powerful but easily approachable. This led us to focus on how learners can “tweak” their models, both by changing snippets of code and by adjusting their actions as a way of making sense of the models and underlying mechanisms in the phenomena without needing to learn to code.

We share analysis of interactions from three implementations that are part of the same ongoing design-based research project with the GEM-STEP software platform (Danish et al., 2022) to illustrate this approach and explore how it supports sensemaking amongst groups of students. Our research question is: How do tweaks in script and in planned actions support learners’ sensemaking in an embodied computational learning environment?

### Methods

The present analysis draws on data from three classroom implementations of embodied ecosystem modeling activities using the GEM-STEP software platform (Danish et al., 2022). In the first two, students were able to embody fish in a pond ecosystem, and rabbits or worms in a garden ecosystem. In the third, students embodied beavers in a wetland ecosystem. Each student controls a character (e.g., a beaver) in a GEM-STEP model by moving in the physical space with a tracking tag or using an iPad. We refer to them as characters rather than agents to highlight how students had agency in controlling them, and often connected their movements to an emerging narrative about the model. The entire model is displayed for all students in the front of the classroom. Students are organized in groups, taking turns to be part of the model or act as observers, and then collectively discuss the emergent model. The GEM-STEP models are developed in a custom scripting language (GEM-SCRIPT), and students can use this script to change how the elements of the model work. Given our focus on supporting iterative modeling cycles and not programming, we designed activities where students would only need to view or edit small sections (“teaspoons”) of just a few lines of code to promote reflection about how the modeled ecosystem works. To examine the interplay between working with script and acting differently to change the model, we first logged all the activities across the implementations. We then focused on the major activities and engaged in Interaction Analysis to explore how each supported students’ modeling in unique ways.



## Findings and discussion

We present summaries of the key interactions with the models to help understand the unique contributions of the “tweaking”

1. *Initial embodied modeling*: When students first approached an embodied model they tended to focus on their own character, and how their actions influenced the character’s behavior. In most of our models, the characters had an energy “meter” and students focused on trying to avoid dying due to lack of energy.
2. *Reflecting and refining a model*: When asked to reflect on the initial model and then change their behaviors to explore new questions or issues (e.g., staying alive longer), students typically focused on how to move their own individual characters differently. Discussions rarely focused on mechanisms.
3. *Examining code snippets*: When presented with code that was part of the model either in small groups or as part of the whole class, students appeared to enthusiastically discuss how the system worked, sometimes referring to their prior embodiment. This increased their focus on the mechanisms in the system. Using a jigsaw method, we were able to leverage this approach to facilitate cross-group reflection about different aspects of the model. Comments within the code also enhanced discussion of this process.
4. *Planning code tweaks*: When discussing how to change one or more variables, students engaged in debates about what variables were most important and hypothesized about how different properties would affect the system. They often used quantities in a rather fluid way, picking larger or smaller numbers without articulating why those specific numbers were chosen. Having seen that code tweaking was possible, students also began suggesting it in later activities even when it was not explicitly offered.
5. *Planning action tweaks*: When discussing how to behave differently, students attended to how different actions of individual characters would lead to new aggregate outcomes. When asked to pair these with code tweaks, they explored how their actions could capitalize on the new code to help the ecosystem.
6. *Reflecting on code changes*: Students appeared to attend closely to the potential outcomes of their code changes, focusing on whether / how they worked or not. They often attended to aggregate outcomes. They also made suggestions for future changes based on these interactions.

Looking across activities, helping learners to frame their behavior as modeling gradually helps them engage with the underlying mechanisms. However, the game-like nature of the environment, and the act of taking on an individual role also meant students didn’t always attend to the mechanisms or aggregate outcomes. Fortunately, the opportunity to “tweak” the code or actions in their models helped re-orient them toward mechanisms.

## References

- Danish, J. A., Enyedy, N., Saleh, A., & Humburg, M. (2020). Learning in embodied activity framework: A sociocultural framework for embodied cognition. *International Journal of Computer-Supported Collaborative Learning*, 15(1), 49–87. <https://doi.org/10.1007/s11412-020-09317-3>
- Danish, J. A., Anton, G., Mathayas, N., Jen, T., Vickery, M., Lee, S., Tu, X., L., Cosic, Zhou, M., Ayalon, E., & others. (2022). Designing for Shifting Learning Activities. *The Journal of Applied Instructional Design*, 11(4).
- Grover, S., & Pea, R. (2018). Computational thinking: A competency whose time has come. *Computer Science Education: Perspectives on Teaching and Learning in School*, 19(1), 19–38.
- Guzdial, M., Dodoo, E., Naimpour, B., Nelson-Fromm, T., & Padiyath, A. (2023). Putting a Teaspoon of Programming into Other Subjects. *Communications of the ACM*, 66(5), 24–26. <https://doi.org/10.1145/3587926>
- Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality. *Educational Researcher*, 42(8), 445–452. <https://doi.org/10.3102/0013189x13511661>
- Pierson, A. E., Keifert, D. T., Lee, S. J., Henrie, A., Johnson, H. J., & Enyedy, N. (2022). Multiple Representations in Elementary Science: Building Shared Understanding while Leveraging Students’ Diverse Ideas and Practices. *Journal of Science Teacher Education*, 0(0), 1–25.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.

## Acknowledgments

We would like to thank the students and teachers who supported this work. We also thank our collaborators on the GEM-STEP project. Portions of this work were funded by the National Science Foundation.

# Explore Preservice Teacher's Interaction Pattern Mediated by Teacher Dashboard in the Video-Based Professional Learning Communities

Huiying Cai, Jiangnan University, caihy@jiangnan.edu.cn  
 Linmeng Lu, East China Normal University, linlu0705@gmail.com,  
 Xiaoqing Gu\*, East China Normal University, xqgu@ses.ecnu.edu.cn

**Abstract:** In order to examine the effect of teacher dashboard generated through intelligent analysis of classroom videos, the quasi-experiment of pilot study was conducted in a university in Eastern China. 48 preservice teachers were split into control and experimental groups. In the experimental condition, their learning process was mediated by teacher dashboard in the video-based professional learning communities. After analyzing group teacher collaborative discourse, it revealed insights into interaction patterns, validating the dashboard's effectiveness in the video-based professional learning communities.

## Introduction

Simulative and sociocultural theorists argue that collaborative examination of teaching practice in Professional Learning Communities (PLC) enhances teacher learning experiences (De Jong, Meirink, & Admiraal, 2022). Right now, video-based PLC, where teachers analyze and discuss teaching videos, fosters deep reflection. (Kang & van Es, 2019). However, the challenge of overwhelming data inherent in classroom video due to the long and excessive content of videos would be not beneficial for preservice learning (Chen, 2020). A solution is a teacher dashboard visualizing classroom video information, providing data-based insights. In this study we intend to explore the effectiveness of teacher dashboard generated through intelligent analysis of classroom videos using artificial intelligence on teacher learning. The research questions are: (1) What are the effects of teacher dashboard on teacher's interaction process during video-based PLC? (2) how did teacher dashboard affect teacher's interaction pattern during video-based PLC?

## Methods

In this study, a quasi-experiment pilot study was conducted in a university in Eastern China. 48 preservice teachers were divided into 16 groups: 8 groups in a control condition (CC) and 8 groups in an experimental condition (EC), with 3 preservice teachers in each group. There was no significant difference in teachers' reflection ability between the two conditions ( $t=0.835$ ,  $p>0.05$ ). The AI-based platform named Quanjin was utilized to analyze classroom videos focusing on teachers' instructional behavior, teacher-student interaction, and students' emotions. The teacher dashboard of classroom videos was generated, as illustrated in Figure 1.

**Figure 1**  
The Interface Example of Teacher Dashboard

### 2. Teacher-student interaction

#### > (3) Types of teacher-student interaction

Using intelligent analysis technology, we can calculate the number of three types of behavior, including teacher-led behavior, student behavior and teacher-student interaction behavior, by analyzing the teacher-student dialogue discourse in classroom teaching. On this basis, we can calculate the share of teacher behavior in the total number of classroom behaviors, and the share of teacher-student interaction behavior in the total number of classroom behaviors.

According to the values of the rate of teachers' behavior and the rate of the teacher-student interaction, we can divide the types of teacher-student interaction into four types: practice type, lecture type, dialogue type and mixed type. The specific contents are shown in Table 1.

Table 1 Types and characteristics of teacher-student interaction mode in classroom

Types of teacher-student interaction	Behavior characteristics of teacher student interaction mode
Practice type	The occupation rate of teachers' teaching behavior is low, and students' activities to complete exercises are the main activities. The impact on students' effective learning is low.
Lecture type	The occupation rate of teachers' teaching behavior is high, and the interaction between teachers and students is low.
Dialogue type	The number of teacher-student interaction is relatively high, indicating that the interaction between teachers and students is sufficient.
Mixed type	The proportion of teachers' behavior is around half, and teachers and students are fully involved.

The type of teacher-student interaction of this lesson is belong to dialogue type.

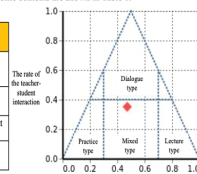


Figure 1 type diagram of teacher-student interaction mode

**Table 1**  
Coding Scheme for Teacher Collaborative Discourse

Learning event during interaction	Description	Code
Unrelated learning activities	Teachers express concern, encouragement, and discuss topics unrelated to the task.	P1
Group planning	Teachers pay attention to how to divide the tasks and how to carry out the tasks.	P2
Group monitoring	Teachers pay attention to monitor the progress of the collaborative task, whether the current task is completed, and whether they can enter the next task.	P3
Describe instructional information	The teacher describes the information about the classroom video from the perspective of the observer.	P4
Sharing/comparing information or proposing similar ideas	Teachers express their understanding approval or doubt of the teaching events in the classroom video.	P5
Negotiation of meaning and co-construction of knowledge	Teachers discuss their views interactively from the perspective of teaching effect.	P6
Testing and modification of proposed synthesis or co-construction	Teachers reach a common understanding of the point of view.	P7
Agreement statement or application of newly constructed meaning	Teachers come up with plans and plans to improve classroom teaching.	P8

In the study, the researchers introduce experiment backgrounds, and each group member logs into the Quanjing platform. During implementation, preservice teachers in CC and EC watch classroom videos on Quanjing, discuss via QQ software to evaluate the instruction design, and redesign lesson plans on the

collaborative website Jinshan. Unlike CC group teachers, EC group teachers use the Quanjing-generated teacher dashboard for discussions. Finally, each group submits a revised instructional design.

In this study, data from teacher collaborative discourse recorded in QQ will be analyzed to investigate interaction patterns in PLC. The coding unit was a learning event, a dialogue episode with uninterrupted interaction and consistent semantic content (Cai, Lin, & Gu, 2016). The coding framework for analyzing teacher collaborative discourse was proposed by Gunawardena, as shown in Table 1 (Gunawardena et al., 2007). Chi-square and lag sequence methods were used to analyze group teacher collaborative discourse, revealing a consistency coefficient of over 83% among three coders.

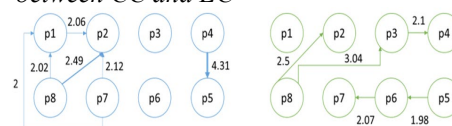
## Findings and discussion

As shown in Table 2, there is a significant difference investment time of the specific learning event during collaboration between CC and EC, with  $\chi^2=2675.00$ ,  $p=0.00<0.05$ . It indicated that the mediation of a teacher dashboard positively influences their interaction process in video-based PLC by encouraging them to engage high-level interactive learning activities such as sharing and negotiating information, while spending less time on low-level tasks such as group planning and task division.

**Table 2**  
*Chi-Square Analysis Results of Interactive Behavior Investment Time between CC and EC*

	P1 <sup>±</sup>	P2 <sup>±</sup>	P3 <sup>±</sup>	P4 <sup>±</sup>	P5 <sup>±</sup>	P6 <sup>±</sup>	P7 <sup>±</sup>	P8 <sup>±</sup>
CC <sup>±</sup>	5808 <sup>±</sup> (11.91%) <sup>±</sup>	5154 <sup>±</sup> (10.57%) <sup>±</sup>	1950 <sup>±</sup> (4.00%) <sup>±</sup>	3720 <sup>±</sup> (7.63%) <sup>±</sup>	16667 <sup>±</sup> (34.19%) <sup>±</sup>	6303 <sup>±</sup> (12.93%) <sup>±</sup>	2706 <sup>±</sup> (5.55%) <sup>±</sup>	6440 <sup>±</sup> (13.21%) <sup>±</sup>
EC <sup>±</sup>	3549 <sup>±</sup> (10.92%) <sup>±</sup>	1433 <sup>±</sup> (4.41%) <sup>±</sup>	2851 <sup>±</sup> (8.77%) <sup>±</sup>	1691 <sup>±</sup> (5.20%) <sup>±</sup>	11822 <sup>±</sup> (36.37%) <sup>±</sup>	6058 <sup>±</sup> (18.63%) <sup>±</sup>	859 <sup>±</sup> (2.64%) <sup>±</sup>	4246 <sup>±</sup> (13.06%) <sup>±</sup>

**Figure 2**  
*Interaction Behavior Transition Diagram between CC and EC*



The results of lag sequence methods are depicted in Figure 2. It reveals two distinct interaction patterns between the two groups. Firstly, compared to CC group teachers, EC group teachers focused more on the reasoning process when viewing classroom videos. For instance, in CC, only one significant interaction path related to group teachers' reasoning process was observed, P4→P5, Z score=4.31. However, in EC, three significant interaction paths were identified (P3→P4, Z-score=2.1; P5→P6, Z-score=1.98; P6→P7, Z-score=2.07). Secondly, EC group teachers exhibited better task cohesion and higher task engagement compared to CC group teachers. In EC, every interaction path was significantly related to learning-related activities, whereas in CC, three interaction paths were not (P1→P2, Z-score=2.06; P7→P1, Z-score=2; P8→P1, Z-score=2.02). It suggests a positive mediating effect of the teacher dashboard on teachers' interaction patterns in video-based PLC.

The results showed the positive meditative effect of the teacher dashboard generated by AI analysis on facilitating teachers' interaction patterns in video-based PLC. Future research should explore practical interventions during AI-supported video-based professional learning communities, such as providing prompts or hints to support teacher understanding of the pedagogical significance of indicators extracted from AI technology.

## References

- Cai, H., Lin, L., & Gu, X. (2016). Using a semantic diagram to structure a collaborative problem solving process in the classroom. *Educational Technology Research and Development*, 64(6), 1207-1225.
- Chen, G. (2020). A visual learning analytics (VLA) approach to video-based teacher professional development: Impact on teachers' beliefs, self-efficacy, and classroom talk practice. *Computers & Education*, 144, 103670.
- De Jong, L., Meirink, J., & Admiraal, W. (2022). Teacher learning in the context of teacher collaboration: connecting teacher dialogue to teacher learning. *Research Papers in Education*, 37(6), 1165-1188.
- Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of educational computing research*, 17(4), 397-431.
- Kang, H., & van Es, E. A. (2019). Articulating design principles for productive use of video in preservice education. *Journal of Teacher Education*, 70(3), 237-250.

## Acknowledgments

This research was funded by National Social Science Fund of China (Grant No. CCA220318).

# Opening the Black Box of the Effects of Instructional Support on Collaborative Diagnostic Processes with Process Mining

Katharina M. Bach, Andreas Wildner

K.Bach@psy.lmu.de, Andreas.Wildner@psy.lmu.de  
Department of Psychology, LMU Munich, Germany

Martin R. Fischer, University Hospital, LMU Munich, Germany, martin.fischer@med.uni-muenchen.de

Frank Fischer, Olga Chernikova, Constanze Richters

Frank.Fischer@psy.lmu.de, O.Chernikova@psy.lmu.de, Constanze.Richters@psy.lmu.de  
Department of Psychology, LMU Munich, Germany

**Abstract:** This study investigates the role of reflection phases and collaboration scripts in collaborative diagnostic reasoning in an agent-based simulation among 117 medical students working on a fictitious patient case. Through process mining, we found that reflection fostered evidence generation and evidence and hypothesis sharing, whereas collaboration scripts led to skipping collaborative diagnostic activities. Neither significantly affected the diagnostic accuracy. The findings challenge the focus on conventional outcome measures like accuracy.

## Collaborative diagnostic reasoning in medicine and how to foster it

According to the Collaborative Diagnostic Reasoning (CDR) model (Radkowitz et al., 2020), evidence generation, drawing conclusions, evidence elicitation, evidence sharing, and hypothesis sharing are collaborative diagnostic activities (CDAs). Simulation-based learning environments can incorporate scaffolding to foster medical students' CDAs. Reflection phases encourage learners to think in-depth about the relevant activities (Ibiapina et al., 2014). As a type of socio-cognitive scaffolding, external collaboration scripts can organize and improve collaboration (Vogel et al., 2017). However, traditional inferential statistics exclusively examine the effects of such scaffolding on CDAs or diagnostic accuracy as quantitative outcome measures, disregarding potential differences in CDR processes that can lead to the same outcome. In contrast, process analyses offer insights into mechanisms behind scaffolding effects (Stadler et al., 2020). This study aimed to open the black box of CDR processes by addressing the following research questions: 1) How do collaboration scripts and reflection phases change learners' CDR processes, and 2) affect diagnostic accuracy during learning in an agent-based simulation?

## Methods

Participants ( $N = 117$ ) were intermediate medical students ( $n_{female} = 81$ ,  $n_{male} = 36$ ;  $M_{age} = 25.34$ ,  $SD_{age} = 3.18$ ) who acted as trainee doctors, working on a fictitious text-based patient case. The simulation-based learning environment collected log data of students' CDAs. First, the students studied the patient's health record (evidence generation). Then, they requested radiological examinations from an agent-based radiologist (evidence elicitation) and had to justify their request with evidence (evidence sharing) and suspected diagnoses (hypothesis sharing). If successful, they received the examination results (evidence generation) and could request additional information, review the results, request a new examination, or proceed to the final case solution, where they were asked to state their most likely suspected diagnosis (drawing conclusions). Otherwise, they had to revisit their original request.

Students were randomly assigned to one of three conditions: In the reflection condition ( $n = 43$ ), they received additional questions before collaborating, encouraging them to think about the case, their suspected diagnosis, and alternative diagnoses. Students in the collaboration scripts condition ( $n = 39$ ) received prompts at the beginning of the interaction with the radiologist. If they insufficiently justified a request, they also received prompts adapted to the source of their error. In the control condition ( $n = 35$ ), students received no scaffolding.

At first, we coded the log files according to the CDAs and evaluated the final suspected diagnosis as correct (1) or incorrect (0) as a measure of diagnostic accuracy. Then, employing process mining, we used ProM 6.6 to generate a process model for each condition from the log files. The models depicted the possible paths between the CDAs students followed. Moreover, we fitted a logistic regression model using *R* with diagnostic accuracy as the outcome and condition as the independent variable.

## Results and discussion

In the control and collaboration scripts conditions, students generated evidence with the health record after starting their first request, resulting in minimal rechecking later. The reflection condition facilitated evidence generation

most; students viewed the health record repeatedly. Thus, the entire model represented one big loop, distinguishing it from the other conditions. When comparing evidence elicitation, it becomes apparent that students in the control and the collaboration scripts conditions frequently skipped parts of the request form. In contrast, students in the reflection condition diligently completed most of their requests and shared evidence and hypotheses with the radiologist. Readers can find the three process models under the following link: [https://osf.io/53zg7/?view\\_only=53e93cc2edfb43ac84a79325252d361d](https://osf.io/53zg7/?view_only=53e93cc2edfb43ac84a79325252d361d). The quantitative analyses showed that neither collaboration scripts ( $b = -0.04, p = .240$ ) nor reflection phases ( $b = 0.12, p = .801$ ) significantly affected diagnostic accuracy.

The process models uncovered variations in evidence generation and sharing of evidence and hypotheses. Students in the reflection condition enhanced their familiarity with the evidence by revisiting the health record, allowing them to justify their requests adequately. Compared to students in the collaboration scripts condition who repeatedly skipped parts, these students showed more efficient CDR processes. The reflection phase has potentially activated prior knowledge of disease patterns (Mamede et al., 2020). Consistent with prior research, however, these stringent CDR processes did not translate into improved diagnostic accuracy (Richters et al., 2022). The repeated generation of evidence in the reflection condition may indicate students' uncertainty due to an insufficient initial case representation. Learners need to apply a bottom-up reasoning approach to build up such case representations before reflection is externally activated, which may cognitively overwhelm novice diagnosticians. Thus, in line with previous research emphasizing the critical role of the initial case representation for diagnostic accuracy (Richters et al., 2022), we suggest additional prompts that stimulate knowledge integration and allow the generation of reasonable hypotheses. Further, potentially, students repeatedly generated evidence to clarify the relation between their initial hypotheses and the evidence - something diagnostic accuracy could not capture. Instead, through careful evidence and hypothesis sharing, the identified CDR process differences induced by reflection may have improved collaboration efficiency, which potentially represent a better measure.

The non-significant effects of both types of scaffolding could also be due to being non-adaptive, as prior research has demonstrated that effectiveness is diminished when support does not align with learners' prerequisites (Chernikova et al., 2020). Reflection phases, for example, were shown to be only effective for students with high prior knowledge (Richters et al., 2022). Future research should repeat the study and consider students' prior knowledge or other individual differences for more adaptive scaffolding.

The present study provides insights into the effects of collaboration scripts and reflection phases on CDR processes and diagnostic accuracy in a sample of medical students. Following the current scientific discourse, our findings challenge traditional outcome measures such as diagnostic accuracy and call for a nuanced approach to assessing the impact of scaffolding approaches on complex skills such as CDR.

## References

- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499–541.
- Ibiapina, C., Mamede, S., Moura, A., Elói-Santos, S., & van Gog, T. (2014). Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Medical Education, 48*(8), 796–805.
- Mamede, S., Hautz, W. E., Berendonk, C., Hautz, S. C., Sauter, T. C., Rotgans, J., Zwaan, L., & Schmidt, H. G. (2020). Think twice: Effects on diagnostic accuracy of returning to the case to reflect upon the initial diagnosis. *Academic Medicine, 95*(8), 1223–1229.
- Radkowsitch, A., Sailer, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2022). Diagnosing collaboratively: A theoretical model and a simulation-based learning environment. In F. Fischer & A. Opitz (Eds.) *Learning to diagnose with simulations: Teacher education and medical education* (pp. 123-141). Springer Nature.
- Richters, C., Stadler, M., Radkowsitch, A., Behrmann, F., Weidenbusch, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2022). Making the rich even richer? Interaction of structured reflection with prior knowledge in collaborative medical simulations. In A. Weinberger, W. Chen, D. Hernández-Leo, & B. Che (Eds.), *Proceedings of the 15th International Conference on CSCL* (pp. 155-162).
- Stadler, M., Hofer, S., & Greiff, S. (2020). First among equals: Log data indicates ability differences despite equal scores. *Computers in Human Behavior, 111*, 106442.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*(3), 477- 511.

## Acknowledgments

This research was funded by Deutsche Forschungsgemeinschaft (COSIMA, FOR 2385) grant for project 6 CoSiMed (FI 792/11-1).

# Can ChatGPT be Always Powerful: Evidence from a Postgraduate Writing Class

Chunlin Lei, Shanghai University of International Business and Economics, leichunlin@sui-be.edu.cn  
 Dan Zhu, Shanghai University of International Business and Economics, 932399208@qq.com  
 Junjie Zhang, the University of Hong Kong, zjjedu@connect.hku.hk

**Abstract:** This study compared textual features of English compositions produced by a group of Chinese postgraduates and ChatGPT respectively. Quantitative analysis was conducted to investigate the lexical and syntactic complexities, and qualitative analysis was used to examine errors and other textual features. The results showed ChatGPT performed better in creating syntactic variations and error-free texts while postgraduates did better in lexical density and rhetoric use.

## Introduction

ChatGPT has human-like text processing capabilities, and it is able to produce coherent, contextualized responses to a variety of cues, including questions, prompts, and textual supplements (Rudolph, 2022). Although the power of ChatGPT is well acknowledged, there are concerns about the way we use ChatGPT. Just as Neem (2023) points out that using a machine to help one to lift dumbbells does not mean that one's muscles will develop, and using a machine to write an essay does not mean that one's mind will develop. Therefore, it is important to understand that if ChatGPT is powerful in producing texts, in what way it performs better than human beings. Can ChatGPT be always powerful and surpass human beings in every aspect of writing? This study explores the differences in lexical and syntactic complexities between compositions produced by ChatGPT and a group of postgraduates. Besides, error and other textual analyses are conducted to unfold the benefits and possible pitfalls of ChatGPT in writing English compositions.

## Research design

This study was conducted in an *Advanced Business English Writing class* with 25 first-year postgraduates majored in Applied Linguistics. One of the writing tasks, cover-letter writing, was used as the dataset in this study. Specifically, this writing task asked the students to compose cover letters based on the instructions. Meanwhile, the same writing prompt was fed to ChatGPT. As ChatGPT can generate different texts each time when the same instruction is given, the process was repeated ten times at different intervals in this study. Eventually 25 pieces of human writings and 10 pieces of machine writings were obtained for analysis.

The textual analysis tool Coh-Metrix was used firstly to compare the lexical and syntactic differences in compositions generated by the postgraduates and ChatGPT. Conventionally, lexical complexity is mainly reflected by six indicators such as average number of syllables in a word, type-token ratio, etc.; while syntactic complexity is mainly reflected by five indicators such as average sentence length, syntactic similarity of neighboring sentences, etc. SPSS was used to examine the statistical differences between writings of humans and machine. Qualitative textual analysis was also conducted to identify the errors made by postgraduates or the machine; and other features such as pragmatic appropriateness or rhetoric devices employed in writing.

## Results

### Comparison of lexical complexity

Table 1 shows the lexical differences in compositions produced by postgraduates and ChatGPT. No significant differences on number of syllables or letters in a word, or word frequency were found. Surprisingly, postgraduates performed better than ChatGPT when Type-Token Ratio is considered, indicating student texts embodying higher vocabulary density and richer vocabulary use ( $p < 0.5$ ).

**Table 1**  
*Lexical complexity: independent samples t-test*

	F	Sig.	t	Sig. (2-tailed)	Mean Difference
Average number of syllables in a word	.0210	.8850	-.9120	.3680	-.0228

Average number of letters in a word	.9690	.3320	-1.1910	.2420	-.0840
Type-Token Ratio (content word)	2.3380	.1360	2.4340	.0210	.0289
Type-Token Ratio	1.4190	.2420	2.8320	.0080	.0359
Content Word Frequency	1.869	.181	.788	.4360	.0254
Word Frequency	.136	.714	.589	.5600	.0131

### Comparison of syntactic complexity

Table 2 shows the syntactic differences in compositions generated by students and ChatGPT. Significant differences were identified on most measures except for ease of reading. ChatGPT produced longer piece of writing and used longer sentences; and more diversified sentence patterns/structures were utilized as well.

**Table 2**  
*Syntactic complexity: independent sample t-test*

	F	Sig.	t	Sig. (2-tailed)	Mean Difference
Number of words	0.013	0.911	-3.292	0.002	-41.22
Average sentence length	2.9580	.0950	-2.1750	.0370	-1.7320
Syntactic similarity of neighboring sentences	5.6960	.0230	2.6790	.0110	.0428
Syntactic similarity of all sentences across paragraphs	4.299	.046	2.682	.0110	.0203
Ease of reading	.001	.972	1.691	.1000	3.7475

### Comparison of errors and other textual features

The authors of this study investigated the errors made by students and ChatGPT in compositions. Postgraduates as human beings, though writing with computers, still made spelling mistakes (N=11) such as “promte” (promote), “continue” (continue). In addition, typical errors such as incorrect word choices (N=45), wrong collocation use (N=38), non-idiomatic Chinglish expressions (N=26), and illogical sentence structures (N=34) were identified in the postgraduates’ writings. In comparison, ChatGPT made 0 spelling mistakes, very few errors on collocation, sentence structure, or logic (negligible or beyond the authors’ identification).

However, in this writing task, it was found that ChatGPT often created some “redundant information” (N=7) that is not supposed to be included in the cover letter if the pragmatic writing context is considered. Students made fewer mistakes in this respect (N=5). Furthermore, students sometimes used rhetoric devices (N=4) to make the writings more live and vivid, while ChatGPT did not generate texts with such a feature.

### Implications

This study shows the advantages of ChatGPT in English composition writing, particularly in generating syntactic complex and error-free texts. However, ChatGPT does not perform better than human in every respect. The study prompts us to think further in the design of foreign language writing class using ChatGPT as a useful tool.

### References

- Rudolph, J. (2022). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 5(1).
- Neem, J. (2023). 89% American college students use ChatGPT to writing their assignments. Available online Jan 30, 2023 <https://finance.sina.com.cn/stock/usstock/c/2023-01-30/doc-imycytui0280714.shtml>

### Acknowledgments

This paper is sponsored by the SUIBE-WES Institute of Holistic Education.

# Understanding Knowledge Building Processes Using Epistemic Network Analysis and Sequential Analysis

Xiaojie Zhang, The University of Hong Kong & Southern University of Science and Technology,  
zxjie@connect.hku.hk

Carol Chan, The University of Hong Kong, ckkchan@hku.hk

Shihui Feng, The University of Hong Kong, shihuiife@hku.hk

Jianhua Zhao, Southern University of Science and Technology, zhaojh@sustech.edu.cn

**Abstract:** This study investigated the knowledge building processes of 50 Grade-Five students in a collaborative inquiry about the history and evolution of Chinese characters. Epistemic network analysis and sequential analysis were employed to investigate how students contribute to the community knowledge advances within the *Initial Phase* and *Theory-Building Phase*. The temporal co-occurrence patterns and sequential patterns between questioning, explanation, and regulation processes were examined. This study sheds light on understanding KB processes from a temporal dynamic perspective.

## Introduction and previous studies

As a principle-based pedagogy in Computer-Supported Collaborative Learning (CSCL), Knowledge Building (KB) is proposed to facilitate idea improvement and productive knowledge development in collaborative communities, in which students propose authentic questions, exchange knowledge and ideas, construct and create community knowledge (Scardamalia & Bereiter, 2014). Some studies analyzed students' discourses crossed different inquiry phases or different groups using the coding and counting approach (van Aalst, 2009). However, the temporal patterns of KB discourse need deep investigation to understand how students contribute to the KB process (Chen et al., 2017; Hod et al., 2020). Epistemic network analysis (ENA) and sequential analysis (SA) have been revealed the potential to compare patterns of the KB processes and demonstrate sequential events' transitional probabilities in KB discourse (Chen et al., 2017; Hod et al., 2020). Therefore, this study aims to understand how students engage in KB processes using ENA and SA.

## Methods

50 Grade-Five students participated in an online collaborative inquiry into the history and evolution of Chinese characters. The collaborative inquiry was conducted using Knowledge Forum® (KF). The activity was designed containing two phases: 1) Phase I: *Initial Phase*, students published initial inquiry interest and generated preliminary ideas; and 2) Phase II: *Theory-Building Phase*, students engaged in deep group inquiry under sub-topics. Based on KB pedagogy principles, the two phases were adopted in most KB literatures. A total of 311 notes generated in the KB processes were analyzed. A coding scheme ( $\kappa = .865$ ,  $p < .001$ ) developed by van Aalst (2009, p. 274) was employed to characterize the notes' categories, such as *Seeking explanations*, *Proposing explanations*, and *Synthesizing diverse ideas*. The ENA web tool (<http://app.epistemicnetwork.org>) was employed to do epistemic network analysis (Shaffer et al., 2009). The units were "phase > subtopic > authors" and the time window was set as 4, as related discourses mostly co-occurred in four consecutive notes. For examining the transitional frequencies, SA was performed (Sackett, 1979). The codings of each subtopic's KB notes were extracted into independent sequences. Then, the sequences were analyzed to get a transitional frequencies table by using GSEQ 5.1 and then converted into transitional graphs by Gephi 0.9.7.

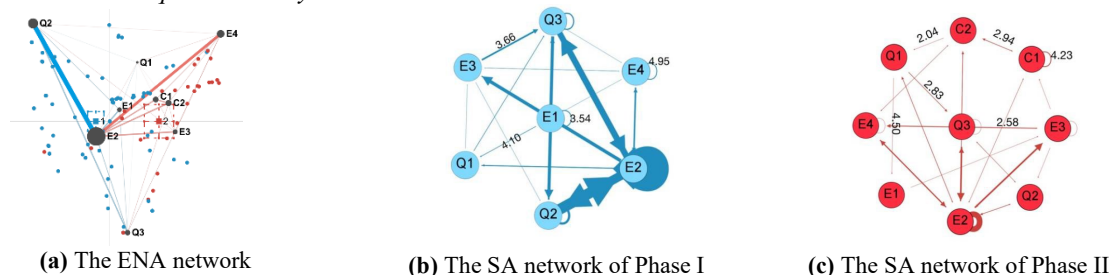
## Results

The Mann-Whitney test showed significant differences between the ENA networks of the two phases ( $Mdn = .00$ ,  $N = 141$ ;  $Mdn = 0.71$ ,  $N = 40$ ;  $U = 971.00$ ,  $p = 0.00$ ,  $r = 0.66$ ) (Figure 1(a)). In Phase I (the *Initial Phase*), the prominent ENA pattern was the co-occurrences of "*Questioning*" and "*Explanation*". Specifically, while "*Seeking explanations*" (e.g., "Will Chinese characters change in future?"), students "*Proposed explanations*" (e.g., "Chinese characters will become simpler") to generate more ideas and create an atmosphere of building community knowledge. The prominent ENA pattern observed in Phase II (the *Theory-Building Phase*) was the co-occurrences of "*Explanation*" and "*Community regulation*". Specifically, while "*Proposing explanations*" (e.g., "Chinese characters changed from circle to square"), students "*Supported explanations*" by providing evidence (e.g., "We can see that change from those pictures") and "*Improving existing explanations*" (e.g., "Square Chinese characters confirm the virtue of Chinese people's obedience to rules"). In addition, students regulated their collective inquiry process during the theory-building processes, such as "*Synthesizing diverse ideas*"



(e.g., “sum up previous ideas - Chinese characters have many expression forms”) and “*regulating*” (e.g., “student Z, please summarize above ideas; student G, please collect pictures”).

**Figure 1**  
*ENA and Sequential Analysis Networks*



**Note.** Phase I and Phase II was shown in blue and red, separately. *Questioning*: Q1: Seeking facts, Q2: Seeking explanations, Q3: Further questioning; *Explanation*: E1: Simple claim, E2: Proposing explanations, E3: Supporting explanations, E4: Improving existing explanations; *Community*: C1: Regulating, C2: Synthesizing diverse ideas.

Furthermore, SA demonstrated more in-depth the directed transition relations of the KF processes (Figure 1 (b) (c)). In Phase I (the *Initial Phase*), students continually “Proposed explanations” after “Proposing explanations” and tended to “Proposing explanations” after “Seeking explanations”. Students proposed “Further questioning” after “Proposing explanations” to the previous questions, and vice versa. In Phase II (the *Theory-Building Phase*), students build profound ideas by the transition from “Proposing explanations” to “Supporting explanations” and then to “Improving existing explanations”.

## Conclusion

The findings of this study provide insight into the co-occurrence and transition of the KB processes. Previous studies encouraged students to post questions in the *Initial Phase* and more community regulation in the *Theory-Building Phase*. This study goes further and suggests that pedagogy design could consider encouraging the alternation of “Questioning” and “Explanation” in the *Initial Phase* and the alternation of “Explanation” and “Community regulation” in the *Theory-Building Phase*. Additionally, this research reports different insights from ENA and SA. ENA analyzes the temporal co-occurrence patterns of KB processes and compares these patterns among different groups and individuals (Hod et al., 2020; Shaffer et al., 2009). On the other hand, SA uncovers the detailed transition of KB processes (Chen et al., 2017). The different insights from ENA and SA complement each other and can be used to study the temporal patterns of KB processes. This study suggests that ENA and SA can complement each other in uncovering the KB processes. Our future study will continue investigating the KB process pattern related to students' performance outcomes.

## References

- Chen, B., Resendes, M., Chai, C. S., & Hong, H.-Y. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, 25(2), 162–175. <https://doi.org/gc7r9q>
- Hod, Y., Katz, S., & Eagan, B. (2020). Refining qualitative ethnographies using Epistemic Network Analysis: A study of socioemotional learning dimensions in a Humanistic Knowledge Building Community. *Computers & Education*, 156, 103943. <https://doi.org/gg66d4>
- Sackett, G. P. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. *Handbook of infant development*, 1, 623–649.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 397–417). Cambridge University Press. <https://doi.org/d9s2>
- Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislavy, R. (2009). Epistemic network analysis: A prototype for 21st-Century assessment of learning. *International Journal of Learning and Media*, 1(2), 33–53. <https://doi.org/dw9h5r>
- van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 259–287.

## Acknowledgment

This study is supported by the National Social Science Foundation of China (Grant No. BCA200090).

# Toward a Learning Agreement for Generative AI Use: First-Year University Student Perspectives

Marc Beardsley, Universitat Pompeu Fabra, Barcelona, marc.beardsley@upf.edu  
Ishari Amarasinghe, Radboud University, Netherlands, ishari.amarasinghe@donders.ru.nl  
Davinia Hernández-Leo, Universitat Pompeu Fabra, Barcelona, davinia.hernandez-leo@upf.edu

**Abstract:** Tensions surrounding Generative Artificial Intelligence (GenAI) use in university learning calls for a process bringing teachers and students together to define acceptable uses of AI in education. This article introduces first year student opinions on GenAI use and learning agreements as a flexible, participatory tool to govern its use. Results show inequality in students' familiarity with GenAI, requests to be able to opt out of its use, and strong support for learning agreements as an instrument helping them learn to properly use AI and preventing them from being unfairly disadvantaged.

## Introduction

The proliferation of readily accessible Generative Artificial Intelligence (GenAI) tools capable of generating diverse content is disrupting university learning (Bozkurt et al., 2023). On one side, several concerns have emerged such as those related to academic dishonesty, privacy issues concerning user data, and an inherent bias in generated content. Moreover, the situation is constantly changing with frequent releases of new tools and improvements to existing tools. On the other side, there are calls stressing the need and opportunities to embrace GenAI in education (Markauskaite et al., 2022). Tensions between opportunities and concerns show the need to establish a process for bringing stakeholders such as teachers and students together to define acceptable uses of AI in education, to gather insights into how AI is being used, and assess whether such uses are effective or not – to model how such discussions can be held in a flexible manner that accounts for the dynamic situation. Learning agreements may offer a flexible, participatory tool for managing the use of AI by students. Learning agreements are negotiated agreements between students and teachers in which the process and outcomes of learning in a course are agreed upon (Creaner & Creaner-Glen, 2014). Through such agreements students are empowered to take ownership of their learning and “become aware of issues to reflect on before, during, and after activities” (Clear et al., 2016). This article presents results of a survey study that explored the following three research questions: R1. How familiar are first-year university students with GenAI? R2. Would first-year students be accepting of a course learning agreement on GenAI use?

## Methods

First year engineering students enrolled in a general studies course at a university in Spain consented to sharing their data for the study ( $f = 26$ ,  $m = 56$ ,  $other = 01$ ,  $prefer\ not\ to\ say = 01$ ,  $M\text{-age} = 18.17$ ,  $SD\text{-age} = 1.37$ ). Data was collected via an online survey (Google Form) given to students in a class taking place the first week of the first trimester. Following the completion of the survey, students were asked to review and complete a learning agreement form, attend a lecture introducing GenAI and Academic Integrity in which the terms of the form were discussed, and then, after the lecture, complete a final version of the learning agreement form. Responses were excluded from students who did not consent to sharing their data for research purposes ( $n = 11$ ).

## Results and discussion

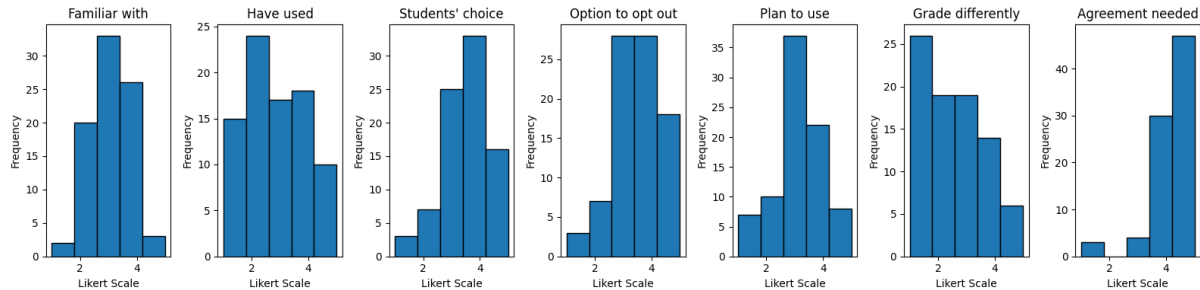
### R1. How familiar are first-year university students with GenAI

Figure 1 shows the count of responses per rated item. For the first survey item, *I am familiar with GenAI tools* ( $M = 3.095$ ,  $SD = 0.886$ ), results show that more students report being familiar with GenAI than not being familiar with it. For the second item, *I have used GenAI tools to assist with my schoolwork* ( $M = 2.809$ ,  $SD = 1.293$ ), more students report not having used GenAI than having used it. For the third item, *I think students should have the option to use GenAI to assist their learning in university courses* ( $M = 3.619$ ,  $SD = 1.004$ ), few agree that students should not have the option to use GenAI. For the fourth item, *I think students should have the option to opt out of using GenAI in class (i.e., professors should not require that we use it)* ( $M = 3.607$ ,  $SD = 1.029$ ), the vast majority feel students should be able to opt out of using GenAI. Ratings of the fifth item, *I plan to use GenAI to support my learning in university courses* ( $M = 3.166$ ,  $SD = 1.039$ ), suggest most students are uncertain on whether they will use GenAI for this course. For the sixth item, *I think that grading criteria should differ for students that use*

AI tools versus those that do not ( $M = 2.464$ ,  $SD = 1.284$ ) the majority do not believe grading criteria should differ between GenAI users and non-users. An analysis of the open-ended question, *Please list the AI tools you are familiar with*, reveals that students are most familiar with OpenAI tools (DaLL-E, ChatGPT) – 64.1% of tools mentioned and appearing in 94% (78/83) of student responses.

**Figure 1**

Results of the 7 Closed Questions (5-point Likert scale) on Generative AI



## R2. Would first-year students be accepting of a course learning agreement on GenAI use?

Results for the seventh item, *I think that each course should have a learning agreement related to the use of GenAI. A learning agreement outlines the ground rules for using GenAI in a course that students read and must agree to before using AI tools* ( $M = 4.404$ ,  $SD = .879$ ), show that an overwhelming majority are in favour of a course learning agreement for governing GenAI use. A thematic analysis of responses to an open question asking for an elaboration on their level of agreement rating, revealed that the most frequently expressed reasoning for having a learning agreement is to prevent being disadvantaged (62.3% of utterances). Many stated that rules, in general, should be applied (24.7%). Specific disadvantages listed included others gaining an unfair advantage from using AI in ways that other students are unaware of (ensure fairness, 13.9%), teachers penalising students for using AI in ways the teacher does not permit (prevent misunderstandings, 13.9%), and cheating with the help of AI (prevent abuse, 9.9%). The second most frequently expressed reasoning is to help students learn to use AI (23.8%) as one needs to learn how to use it properly (12.9%) and one needs to be aware of the potential, pros and cons of using AI as it is the future (10.9%). Only 4% state that they feel a learning agreement is not necessary.

All in all, results provide further evidence to the dynamic nature of the situation as differences were found across the level of familiarity and use of students with GenAI with many wanting to be able to opt out of using it. Also, with the overwhelming majority being in favour of a course learning agreement for governing GenAI use both to prevent being unfairly disadvantaged in their coursework and to help them learn to use AI properly, this work presents learning agreements as a flexible, participatory instrument with potential to bring stakeholders such as teachers and students together to define acceptable uses of AI in education.

## References

- Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S. & Bali, M. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, 18(1).
- Clear, T., McDermott, R., Parsjö, E., Cajander, Å., Daniels, M., & Lagerqvist, N. (2016, October). A framework for writing learning agreements. In *2016 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE.
- Creaner, M., & Creaner-Glen, J. (2014). Learning agreements: road maps to integrative learning. In *Integrative Learning* (pp. 105-117). Routledge.
- Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, 3, 100056.

## Acknowledgements

This work has been partially funded by UPF's PlaCLIK (E2023015714), Erasmus+ KA220-SCH-92001A07), the Government of Catalonia (SGR 00930), Spanish Ministry (PID2020- 112584RB-C33 granted by MICIN/AEI/10.13039/501100011033). DHL (Serra Húnter) acknowledges the support by ICREA (Academia).

## Collaborative Diagnostic Reasoning in a CSCL environment: A Multi Study Structural Equation Model

Laura Brandl, Department Psychology, LMU Munich, l.brandl@psy.lmu.de  
Matthias Stadler, LMU University Hospital, LMU Munich, matthias.stadler@med.uni-muenchen.de  
Constanze Richters, Department Psychology, LMU Munich, constanze.richters@psy.lmu.de  
Anika Radkowsitsch, IPN, radkowsitsch@leibniz-ipn.de  
Ralf Schmidmaier, LMU University Hospital, LMU Munich, ralf.schmidmaier@med.uni-muenchen.de  
Martin R. Fischer, LMU University Hospital, LMU Munich, martin.fischer@med.uni-muenchen.de  
Frank Fischer, Department Psychology, LMU Munich, frank.fischer@psy.lmu.de

**Abstract:** In Collaborative Diagnostic Reasoning (CDR), good diagnostic outcomes depend on high quality diagnostic activities influenced by social skills, content, and collaboration knowledge. Analyzing data from three studies on simulation-based learning (504 medical students) using a structural equation model, our results challenge the current CDR model. We suggest prioritizing collaboration knowledge over social skills, emphasize the reduced impact of content knowledge in simulations, and distinguish between information elicitation and sharing, with the latter being more transactive.

### Collaborative diagnostic reasoning

As in other collaborative settings, physicians need to collaborate with physicians from different fields of expertise to diagnose a patient's illness. The Collaborative Diagnostic Reasoning (CDR) model focuses on the diagnostic process in collaborative settings within knowledge-rich domains and describes the relationship between individual characteristics, the collaborative diagnostic process, and the diagnostic outcome: Content and collaboration knowledge as well as social skills are thought to influence collaborative diagnostic activities (Radkowsitsch et al., 2022). To construct a shared problem representation (Rochelle & Teasley, 1995) and achieve a high-quality diagnostic outcome, it is crucial to clearly communicate the type of evidence needed to reduce uncertainty (evidence elicitation) and to share relevant information (evidence sharing) and suspected diagnoses (hypothesis sharing) that may influence the collaborative diagnostic process. Thus, the CDR model emphasizes the importance of high-quality collaborative diagnostic activities influenced by content and collaboration knowledge as well as social skills to achieve accurate, justified, and efficient diagnostic outcomes.

### Current study and research question

The CDR model has only been partially tested, a comprehensive model test is still needed. The aim of this study is to test the CDR model in a simulation-based learning environment by analyzing data from three studies in the medical domain, *asking to what extent the relations in the CDR model are applicable across studies*. We expected the individual characteristics to be positively related to the quality of collaborative diagnostic activities, as well as the collaborative diagnostic activities to be positively related to diagnostic outcomes.

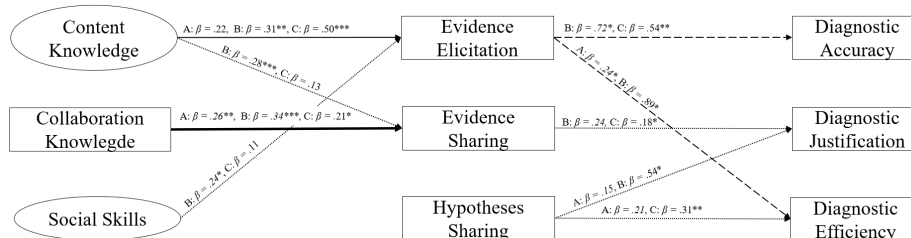
### Method and results

We conducted three studies involving medical students: Study A (n = 157) took place in a laboratory setting, Study B (n = 155) collected data online, and Study C (n = 192) was integrated into the curriculum. Despite the different data collection modes, the simulation was the same, so the same activities are required for all studies. Before entering the simulation students answered questions regarding their individual characteristics (i.e., social skills, content and collaboration knowledge;  $\omega = .28 - .83$ ). The students' task was then to diagnose an authentic (though fictitious) patient case in the role of an internist while collaborating with an agent-based radiologist. Developed and validated with medical experts, the case structure and content are closely aligned with the real-life clinical context. Collaborative diagnostic activities were measured using quality indicators constructed from log-file data storing each click in the simulation. The diagnostic outcome was operationalized as accordance of the given diagnosis with an expert solution (diagnostic accuracy); whether relevant information is used to justify it (diagnostic justification), and how much time the student needed to reach it (diagnostic efficiency). For a comprehensive model test, we estimated a structural equation model for each study using Mplus Editor. To conclude support for a hypothesis, path coefficients of two studies  $> 0.1$  are required.

The structural equation model had a good fit for every study ( $X^2(75) = 68.31 - 93.82$ , RMSEA = 0.00 – 0.04, CFI = 1.00). Looking at supported paths (see Figure 1) content knowledge is positively related to quality in evidence elicitation and evidence sharing, collaboration knowledge is positively related to evidence sharing, and

social skills are positively related to evidence elicitation. Which in turn is positively related to diagnostic accuracy and efficiency. Evidence sharing is positively related to diagnostic justification. Hypotheses sharing is positively related to diagnostic justification and efficiency. Further, we found no support for any indirect effects.

**Figure 1**  
*Relations between Individual Characteristics, Collaborative Diagnostic Activities, and Diagnostic Outcome*



*Note.* The full thick line represents a positive significant relation in three studies, the full line represents a positive relation in three studies of which two are significant, the dashed line represents a positive significant relation in two studies, and the dotted line represents a positive significant relation in two studies of which one is significant.  $*p < .05$ .  $**p < .01$ .  $***p < .001$ .

## Discussion

In conclusion, the relations in the CDR model (Radkowsch et al., 2022) are applicable across studies. However, the results challenge the CDR model in three ways. First, collaboration knowledge is crucial for effective evidence sharing, as indicated by the significant positive relation in all three studies—an often-lacking skill (Tschan et al., 2009). Therefore, it is imperative to give higher priority to collaboration knowledge compared to early-acquired social skills when revising the CDR model. Second, content knowledge seems to not play a crucial role in predicting diagnostic success, as indicated by the lack of indirect effects of content knowledge on diagnostic outcomes, compared to many other educational settings (Boshuizen et al., 2020). At least in simulation-based CDR, it is not enough to have sufficient knowledge, but it is also necessary to be able to engage in high-quality collaborative diagnostic activities in order to achieve high-quality outcomes, probably because of the opportunities to repeat, retry, and revise in simulation-based environments. Third, based on the relations we found for the different collaborative diagnostic activities, we propose to distinguish between broader categories of collaborative diagnostic activities: those that primarily focus on individual goals and needs (e.g., elicitation) and more transactive activities that directly target the collaborator (e.g., sharing). One advantage of the analyses is the high congruence of the used studies, which only differed in the mode of data collection. Thus, they were all in the same domain, so it remains an open question whether these findings are applicable across domains (e.g., teacher education). By considering all these aspects, we can take an important step in optimizing CDR and ultimately contributing to improved diagnostic outcomes in collaborative settings.

## References

- Boshuizen, H. P., Gruber, H., & Strasser, J. (2020). Knowledge restructuring through case processing: The key to generalise expertise development theory across domains? *Educational Research Review*, 29, 100310.
- Radkowsch, A., Sailer, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2022). Diagnosing collaboratively: A theoretical model and a simulation-based learning environment. In F. Fischer & A. Opitz (Eds.), *Springer Briefs in Education. Learning to diagnose with simulations - Examples from teacher education and medical education*. (pp. 116–134). Springer.
- Rochelle, J., & Teasley, S. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. O'Malley (Ed.), *Computer-Supported Collaborative Learning* (pp. 66–97). Springer.
- Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory: A Simulation Study of Group Medical Decision Making. *Small Group Research*, 40(3), 271–300.

## Acknowledgments

This research was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG, FOR 2385, Project 6 CoSiMed, FI 792/11-1 + 11-2).

# Design, Development, and Evaluation of an Immersive Virtual Reality Learning Environment (IVRLE) for Motor Learning in Fencing

Jonathan J. Foo, Wenli Chen, Wei Peng Teo  
jonathan.foo@nie.edu.sg, wenli.chen@nie.edu.sg, weipeng.teo@nie.edu.sg  
National Institute of Education, Nanyang Technological University

**Abstract:** The ability to engage and motivate and to support distance learning suggests the Immersive Virtual Reality Learning Environment (IVRLE) as an effective tool for the support and enhancement of collaborative learning. This study describes the theories, processes, and considerations in the design and development of an IVRLE with automated augmented feedback for novice-learners' motor learning of basic skills in fencing. The experimental design approach was used for the evaluation of the effectiveness of the IVRLE. The results show that the IVRLE improved the participants' motor learning of basic skills in fencing.

## Introduction

The ability to engage and motivate and to support distance learning suggests the Immersive Virtual Reality Learning Environment (IVRLE) as an effective tool for the support and enhancement of collaborative learning. This study describes the process of the design, development, and evaluation of an IVRLE for learning a series of motor skills in the context of the sport of fencing called the VR Fencing Coach Simulator. The research questions for this study are: RQ1: Does the performance of participants improve after they go through the motor learning in an IVRLE? RQ2: Are participants able to retain motor skills they have learnt within an IVRLE and be able to perform these motor skills in the real world? This study was submitted under the scope of Human Biomedical Research Framework (HBRA) and was approved by the Nanyang Technological University Institutional Review Board (NTU-IRB-2020-131).

## Immersive VR in sports education

The Immersive Virtual Reality Learning Environment (IVRLE) has significantly impacted the world of sports. As a result, modern athletes and coaches regard the knowledge gained from VR technological advances as invaluable. This belief may have some connection to the idea of feedback, which was first introduced in the field of mechanical control theory. The feedback provided through well-designed computer instruction can ascertain to the degree which the athlete's response was correct. As such, IVRLEs are well-suited for acquiring new skills.

## Design and development of the VR fencing coach simulator

Developed using the Unity game engine and distributed through the Steam VR platform, the VR simulator is built as an interactive game with six degrees of freedom (6DoF), in which participants learn to fence by controlling an avatar of themselves in the virtual environment. Both a front-end VR Head-mounted display (HMD) system and a back-end processor generating the virtual coach and collecting data from the participant's performance in the simulator have been incorporated into the system.

## Gameplay

In the gameplay, participants learn and practise basic fencing moves and stances. To acclimatise to the environment, they participate in a warm-up session where they poke at target objects flying towards them. After warming up, participants will be asked to select "Coaching" to start the proper coaching simulation session. Once in the coaching mode, they will have the choice of 5 fencing conditions. While in each condition, the participant will go through the instruction and practice of three fencing moves: En Garde, Direct Hit and Beat 4 Attack. Each move includes two demonstrations and three chances to practise the move and, in so doing, receive augmented feedback after each practice.

## Automated augmented feedback

Feedback is the element that allows the IVRLE to mimic the role of the coach. This IVRLE design positions augmented feedback within the VR simulator as a substitute for coaching sessions with a human coach. Auditory augmented feedback, also known as unimodal (Uni) feedback, will be provided to participants in the form of

computer-generated speech sounds programmed to invoke a response based on whether or not participants deliver a good performance. The animation of a mirror figure performing the move as a visual reference and reminder is how the visual component of the Auditory-Visual Augmented Feedback or multimodal (Multi) modes is conveyed to the participant.

## Research design

To answer RQ1 and RQ2, this study employed an experimental design to evaluate the effectiveness of IVRLE on the participants' motor performance. Paired-samples t-tests were used to calculate if there was a significance in the difference of quality of performance levels. A sample of fifty-eight (58) participants was recruited. 66% identified themselves as male and 34% as female, and they were between the ages of 22 and 39 ( $M = 29.98$ ,  $SD = 4.44$ ). The IVRLE learning task was a 150 mins session consisting of a warmup, three training conditions with three associated post-tests and a retention period of 30 mins followed by a retention test.

## Performance measurement

In order to measure the performance outcome of motor skills, this study utilised the Perception Neuron Pro system as an alternative to the traditional motion capture system to calibrate the participant's full-body inertial motion capture in real-time. The kinematic data of the captured joint angles are used to calculate the measurement of performance in Absolute Error (AE), which will be done employing Root-Mean-Squared Error (RMSE). As the motor skills deployed in this study are three fencing moves, and the measurement sought is the quality of performance, error measurement was deemed the most suitable. In addition, as these fencing movements are continuous skills, the RMSE score is the ideal measure for comparing the quality of performance of each participant's move to the gold-standard moves performed by an expert model in this study.

## Results

To answer RQ1, two paired-sample t-tests were conducted to compare the quality of performance between the participants' first no feedback learning condition and the second and third learning conditions. The results showed that there was a significant difference on the RMSE Quality of Performance levels between the first condition ( $M = 640.99$ ,  $SD = 133.17$ ) and the second and third learning conditions ( $M = 610.05$ ,  $SD = 131.17$ );  $t(57) = 2.284$ ,  $p = .013$ . where the effect size was small, with a Cohen's  $d$  of 0.300 and ( $M = 609.97$ ,  $SD = 124.55$ );  $t(57) = 2.637$ ,  $p = .005$ . where the effect size was moderate, with a Cohen's  $d$  of 0.346 respectively. The findings suggest that the participants delivered a better quality of performance after learning the fencing motor tasks in the IVRLE.

To answer RQ2, two paired-sample t-tests were conducted to examine the difference in Quality of Performance (RMSE) scores between the retention test and the second and third conditions. The results indicated that there was no statistically significant difference in the quality of performance between the retention test ( $M=619.36$ ,  $SD=127.28$ ) and the second and third learning conditions ( $M=610.14$ ,  $SD=131.17$ );  $t(57)=-1.071$ ,  $p = .144$  and ( $M=609.97$ ,  $SD=124.55$ );  $t(57)=-1.005$ ,  $p = .160$  respectively. Both results suggest that there is no significant difference between the quality of performance of the two learning conditions and that of the retention test. The evidence strongly supports the hypothesis that motor skills learnt in an IVRLE are retained even after a 30-minute break and can be transferred and performed in a real life setting.

## Limitations and challenges

Factors such as the particular VR system used, the task or skill learnt, the participant characteristics, and the instructional design elements can all influence the outcomes. Thus, caution should be exercised when extrapolating the results to broader populations or diverse VR learning scenarios. The study's duration and the participants' learning curve can influence the outcomes. Short-term interventions may not capture IVRLE's full potential or long-term effects. Participants' unfamiliarity with the VR system may also impact their performance and learning outcomes. Longer-term studies that account for the learning curve and prolonged exposure to the VR environment are necessary to understand the effects of IVRLE.

## Discussion and conclusion

This study carries theoretical significance by advancing our understanding of motor learning in IVRLE. The results of the study shows the effect of IVRLE in enhancing the motor learning outcomes. This research contributes to the design and development of virtual reality (VR) solutions and interventions. Even though this study was in the limited context of fencing, it will serve as a more significant reference for the further effective design and deployment of Virtual Reality (VR) and other related Extended Reality (XR) technologies in education and skills training.

# Evaluating Video Conferencing and Desktop Virtual Platforms for Supporting Remote Classroom Activities

Matin Yarmand, Ru Wang, Haowei Li, Nadir Weibel  
myarmand@ucsd.edu, ruw001@ucsd.edu, hal065@ucsd.edu, weibel@ucsd.edu  
UC San Diego

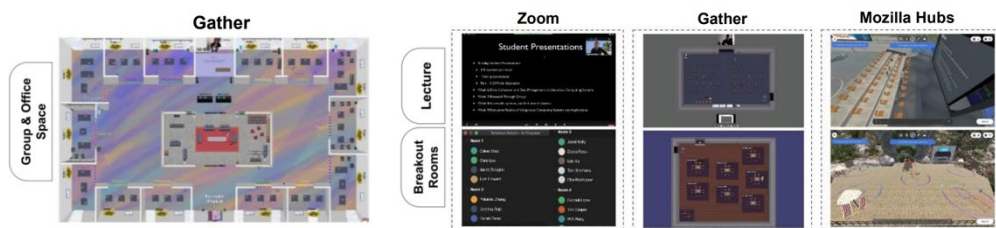
**Abstract:** Facilitating effective classroom interactions is especially challenging in online classrooms where participation is entirely remote and subject to tool features. This work explores three learning platforms during a six-week field study at a college class (N=30 students): video conferencing system, video chat with 2D map, and Desktop VR with 3D map. Post-study interviews and survey results revealed that students strongly preferred the plain video conferencing system for lectures due to superior student-content interactions. Meanwhile, they indicated a higher sense of co-presence and social awareness via top-down viewing in 2D map, shape-changing and moving avatars, and spatial audio. This study suggests tailoring features to optimize one type of interaction informed by characteristics of the learning activities.

## Introduction

When teaching university classes online, many instructors opt in for synchronous approaches due to real-time interaction. When facilitated online, lecture delivery relies on video conferencing platforms which (on the contrary) can offer limited support for other classroom interactions, such as pre/post lecture chats. In addition to video conferencing like Zoom (Yarmand et al., 2021), some works incorporated tools with better flexibility in forming interactions: for instance, Gather – a video call system with 2D map – can facilitate group learning activities more seamlessly (Fitria, 2021). Recent works also used Desktop-VR platforms (Fuller, 2021), that not only promote forming connections, but can also enhance learning experience via realistic 3D environments.

Facilitating classroom interactions is a significant challenge, especially when students' use of technology is discretionary: in video conferencing tools, learners avoid sharing videos which further exacerbates already limited interactions (Yarmand et al., 2021). While prior works explored online classrooms via short-term and single-platform interventions (e.g., Chen, 2003), we present results of a six-week field study with Zoom, Gather, and Hubs, as shown in Figure 1. We investigate how features of these platforms influence classroom interactions.

**Figure 1**  
*Common Space, Lecture, and Breakout Room Environments on Zoom, Gather, and Hubs*



## Methods

A six-week field study in an Engineering classroom investigated students' perceptions with different platforms. The classes were conducted using the same assigned platform and alternated weekly. The first 15 minutes of class were allocated to optional participation in Gather's spaces, where students could meet up with their peers or the instructor. The last 20 minutes of every week assigned students to discuss a particular topic in breakout rooms.

We collected two data sources. First, weekly surveys assessed perceptions on the assigned platform of the week, administered during the last three weeks. The survey measured eight dimensions of the psychological factors of education (Lee et al., 2010). Cronbach's alpha measured the internal consistency of the questions and determined high reliability scores (>0.7) which prompted Kruskal-Wallis and Mann-Whitney's U tests. Second, after six weeks, all students were invited to participate in semi-structured interviews, of whom four participated (2F, 2M; 21-26 years old). Interviews took 1.5 hours and consisted of three segments for each platform. To enhance recall, each segment took place on the corresponding platform and followed a lecture role-play.

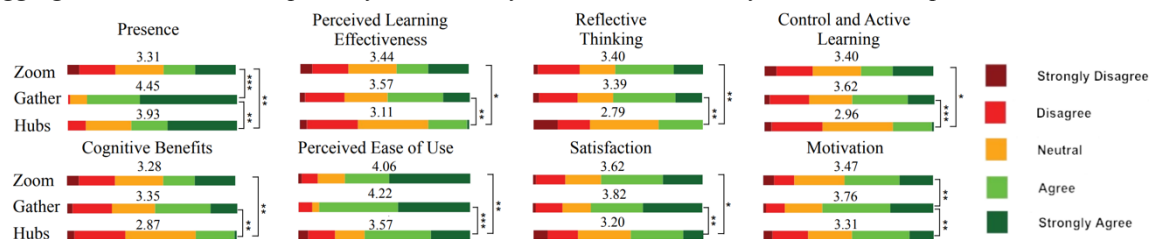
## Results



Figure 2 shows that Gather is generally the most preferred platform. The remaining describes the interview results.

**Figure 2**

*Aggregated Likert-scale Responses for Each Platform and Dimension of Educational Experience*



Note: extracted from prior surveys (Lee et al., 2010). Bars include mean scores, from 1 (lowest) to 5 (highest).

The participants mentioned that the Zoom interface suits lecture formats well due to “larger screen space for slides” (P2) and “high quality audio” (P3). When comparing Zoom to other modalities – P1 felt less distracted: “everything is black. There is just the video playing. I can more easily pay attention to this.” However, Hubs provided low quality visuals. P2 pointed out that the smaller size of slides made viewing the content more difficult. Other participants added that the contrast between the 2D screen and 3D space created a “twisted plane which had some rotation and affected the quality of the video when sitting in the corner” (P3).

Students benefited from the 2D map portion of Gather mainly due to the ability to move around and gain awareness over other students’ activities: “physical presence was dramatically improved, because there’s a consciousness behind movements. They’re sitting next to me, they *are* there. They did not just join the Zoom call and went to cook lunch” (P1). The 2D map also allowed students to tailor their class activities according to others, such as selectively joining popular discussion groups with more students. Hubs lacked a designated video feature – unlike both Zoom and Gather – which diminished satisfaction in small-group interactions: “It’s basically a phone call and, then I see this little avatar’s head getting bigger and smaller as he’s talking” (P1). The participants also strongly preferred the spatial audio of Hubs which allowed students to “listen to what the professor says to the other teams” (P4) and “help get to know more people in class by listening in” (P2).

## Discussion and conclusion

This study highlights a clear distinction between characteristics of lectures and group interactions which might require different tool features. Aiming to optimize all types of interactions simultaneously can not only introduce unwanted complexity and diminish usability, but also it can degrade the overall educational experience. Designing an ideal platform for remote classrooms requires considering existing trade-offs, aligned with Anderson’s Equivalency Theorem (Anderson, 2003) which states: *meaningful learning is supported as long as one of the three forms of interaction [student-student, student-instructor, and student-content] is at a high level. The other two may be offered at minimal levels, or even eliminated, without degrading the educational experience.*

This study explores classroom interactions by conducting a six-week field study at a university class. Students experimented with lecture and breakout room spaces in three platforms: Zoom, Gather, and Hubs. Interview results and surveys revealed students’ strong preference for attending lectures on Zoom, and higher co-presence and social awareness in Gather and Hubs due to top-down viewing of peers’ movements and spatial audio. This work lastly suggests tailoring tool features to unique classroom activities.

## References

- Anderson, T. (2003). Getting the mix right again: an updated and theoretical rationale for interaction. *The International Review of Research in Open and Distributed Learning*, 4(2).
- Chen, M. (2003). Visualizing the pulse of a classroom. *Proceedings of the Eleventh ACM MM*, 555–561.
- Fitria, T. N. (2021). Creating Sensation of Learning in Classroom: Using “Gather Town” Platform Video Game-Style for Virtual Classroom. *Education and Human Development Journal*, 6(2), 30–43.
- Fuller, E. G. (2021, July). WIP: Mozilla hubs classes fight feelings of isolation and online fatigue. In 2021 ASEE Virtual Annual Conference Content Access.
- Lee, E. A., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, 55(4), 1424–1442.
- McAuley, E., Duncan, T., & Tammen, V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport: A confirmatory factor analysis. *Research for exercise and sport*, 60(1), 48–58.
- Yarmand, M., Solyst, J., Klemmer, S., & Weibel, N. (2021). “It feels like I am talking into a void”: Understanding interaction gaps in synchronous online classrooms. The 2021 CHI Conference.

## Youth Perspectives on the Roles and Risks of AI in Their Classrooms

Megan Humburg, Indiana University, mahumbur@indiana.edu  
Dalila Dragnić-Cindrić, Digital Promise, daliladc@digitalpromise.org  
Cindy Hmelo-Silver, Indiana University, chmelosi@indiana.edu  
Krista Glazewski, North Carolina State University, kdglazew@ncsu.edu  
James Lester, North Carolina State University, lester@ncsu.edu

**Abstract:** This study uses thematic analysis of focus groups to highlight how middle schoolers discuss the benefits and drawbacks of educational AI. Students reflected on AI's benefits for learning, while also addressing the unsettling risks of being surveilled. Students also noted how AI falls short compared to human counterparts. Overall, we argue that middle schoolers can articulate complex understandings of AI, and their voices should be central in AIED design.

### Introduction and literature review

As the development of AI technologies for education continues at a rapid pace (Prahani et al., 2022), it is vital for researchers, educators, and students to be aware of the varied benefits and risks of AI tools and the forms of learning that these innovations seek to promote in classrooms. Previous studies of youth perspectives on AI highlight that while students notice the presence of AI in different aspects of their lives, they do not always understand how these technologies function (Greenwald et al., 2021). Researchers have documented how commercial AI software is plagued by issues of algorithmic bias and discrimination along gendered and racialized lines (e.g., Buolamwini & Gebru, 2018), and youth are increasingly aware of the negative impacts that biased technologies can have on their lives, even when they lack the formal vocabulary to describe it (Coenraad, 2022). If we want to ensure more just and ethical AI-driven educational technologies, students' voices must be centered in the design process to help shape AI technologies that impact their classrooms and lives (Hasse et al., 2019). Towards this end, we conducted focus groups with youth interested in AI technologies to explore the question: How do middle school students discuss the roles, risks, and benefits of AI technologies for their classrooms?

### Methods

Data were drawn from two different focus groups conducted with middle school aged students (n=15 students, ages 11-14) in the United States. The focus group structure involved students playing an educational game demo that uses AI-driven embodied conversational agents to give students tailored feedback. Afterwards, students participated in researcher-facilitated, semi-structured discussions. Researchers asked questions to jumpstart discussion (e.g., "If you could design an AI helper for your classroom, what would you want it to do?"), but conversations were ultimately student-led. Three hours of audio data were transcribed, and thematic analysis (Braun & Clarke, 2012) was used to draw together ideas from both focus groups into categories of meaning that reflected the various student-articulated benefits, risks, and roles related to AI classroom integration.

### Results

Overall, four key themes characterized students' conversations: 1) AI as a provider of engaging learning activities, 2) AI as a tool for surveillance and control, 3) AI as an ineffective replacement for human interaction, and 4) AI as a tool for adaptation and support. When asked how they would design AI for learning, students in both focus groups returned repeatedly to the idea that a well-designed AI agent would encourage their emotional engagement. Multiple students mentioned wanting AI to make learning "more fun" and encourage active participation. This interest in engaging activities led Caleb to propose "make all teachers robots [...] but they have a terrible code that you can hack." This proposal was met with mixed responses from peers; Arun, agreed that a hackable robot teacher "would make the kids learn and would make it more fun" because the activity could be "like an escape room" where students could practice their coding skills. The thought experiment around "should we make all teachers robots?" led students to return repeatedly to the core goal of their robot teacher design – a desire for agency over their learning experiences. Whether or not an AI educator could fulfill the goal of making learning more active, fun, and engaging, students clearly felt that advances in AI technology offered them possibilities to redesign their school experiences to better reflect their own ideals for learning activities.

A central concern students raised was that AI tools cannot always be trusted to keep the information they process private. Students noted that the power of AI could be "kind of terrifying" and that it was important to obtain permission to use people's art, voice recordings, and other data. Sara summarized the group's privacy

concerns by saying, “If [a student is] talking to the robot teacher, the robot teacher might as well just be listening or report to the government on what's happening. And that might be like the person's personal information. So then I think that would lead to the kids feeling like they can't really talk to very many people about what's going on.” In this way, students’ discussions mirrored the broader conversations currently taking place in the public sphere about data security, data ownership, privacy, and trust in the design of AI tools. While students saw power and potential in the ability to design AI tools for learning, they also saw risks in allowing AI-driven agents to have access to their data, especially when they were unsure how their information would be used. Another layer of students’ concerns centered on the inability of AI technologies to adequately mimic human qualities such as emotionality, social support, and intelligence. David argued that the AI chatbot in the demo game, which was designed to answer students’ science questions, was not actually intelligent because the AI tool did not offer more information than a human with a search engine. Ryan noted that “humans are more comfortable with humans”, so AI agents might not be as effective for supporting learning without that sense of social and emotional support. All of these comments suggest that students see clear distinctions between the tasks that AI tools can effectively support, and the more complex parts of teaching that require intellectual and socioemotional skills.

Finally, despite their reservations in terms of how AI can be misused or overused in the classroom, students in both groups noted that AI technologies have the potential to offer useful differentiation for a variety of learners based on their particular interests, skills, and prior knowledge. For example, Mara explained that when playing the game, “if you’re really really knowledgeable in those topics, you would want something more advanced to challenge you.” This focus on tailoring students’ learning experiences ties back to the overarching design goal that students articulated throughout their discussions, which was to generate learning experiences that were active, agentic, enjoyable, and engaging for each individual student.

## Discussion

While students’ designs pushed ethical and technological boundaries, at the core of these conversations was a desire for control over their learning experiences. These results suggest that we should not underestimate the complexity of students’ emerging understandings of AI technologies, even when they are still coming to understand how AI functions. Students were able to hold ethical, economic, socioemotional, and educational concerns in tension with one another as they workshoped design ideas together and navigated what the role of AI should be in their classrooms. Centering complex ethical dilemmas in discussions can help youth develop deeper understandings of AI as they express their concerns and their hopes for how these technologies will impact their lives (Lee et al., 2022). Working with students to articulate together what values and risks AI brings to their classrooms can help them to envision new possible futures and the technologies that these futures require (Rasa & Laherto, 2022). Centering students’ voices in the design of AIED technologies offers them agency to imagine and design towards alternative futures where all learning is active and meaningful for their lives.

## References

- Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.
- Coenraad, M. (2022). “That’s what techquity is”: youth perceptions of technological and algorithmic bias. *Information and Learning Sciences*, 123(7/8), 500-525.
- Greenwald, E., Leitner, M., & Wang, N. (2021, May). Learning artificial intelligence: insights into how youth encounter and build understanding of AI concepts. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 17, pp. 15526-15533).
- Hasse, A., Cortesi, S., Lombana-Bermudez, A., & Gasser, U. (2019). *Youth and artificial intelligence: Where we stand*. Berkman Klein Center Research Publication.
- Lee, C. H., Gobir, N., Gurn, A., & Soep, E. (2022). In the black mirror: Youth investigations into artificial intelligence. *ACM Transactions on Computing Education*, 22(3), 25.
- Prahani, B. K., Rizki, I. A., Jatmiko, B., Suprpto, N., & Amelia, T. (2022). Artificial Intelligence in Education Research During the Last Ten Years: A Review and Bibliometric Study. *International Journal of Emerging Technologies in Learning*, 17(8).
- Rasa, T., & Laherto, A. (2022). Young people’s technological images of the future: Implications for science and technology education. *European Journal of Futures Research*, 10(4).

## Acknowledgments

This work is funded by the National Science Foundation as a part of the EngageAI Institute (DRL-2112635). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## Reflecting on Prior Collaborations to Prepare for Group Work

Joshua Katz, University of Illinois Urbana-Champaign, jekatz2@illinois.edu  
Molly Goldstein, University of Illinois Urbana-Champaign, mhg3@illinois.edu  
Emma Mercier, University of Illinois Urbana-Champaign, mercier@illinois.edu

**Abstract:** We report on a study of the role of reflection on prior experiences on the quality of group contracts and differences in outcomes between groups who reflected before creating their contract and groups who did not receive a prompt to reflect. Initial results indicate that group contracts were longer and more specific if individuals had not. While there was little difference in project quality, teams were less satisfied with their collaboration in the reflection condition.

### Introduction

Collaborative learning is a powerful tool and yet students often report problems and experience high levels of interaction difficulties, which can impact their learning outcomes (Takai & Esterman, 2019). Research highlights the importance of preparing students to engage in group activities, yet instructors often lack the skills or knowledge to conduct these interventions (Mercier et al., 2023). Reflective practices are also vital to collaborative learning (Aldmir & Borge, 2020). We report on a pilot study to exploring how small interventions, which are simple to implement and do not consume large amounts of class time, can support collaborative groups.

One strategy used to guide groups is creating group contracts (Brannen et al., 2021). The best way to create contracts is still to be determined. In these contracts, students prepare and follow the agreed-upon guidelines to complete the project successfully. Prior work on the role of group contracts indicate that they can be supportive in improving the group experiences and project outcomes for groups, however, without guidance, students can create excessively simple contracts that do provide any benefit to their processes and outcomes (Cartwright et al., 2020). We explored whether giving individuals 5 minutes to reflect on prior group experiences before creating a contract led to different contracts and outcomes than groups who did not have time to reflect.

### Research questions

1. Do groups who reflect on their prior collaborative experiences create qualitatively different contracts than groups who do not reflect on their prior experiences?
2. Do groups who reflect prior to creating a contract have different satisfaction outcomes from groups who do not reflect on their contract?

### Methods

#### Design and participants

A mixed methods design was used for this study. Sections of the courses were randomly placed in one of two conditions – students in three sections were told to reflect on their prior experiences before creating a group contract, while students in the other three sections were just told to create their contracts.

A range of data was collected during the semester. For this proposal, we describe the team contracts (N = 24; 12 per condition) and final reflections (39 in no reflection; 32 in reflection condition). During the second week of the course, the students were introduced to their final projects and teams. They were given options of which building on campus they were to construct during the final project.

The research took place in an introductory engineering design and building information modeling course at a large Midwestern University in the United States in the Spring of 2023. The class is structured with a full-class lecture of 107 students and six individual lab sections of up to 20 students. There were 107 (80 consenting) undergraduate students in the class (65.4% male, 34.6% Female). In the team design project, teams of four to five students worked with blueprints of a campus building and created a digital twin of the building in Revit.

#### Coding and analysis

The primary author developed the coding schemes, in consultation with the other authors. An emergent inductive coding scheme was created for the coding of contract quality by the authors. A second emergent inductive coding scheme was created for the response to the question of how satisfied students were with their team. For this paper, we report the overall satisfaction reported (from not satisfied to very satisfied) but not the qualitative differences.

## Results

There was no statistical difference in the average word count for contracts (no-reflection  $M = 49.17$ ;  $SD = 22.53$  in reflection condition:  $M = 35.83$ ;  $SD = 13.51$ ;  $t(22) = 1.76$ ,  $p = .09$ ). In the reflection condition, there were more references to procrastination, communication, and specific modes of communication, as well as helping each other. In the no reflection condition, there were more references to meeting times, and specific meetings times, as well as deadlines and distribution of work (See Table 1).

The final reflections were coded for a number of themes. The simplest was a four-point scale from not satisfied to very satisfied with their group and the distribution of labor through an online survey. Similar percentages of students across conditions were satisfied or very satisfied (54% in the reflection and 53% in the no reflection condition). More students in the reflection condition were not satisfied with their group (11 of 33; 33%) when compared to those in the no reflection condition (1/39; 2%).

**Table 1**  
*Number of Themes Identified between Cconditions*

Theme	Reflection (N = 12)	No Reflection (N = 12)	Theme	Reflection (N = 12)	No Reflection (N = 12)
Procrastination	4	2	Helping	3	1
Meeting times (not specific)	6	9	Conflict management	4	8
Meet time/place specified	1	4	Deadlines	7	10
Communication	11	7	Distribution of work	6	9
Communication medium specified	7	4			

## Discussion

While just a small amount of the data has been reported in this paper, it appears that there are differences between the two conditions. Participants who reflected on their prior experiences created shorter contracts than those who did not reflect, and overall, focused on various aspects of collaboration. Groups who reflected were more likely to discuss the mode of communication as well as making a commitment to helping each other, while groups who did not reflect were more likely to specify where and when they would meet and how they would deal with conflict. These differences may be because reflections led students to identify areas of failure in the past (e.g., form of communication) or it may have reduced their perception of a need to discuss a range of issues with their team, leading to shorter contracts. Further analysis of the initial reflections will shed light on these hypotheses.

Groups who reflected were less satisfied with their group processes or distribution of work than groups who had not reflected. It may be that the intervention was too short (5 minutes) and too distant to have any impact on this, but further analysis of the intermediate reflections will be used to explore these issues in more detail. While this study set out to explore whether simple interventions could help groups in their collaborative processes, the data so far indicates more nuanced interventions may be necessary to truly support collaboration.

## References

- Aldemir, T., & Borge, M. (2020). Unpacking collaborative sense-making: The role of reflective accuracy in collaborative process quality. *In Proceedings of the International Society of the Learning Sciences (ISLS) Annual Meeting* (pp. 1-4). Online. Retrieved from <https://repository.isls.org/handle/1/6405>.
- Mercier, E., Goldstein, M. H., Badigar, P., & Rajarathinam, R. J. (2023). Collaborative learning in engineering education. *In International handbook of engineering education research* (pp. 402-432). Routledge.
- Brannen, S. F., Beauchamp, D., Cartwright, N., Liddle, D., Tishinsky, J., Newton, G., & Monk, J. (2021). Effectiveness of Group Work Contracts to Facilitate Collaborative Group Learning and Reduce Anxiety in Traditional Face-to-Face Lecture and Online Distance Education Course Formats. *International Journal for the Scholarship of Teaching and Learning*, 15(2).
- Cartwright, N. M., Patil, P., Liddle, D. M., Newton, G., & Monk, J. M. (2021). Enhancement of Professional Behaviours and Perceptions of Critical Skill Job Preparedness through the Use of a Group Work Contract in Fourth-Year Nutritional Science Students. *International Journal of Higher Education*, 10(2), 27-41.
- Takai, S., & Esterman, M. (2019). A review of team effectiveness models and possible instruments for measuring design-team inputs, processes, and outputs. *The International journal of engineering education*, 35(6), 1684-1697.

## Evaluating a New Knowledge Building Analytics Tool (KBAT)

Calixto Gutiérrez-Braojos, calixtogb@ugr.es, University of Granada, Spain  
Carlos Rodríguez-Domínguez, carlosrodriguez@ugr.es, University of Granada, Spain  
Daniela Linda, linda.daniela@lu.lv, University of Latvia  
Paula Rodríguez-Chirino, paularodrich@correo.ugr.es, University of Granada, Spain

**Abstract:** The study explores the appraisal students make of the visualization tool Knowledge Building Analytics Tool (KBAT) to facilitate reflective self-assessments in Knowledge Building communities supported by the Knowledge Forum. This study carries out a design-based research methodology. Data collection involved a survey evaluating students' perceptions of KBAT's value. The study involved 122 students from an educational research course at the University of Granada, Spain. The results show that KBAT is highly valued by students.

### Introduction

Learning analytics is an emerging field with tremendous promise in enhancing educational processes. One of the educational theories that is addressing the challenge of developing Analytics tools is Knowledge Building, KB (Zhu & Kim, 2017). This educational theory aims to empower students to collaborate in improving ideas (Scardamalia & Bereiter, 2021). The KB pedagogy proposes that teachers should gradually entrust students with responsibilities linked to knowledge construction. The teacher should create an educational environment that encourages students to share, question, and refine ideas about authentic problems. Students need to engage in knowledge improvement, which involves reflecting on the extent to which ideas are progressing and which issues should be addressed to continue deepening their understanding. Many recent studies affirm that conducting concurrent and reflective assessment sessions helps coordinate individual efforts to enhance shared knowledge (e.g., Yang et al., 2022). KB is often implemented in educational hybrid environment supported by technology named KF. This educational platform offers features that help students collaborate and enhance their knowledge, such as knowledge building scaffolds, (...), ideas promising tool. Over the past decade, researchers and educators have worked together to develop associated analytics tool and visualization for the Knowledge Forum. The progress made in the field of KB analytics tools is praiseworthy for its the innovative value it brings to the field of education. KB Analytics are still at a stage of development (see Zhu & Kim, 2017).

This study builds upon funded research with the purpose of constructing and evaluate the Knowledge Building Analytics Tool (KBAT). *Our objective is to explore students' perceptions of the practical use of the KBAT and discuss how these perceptions can guide the improvement of future versions of the software.*

### Method

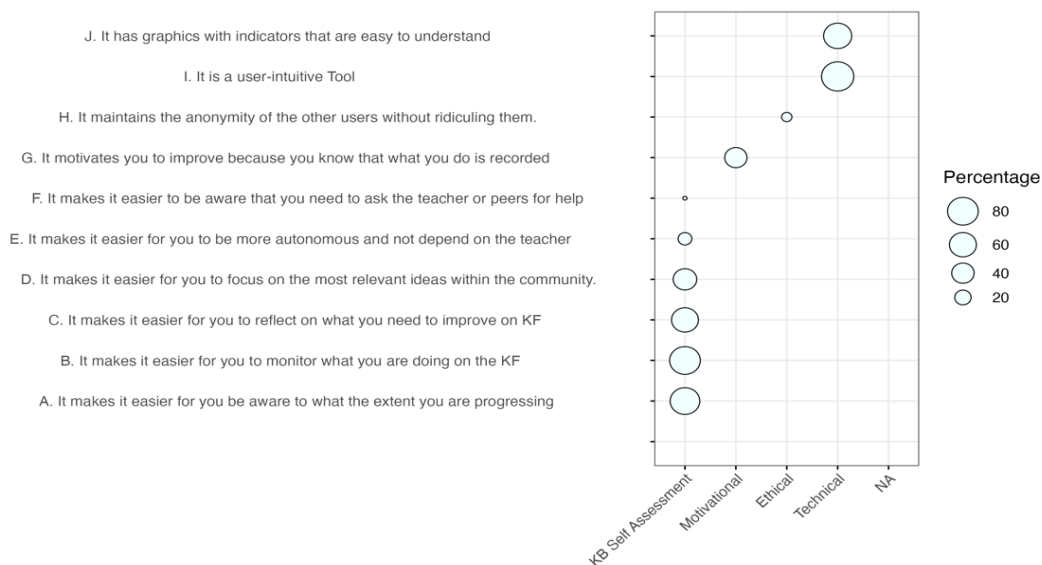
This study carries out a design-based research methodology. Participants were 122 (87.70% female) enrolled in an education program at the University of Granada, Spain. The KB pedagogy was employed during the course in a blended mode supported by KF. Over span of 16 weeks, students engaged in knowledge building discourse to advancing collective understanding around action-research topics. A sequence conformed by modules was carried out to implement KB in the classroom. In addition, students' self-evaluation was supported by Knowledge Building Analytics Tools (KBAT) in weeks 6, 12, and 16. KBAT provides a dashboard that is divided into 12 subsections, explicitly designed to facilitate members' reflection on their own activity in the KF. These subsections consist of analyses, visualizations, and indicators to support reflective consideration of participation in the KF, reading of other notes, equidistribution of engagement among members, member relationships, discursive movements, discourse leadership, conceptual progress over time, promising ideas, and more. Using the information and visualizations provided by KBAT, students described their forum participation, conducted self-assessments, and suggested improvements for more significant contributions in line with KB principles for future participation in the KF.

A 27-item Likert scale, organized into 4 dimensions (self-assessment, motivational, ethical, and technological, based on Authors, 2023), was applied to collect data about students' perceptions of the value they assign to KBAT in obtaining information about their knowledge-building process. In addition, 3 open-ended questions were posed to gather specific suggestions for potential enhancements of KBAT. The data obtained from the Likert scale was analyzed using the RStudio software generating graphs with descriptive information, while the data collected from the open survey was co-coded between the authors (reaching a full agreement).

## Results

The Likert responses to the questionnaire revealed a consensus among the students. They provided positive evaluations of the dashboard across its various dimensions: self-assessment ( $X=4.08$ ;  $SD=.50$ ), motivational ( $X=4.17$ ;  $SD=.56$ ); ethical ( $X=4.63$ ;  $SD=.39$ ); technical ( $X=4.58$ ;  $SD=.43$ ). The open-ended questions confirm the positive results obtained with the Likert scale and provide reasons for the value of KBAT (Figure 1). Although all students stated that using KBAT for reflecting on collaboration and the progress of collective ideas is positive, 18.85% of the participants did not justify their response or providing very brief with valueless answers. The remaining responses (81.85% of students) were jointly categorized by two of the researchers, reaching complete agreement. In fact, only 4.91% of students mentioned any negative appraisal regarding KBAT. Unfavorable assessments were focused on the quantity of graphics and motivational aspects. A few participants expressed concern about the continuous monitoring of their performance, fearing that this would negatively impact their final grade.

**Figure 1**  
*KBAT Value according to Students (Opened-Questions)*



## Conclusions

KBAT is a tool that assists students in achieving their Knowledge Building goals. KBAT allows students to monitor, among other things, the consistency of their engagement, discern patterns in their discourse, and track their conceptual development over time, although further improvements are still possible. During the ISLS-conference, colleagues can use KBAT on a tablet and will be invited to provide feedback for improvements.

## References

- Scardamalia, M., & Bereiter, C. (2021). Knowledge Building: Advancing the State of Community Knowledge. In U. Cress, C. Rosé, A.F. Wise, & J. Oshima (eds). *International Handbook of Computer-Supported Collaborative Learning. Computer-Supported Collaborative Learning Series, 19*. Springer, Cham.
- Yang, Y., Yuan, K., Feng, X., Li, X., & van Aalst, J. (2022). Fostering low-achieving students' productive disciplinary engagement through knowledge-building inquiry and reflective assessment. *British Journal of Educational Technology, 53*(6), 1511-1529. <https://doi.org/10.1111/bjet.13267>
- Zhu, G., & Kim, M. S. (2017). A review of assessment tools of knowledge building: Towards the norm of embedded and transformative assessment. *Paper presented in Knowledge Building Summer Institute, 18–22 June 2017, Philadelphia, PA.*

## Acknowledgments

Project (PID 2020-116872-RA-100), financed by: Spanish-Ministry of Science and Innovation-State Research Agency.

# Cultivating Epistemic Doubt: a Key Competence for Productive Participation in the Era of Artificial Intelligence

Sini Davies, Kati Sormunen, Kaiju Kangas  
sini.davies@helsinki.fi, kati.sormunen@helsinki.fi, kaiju.kangas@helsinki.fi  
University of Helsinki

**Abstract:** The rapid development of increasingly sophisticated AI systems challenges all fields of education. In this paper, we introduce and emphasize the importance of cultivating young people's epistemic doubt capabilities in order to be able to productively participate and work in the emergent era of AI – the era of black boxes and solving agency without intelligence.

## Introduction

Artificial intelligence (AI) is quickly rushing into our lives through the rapid development of increasingly sophisticated AI systems that rely on complex neural networks, also known as deep learning, that can be loosely compared to the human brain. Similarly, to the human brain, it is impossible to understand from the outside, how it functions exactly and how it derives the outcomes it produces (Voosen, 2017). More often than not, AI is ambiguous and opaque – it is a black box. On the other hand, users tend to perceive AI systems, particularly generative ones as human-like entities, and manufacturers seek to enhance this perception to promote higher acceptance and interaction with them (e.g., Hu et al., 2021). Driven by massive amounts of data, AI can learn at a scale and speed that people are unable to achieve and produce outcomes that are very hard to distinguish from those created by humans. However, AI does not understand or know anything (Floridi, 2023). AI systems are prone to produce biased and discriminating outcomes, struggle with basic logical reasoning, and make up things and present them as facts (Lambert & Stevens, 2023). Within this context, AI can be defined as a new form of agency – solving agency – that, unlike any other agency we are accustomed to, has the ability to ‘learn,’ change its behavior, improve itself, and produce human-like outcomes, without being intelligent nor a human (Floridi, 2023).

We align ourselves with many other researchers that, to be able to live and participate productively in the era of AI people crucially need personal and shared competencies of knowledge creation as well as epistemic fluency (Markauskaite et al., 2022; Paavola & Hakkarainen, 2014; World Economic Forum, 2023). However, we argue that in order to achieve this when working with, around, and sometimes despite AI – black boxes and solving agencies without intelligence – young people also need to cultivate sophisticated competencies of epistemic doubt.

## Epistemic doubt: a key capability for productive participation in the AI era

Epistemic doubt has its roots in the theories of doubt in philosophy (see e.g. Peirce, 1877). We define epistemic doubt, in accordance with Bearman and Ajjawi (2023), as both a cognitive and affective state of discomfort and uncertainty where there is a profound understanding that information may be partial, biased, or incorrect. The ability to cultivate epistemic doubt includes two key stages: 1) *Interactions with AI, or with other forms of agency, knowledge, or artifacts, can be held in ‘epistemic doubt,’* and 2) *While holding epistemic doubt, information can be taken ‘in trust’ simultaneously.* When cultivating epistemic doubt, AI systems, AI-generated judgements, content, and ideas emerging through them become considered through three philosophical perspectives: a) ontology (What is there?), b) epistemology (How do we know?) and c) axiology (Should we care?). Thus, epistemic doubt is skillful balancing and moving between trust and distrust (see Bearman & Ajjawi, 2023). Further, we consider wonder and questioning and moving from those to deeper understanding and creation of knowledge as the key characteristics of epistemic doubt, in accordance with previous research on the role of epistemic doubt in collaborative philosophical inquiry (Burgh et al., 2018).

Critical thinking can be seen as closely related to epistemic doubt. However, critical thinking tends to focus on the capabilities of weighing strengths and weaknesses of a proposed idea or knowledge and aims to solve the question of whether something is true or not by evaluating its validity and reliability through inductive and deductive reasoning (e.g., Pithers & Soden, 2000). When working with AI ‘black boxes,’ it is impossible to rely on such evaluation. Previous research has emphasized the importance of evaluating and understanding ‘inputs’ and ‘outputs’ as a way to understand and work with AI systems (Markauskaite et al., 2022; Zednik, 2021). However, as Bearman and Ajjawi (2023) point out, the inputs and outputs in AI systems are the same thing because outputs become inputs in successive iterations of machine learning. Furthermore, epistemic doubt differs significantly from epistemic self-doubt, which refers to “*the special case where what we doubt is our ability to*



achieve an epistemically favorable state, for example, to achieve true beliefs” (Roush, 2017, p. 1). Where epistemic doubt provides for the processes of knowledge creation, epistemic self-doubt disrupts them.

## Concluding remarks

In this paper, we have introduced and emphasized the importance of cultivating sophisticated epistemic doubt in order to be able to productively participate and work in the emergent era of AI. Epistemic doubt, as we define it, is both a personal and shared competence that we consider essential in the era of AI. Being able to take information ‘in trust’ while simultaneously holding epistemic doubt enables functioning and advancing knowledge-creation (Paavola & Hakkarainen, 2014) without getting stuck in endless skepticism when interacting with AI ‘black boxes,’ leaving the door open for further iterations of fluid shifts between trust and distrust. Furthermore, we argue that epistemic doubt cannot be considered something that naturally emerges from interactions with AI. Epistemic doubt needs to be deliberately fostered in education in the AI era. Although in education, the aim is not to support the human-like perceptions of AI, we should acknowledge that such perceptions might not be entirely possible to avoid because of the strive from the designers and manufacturers of AI systems to enhance such characteristics (e.g., Hu et al., 2021). This we consider emphasizing the need for competencies of epistemic doubt even further. Finally, we want to highlight that competencies of epistemic doubt are essential in the broader picture of the world to avoid division and extremism that emerge when people abandon or cannot cultivate their epistemic doubt (see Guerrero, 2021).

## References

- Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13337>
- Burgh, G., Thornton, S., & Fynes-clinton, L. (2018). “Do Not Block the Way of Inquiry”: Cultivating collective doubt through sustained deep reflective thinking. In E. Duthie, F. García Moriyón, & R. Robles Loro (Eds.), *Parecidos de familia. Propuestas actuales en Filosofía para Niños / Family Resemblances: Current trends in philosophy for children* (pp. 47–61). ANAYA.
- Floridi, L. (2023). AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, 36(1), 15. <https://doi.org/10.1007/s13347-023-00621-y>
- Guerrero, A. (2021). Doubt and the Revolutionary. *Midwest Studies in Philosophy*, 45, 423–456. <https://doi.org/10.5840/msp20219288>
- Hu, Q., Lu, Y., Pan, Z., Gong, Y., & Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 102250. <https://doi.org/10.1016/j.ijinfomgt.2020.102250>
- Lambert, J., & Stevens, M. (2023). ChatGPT and Generative AI Technology: A Mixed Bag of Concerns and New Opportunities. *Computers in the Schools*, 1–25. <https://doi.org/10.1080/07380569.2023.2256710>
- Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, 3. <https://doi.org/10.1016/j.caeai.2022.100056>
- Paavola, S., & Hakkarainen, K. (2014). Triological Approach for Knowledge Creation. In *Knowledge Creation in Education* (pp. 53–73). [https://doi.org/10.1007/978-981-287-047-6\\_4](https://doi.org/10.1007/978-981-287-047-6_4)
- Peirce, C. S. (1877). The Fixation of Belief. *Popular Science Monthly*, 12, 1–15.
- Pithers, R. T., & Soden, R. (2000). Critical thinking in education: a review. *Educational Research*, 42(3), 237–249. <https://doi.org/10.1080/001318800440579>
- Roush, S. (2017). Epistemic Self-Doubt. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University.
- Voosen, P. (2017). The AI detectives. *Science*, 357(6346), 22–27. <https://doi.org/10.1126/science.357.6346.22>
- World Economic Forum. (2023). *The future of jobs report 2023*. <https://www.weforum.org/reports/the-future-of-jobs-report-2023/>
- Zednik, C. (2021). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy and Technology*, 34(2), 265–288. <https://doi.org/10.1007/s13347-019-00382-7>

## Acknowledgments

This study was funded by the Strategic Research Council (SRC) established within the Research Council of Finland, grants #352859, #352876

# Reimagining “The Lantern” as a Web-Based Tool for Enhanced Supervised Collaborative Problem Solving

Héctor Florido, Universitat Pompeu Fabra, Barcelona, hector.florido@upf.edu  
Davinia Hernández-Leo, Universitat Pompeu Fabra, Barcelona, davinia.hernandez-leo@upf.edu

**Abstract:** The goal of this research is to offer a web-based lower cost tool version with few improvements applied to an already proposed ambient display (the so-called Lantern) for classroom orchestration of collaborative problem solving. By providing students with a queuing system and enabling teachers to monitor waiting and problem-solving times, the tool offers a comprehensive view of classroom dynamics. The research is divided into a first study with a pre-survey to identify needs and a study that registers the students' usage of the web-based tool to measure waiting times.

## Introduction

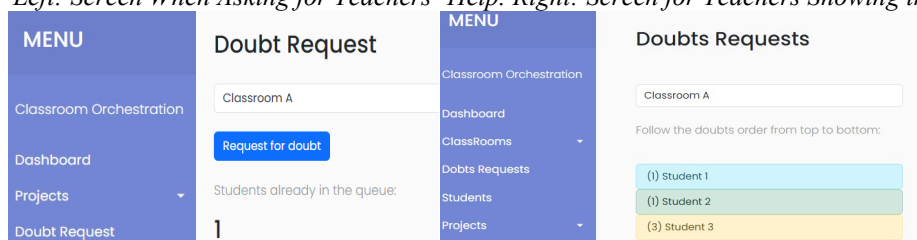
Considering the classroom as a learning community brings benefits when fostering active learning (Tissenbaum & Slotta, 2019). However, the orchestration of active learning classrooms such as recitation sessions, where students work on tasks and seek support from other students and the teacher, can be challenging and are usually far from optimal implementation (Alavi & Dillenbourg, 2012). Orchestration tools aim to support teachers in the real-time management of interactive classrooms, which usually involve collaboration at different social planes. Yet, surprisingly, despite the advancements in CSCL orchestration (Amarasinghe et al., 2023), the “Lanterns” approach by Alavi & Dillenbourg (2012) is still state-of-the-art orchestration technology for the specific case of co-located collaborative problem solving recitation sessions. The Lanterns is a hardware device designed to enhance ambient awareness of students' work status within a classroom environment. In this paper we aim at exploring how a lower cost web-based approach can address similar objectives when compared to the Lanterns, while also exploring current needs in recitation sessions and the avenues for improving the approach. The approach implements a queuing system that allows students to request assistance from the teacher, when facing a dashboard. The research questions explored in the paper are: RQ1) How do students believe orchestration tools for problem solving recitation sessions can improve their learning experience? RQ2) Does a web-based approach decrease the waiting time when compared to a control situation, as the Lanterns does?

## Methodology

The study group entailed two groups of high school vocational training students within the ages from 18 to 50 (total of 54 students). For the study, we implemented a needs analysis survey and an experimental study with two conditions. The control condition used a simplified version of the web-based tool functionalities for the students (not the teachers) to add themselves to a queue when having a doubt. This data was used only to collect timestamps, request of attention to the teacher was done orally/physically and is comparable to the “control setting” in the Lanterns study (Alavi & Dillenbourg, 2012). In the experimental sessions, the teacher was able to use the web-based tool (Figure 1) to see the exact order of the students who have raised their doubts through a queue. Once the teacher addresses a student's doubt, that student is removed from the queue and can rejoin the queue if they have further questions. This approach facilitates teachers to strictly follow the order of the students' doubts or to decide their own attention-to-student strategy based on the visualized data.

**Figure 1**

*Left: Screen When Asking for Teachers' Help. Right: Screen for Teachers Showing the Queue*



## Results and discussion

The collected data shows that there is a clear need to improve efficiency in the classroom: (1) The importance of improving the quality of responses of the teachers in the classroom, 50% rated it only as important. (2) The

importance of improving the clarity of responses of the teachers in the classroom, around 30% voted that it was very important. (3) The importance of improving efficiency in the classroom, students determined with more than the 50% of the votes that it was very important. The students were also asked whether they thought that a computer tool could improve class efficiency. Qualitative analysis showed recurrent comments pointing out the interest in having a queue-oriented system for teachers to consider resolving student's doubts in order. The need for using a tool for managing assistance directly ties into the first research question, thereby attributing significance to the orchestration tools for effective problem-solving and providing a better learning experience in active classrooms. Consequently, students can focus on the task with minor distractions related to asking for assistance and consistently monitor the teachers' locations. In the first session, teachers in the classroom did not check the web-based tool to see who was in the queue and who was not, so students had to raise their hand while adding themselves to the queue. The classroom doubts worked as normally, where students had to wait their turn with their hands raised while the teacher attended to them, and the teacher had to try to maintain a somehow fair order. Data was stored on a database and the waiting times of students, in the two conditions, were analyzed and grouped in 5-minute intervals.

**Table 1**  
*Waiting Time Comparison between Control and Experimental Sessions in Minutes.*

Time	Control (raise hand)	Experimental (waiting)	Experimental (resolving)
15-20	10%	8%	4%
10-15	34%	13%	2%
5-10	34%	21%	26%
0-5	22%	58%	68%

Table 1 shows all the waiting times and the calculated average waiting time was found to be 7.9 minutes since the student raised their hand until the teacher came. After these control sessions, the web-based tool was changed allowing the learning environment to be more productive by showing the teachers the strict order of the doubts into the queue. Figure 1 Right shows the interface that teachers can see when solving doubts. The data represents the waiting times of students using the web-based tool and the average wait time for students was 6.4 minutes. Also the time spent by a teacher helping students was observed and calculated in minutes (Table 1). The data we collected clearly shows that using the web-based tool reduces waiting times compared to a control session where students raise their hands for help. At this stage, the study's second question is addressed as the waiting time has been significantly reduced. Looking ahead, the tool aims to help both students and teachers.

The reimagination of the Lanterns into a web-based tool has been successfully executed, providing a cost-effective and portable tool for education communities. The benefits derived from the collected data affirm the effectiveness of the tool in improving classroom dynamics by significantly reducing the average waiting times for students during doubt-solving processes. In alignment with Alavi & Dillenbourg (2012) observations, our findings suggest that when teams are required to wait for the tutor, they typically opt for either immediately raising their hands, keeping them raised while continuing problem-solving, or alternatively, attempting to get the tutor's attention visually and waiting for an opportune moment. Moreover, a web implementation opens the door to introduce new smart functionalities to further enhance classroom and, also, peer-support orchestration.

## References

- Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies*, 5(3), 264-274.
- Amarasinghe, I., Dimitriadis, Y., Raes, A., Hernández-Leo, D., Martínez-Monés, A., van Leeuwen, A., et al. (2023). Orchestrating Hybrid Learning Scenarios: Challenges and Opportunities. In *ISLS Annual Meeting 2023* (p. 72).
- Tissenbaum, M., & Slotta, J. (2019). Supporting classroom orchestration with real-time feedback: A role for teacher dashboards and real-time agents. *Int. Journal of Computer-Supported Collaborative Learning*, 14, 325-351

## Acknowledgements

This work has been partially funded by UPF's Planetary Wellbeing (PLAWB00322), Spanish Ministry (PID2020-112584RB-C33, CEX2021-001195-M granted by MICIU/AEI/10.13039/501100011033), the Government of Catalonia (SGR 00930), DHL (Serra Hünter) acknowledges the support by ICREA (Academia). 51.

# Is a Collaboratively Created Visualization Worth 100,000 Words? A Case Study on Collaborative Learning Processes of Expert Groups in a Jigsaw Classroom

Julia Eberle, Paris Lodron Universität Salzburg, [julia.eberle@plus.ac.at](mailto:julia.eberle@plus.ac.at)  
Martin Bordewieck, Ruhr-Universität Bochum, [martin.bordewieck@rub.de](mailto:martin.bordewieck@rub.de)

**Abstract:** The jigsaw technique is a widely used macro-script for collaborative learning, albeit its effectiveness for learning is questionable, which might be due to high degrees of freedom concerning the details of its implementation. This study explores how scaffolds regarding the collaborative creation of visualizations, specifically concept maps and infographics, in expert groups are related to collaborative knowledge creation processes regarding the topic and effective preparation for teaching jigsaw groups. We conducted a case study on this scaffolding approach in several expert groups in a field setting in higher education courses. While the study is still work in progress, the paper provides a detailed outline of the theoretical reasoning for the study and its design.

## Introduction

The jigsaw technique (Aronson & Patnoe, 1997) is a widely used macro-script for collaborative learning. While many studies have shown positive effects of the jigsaw technique on self-esteem, self-efficacy, and prejudice reduction, mixed effects have been found for learning outcomes (c.f. Stanczak et al., 2022). The jigsaw technique consists of three phases (Aronson & Patnoe, 1997): 1) Students become temporary “experts” in a specific topic within a course or lesson, e.g. by individually reading learning material. 2) Experts form expert groups to discuss their topic and make sure that every expert has fully understood the topic. 3) Students form “jigsaw groups”, i.e. small groups of different experts with complementary knowledge, and each student teaches the group their unique expert knowledge. However, this broad design leaves much room for the concrete design of each phase so the implementation of a jigsaw script can differ substantially, and consequently, affect collaborative processes and learning outcomes (e.g. Deiglmayr & Schalk, 2015). An often observed but unwanted effect of the jigsaw technique is that learners know much about their own expert group’s topic but little about the topics of other expert groups (Deiglmayr & Schalk, 2015). A crucial point for avoiding this effect is how well experts understand their own topic and how they prepare for teaching their jigsaw group. Especially when complex learning material and core concepts are distributed between expert groups, experts’ understanding, and the quality of their teaching is vital. Providing the expert groups with scaffolds on how to collaboratively develop a good understanding of their topic and on how to design their “lesson” in the jigsaw group may be helpful to improve not only experts’ but also all other students’ learning in a jigsaw classroom.

Creating external visual representations has repeatedly shown to be helpful for understanding core concepts of a topic provided in complex learning and may also be helpful for preparing to teach these core concepts. Therefore, instructing expert groups in a jigsaw classroom to collaboratively create visualizations about their topics seems like a promising approach. Specifically, the creation of *concept maps (CMs)* is a useful means to promote understanding and teaching the core concepts and their relations. CMs are node-link-diagrams that represent concepts as nodes and their relations as links with labels to specify the relationships. They are especially beneficial for processing texts (Hilbert & Renkl, 2008) and collaboratively constructing CMs has shown to be very effective for learning (Schroeder et al., 2018). The creation of *infographics (IGs)*, in addition, is a means for providing complex information in very limited space and, as an aesthetically appealing artifact, for motivating learners and engaging them with the contents (Dunlap & Lowenthal, 2016). Collaboratively constructing IGs promotes visual, digital, and information literacy and facilitates learning and enhancing memory (Çifçi & Elaldi, 2021). However, the effects of scaffolds to collaboratively create CMs and IGs have not been investigated as tools to promote expert groups’ understanding of core concepts and promote their teaching preparations in jigsaw classrooms. As a first attempt in this direction, we conducted a case-study with expert groups which received such scaffolds in jigsaw-based higher education courses to investigate, how the scaffolds affect a) collaborative knowledge building about the core concepts and b) preparations for teaching their expert groups’ core concepts in jigsaw groups.

## Methods

The case study was conducted in the context of three introductory courses on educational psychology as part of a

B. A. program on educational studies. The courses were fully designed as jigsaw classrooms with an introductory session, four to nine expert group sessions, and eight jigsaw group sessions (each focusing on one of eight different topics). Students were divided into expert groups of two to four members, depending on the number of participants in the course. For the case study, two expert groups in each course were selected. These groups were audio-recorded during the expert group sessions, while they worked on understanding their topics and preparing to teach it in the jigsaw groups. All expert groups received two batches of text-based learning material that they were required to study at home and prepare for discussion in the expert groups. In class, the expert groups were first instructed to identify the important terms and prepare a shared glossary in a googledoc as preparation for the visualizations. Next, expert groups were instructed to collaboratively construct a digital concept map of the terms in the glossary using the graphic design website canva.com. Then, students were instructed to prepare infographics (using again canva.com) about their topic. The final expert group meeting was dedicated to collaboratively develop a lesson plan for teaching the jigsaw groups. For each of these tasks, students received instructions about its goal and strategies on how to approach it and were regularly reminded that the goal was that every expert was able to teach the topic to a jigsaw group on their own and that all students would be required to take a test about all eight topics later. In addition, the instructor coached each group individually during class. The expert group sessions were then followed by eight jigsaw group sessions, in which each expert had 60 minutes to teach their topic to their own jigsaw group. Students of the six voluntarily participating groups signed an informed consent. In addition to the audiotapes of their expert group discussions in class, these students received a questionnaire about demographic and psychometric data at the beginning and about questions concerning the collaboration and creation of visualizations after the expert group phase.

The audiotapes will be coded using a slightly modified version of the coding scheme proposed by Gijlers & de Jong (2013) in which all on-topic utterances are coded as either technical (related to digital tool use and the technical creation of the visual representations), regulative (related to regulation of learning and collaboration), transformative (related to knowledge building) or instructional (related to instructional considerations for teaching the topic to the jigsaw group). In addition, a coding scheme will be developed to analyze the visual representations regarding clarity, aesthetical effort, scope, and correctness. Based on these codings, collaborative processes of the individual groups will be visualized and descriptive results of distributions of the codes will be presented. Pre- and post-questionnaire, as well as the coded visual representations will be related to process data in order to gain understanding of the relation of collaborative group processes and group performance / satisfaction.

## Expected results and discussion

From the results of this study, we expect to gain insights whether instructing expert groups in a jigsaw classroom to collaboratively create visualizations is a helpful means to foster students understanding of the core concepts and enables them to prepare thoroughly for teaching the core concepts to their jigsaw groups or if this type of scaffold that focusses on the creation of an artifact rather distracts students from meaningful discussion and preparation activities. Consequently, this study will contribute to our understanding of the vital details that make the design and implementation of the jigsaw technique effective or ineffective regarding learning outcomes.

## References

- Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom: Building cooperation in the classroom*. New York: Longman.
- Çifçi, T., & Elaldi, Ş. (2021). The effectiveness of using infographics on academic achievement: A meta-analysis and a meta-thematic analysis. *Journal of Pedagogical Research*, 5(4), 92–118. <https://doi.org/10.33902/JPR.2021473498>
- Deiglmayr, A. & Schalk, L. (2015). Weak versus strong knowledge interdependence: A comparison of two rationales for distributing information among learners in collaborative learning settings. *Learning and Instruction*, 40, 69-78. <http://dx.doi.org/10.1016/j.learninstruc.2015.08.003>
- Dunlap, J. C., & Lowenthal, P. R. (2016). Getting graphic about infographics: design lessons learned from popular infographics. *Journal of Visual Literacy*, 35(1), 42–59. <https://doi.org/10.1080/1051144X.2016.1205832>
- Gijlers, H. & de Jong, T. (2013). Using concept maps to facilitate collaborative simulation-based inquiry learning. *Journal of the learning sciences*, 22(3), 340-274, <https://doi.org/10.1080/10508406.2012.748664>
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and Constructing Concept Maps: a Meta-Analysis. *Educational Psychology Review*, 30(2), 431–455. <https://doi.org/10.1007/s10648-017-9403-9>
- Stanczak, A., Darnon, C., Robert, A., Demolliens, M., Sanrey, C., Bressoux, P., Huguet, P., Buchs, C., & Butera, F. (2022). Do Jigsaw classrooms improve learning outcomes? Five experiments and an internal meta-analysis. *Journal of Educational Psychology*, 114(6), 1461-1476. <https://doi.org/10.1037/edu000073>

# Teaching Multimedia Design through Online Collaboration during COVID-19: Laurillard's Conversational Framework

Blessed A. Mhungu, University at Buffalo, [blessedm@buffalo.edu](mailto:blessedm@buffalo.edu)  
Ralitsa D. Debra, Kwame Nkrumah University of Science and Technology, [rdebrah.art@knust.edu.gh](mailto:rdebrah.art@knust.edu.gh)  
Dick Ngambi, University of Cape Town, [dick.ngambi@uct.ac.za](mailto:dick.ngambi@uct.ac.za)

**Abstract:** The COVID-19 pandemic in 2020 posed challenges for instructors who relied on using traditional teaching methods, especially in undergraduate courses. To address this, an online course intervention was created to enhance communication and collaboration skills using digital technology tools. Employing Laurillard's Conversational Framework, the study investigated the feasibility of fostering online collaboration and communication. Data collection methods included interviews, focus groups, and online course observations. Findings underscored the significance of instructor guidance, social presence, discussion boards, clear communication channels, and technology integration.

## Context and background

The COVID-19 pandemic in 2020 increased the demand for emerging technology and digital technology learning methods. However, this shift posed significant challenges for instructors accustomed to traditional contact teaching methods. At a university in Cape Town, South Africa, many instructors struggled to adapt to virtual teaching during the pandemic-induced lockdown (Schildkamp et al., 2020). Consequently, some courses were hastily developed and inadequately prepared for emergency remote online education (Hodges et al., 2020). Although some instructors struggled to adjust, a few rose to the occasion, "becoming instructional 'MacGyvers', having to improvise quick solutions in less-than-ideal circumstances" (Hodges et al., 2020, p. 3).

This study is about the experiences of instructors who developed a student-centered learning multimedia foundations online course using Laurillard's conversational framework, fostering deep engagement characterized by collaboration and communication. The instructors in this study employed a problem-based learning (PBL) approach to design the course, promoting a student-centered learning experience that nurtured critical 21st-century skills among undergraduate students (Lynch et al., 2017). Unlike conventional multimedia foundations courses that require students to develop physical print portfolios as part of on-going formative assessments, this course redesign required students to develop a digital e-portfolio website to showcase their work.

## Theoretical framework

In this study, we applied Laurillard's Conversational Framework to identify learning types that enhance engagement, collaboration, and communication in an undergraduate multimedia foundations course during the COVID-19 pandemic. Originally designed to analyze learning processes in technology-enhanced environments, Laurillard's framework focuses on educational technology and digital learning contexts (Laurillard, 1993). Our adaptation of Laurillard's framework aimed to "capture the communicative, reflective, and goal-oriented actions with feedback necessary to support the complete online learning process" (Neo et al., 2013, p. 8).

## Research methodology

To address the research question, '*Which learning types promote engagement, collaboration and communication?*', we employed a phenomenological case study approach, allowing us to draw insights from the perspective of instructors and students who participated in the online multimedia foundations course we had developed (Merriam & Tisdell, 2015, p.110). Data collection spanned over a fifteen-week period and included interviews, focus groups, and online course observations.

## Data analysis

We utilized MAXQDA, a qualitative computer-aided data analysis software, to generate themes and codes via open coding and axial coding. This process involved identifying similar patterns, linkages, or contradictions in the data, enabling us to determine whether to amalgamate, segment, or eliminate themes. The themes underwent continuous refinement, with any emerging subthemes being appropriately labeled and categorized.

## Preliminary findings

In our study, students communicated and collaborated with one another through instructor guidance. Initially, they faced challenges in forming and coordinating their own working groups due to physical distance and uncertainty about technology tools to use. However, with guidance from the instructor, they successfully organized their groups, by selecting group leaders and assigning tasks. Additionally, we discovered that adopting Laurillard's conversational framework and the problem-based learning (PBL) approach facilitated diverse learning types and engagement through digital technology tools. These learning types included pedagogical conversations (*instructor-student interactions*) through formal tools like Blackboard and MS Teams, conversational exchange (*student-student interactions*) through informal tools such as Google Keep and WhatsApp, and interactive conversations (*student-tool interactions*). Digital technology enabled us to identify these learning types as prominent features of our redesigned online multimedia foundations course (Ngambi et al., 2016).

## Discussion

The study found that the students in our newly designed multimedia course maintained positive attitudes toward the online learning environment we created during the COVID-19 pandemic. Laurillard's Conversational Framework guided our efforts to transform the traditional course to an online course. Our online multimedia course encouraged transformation of traditional instructional practices, shifting the instructor's role from master to mentor. Technology played a key role in facilitating communication, feedback, and collaboration between the instructors and students. In teaching multimedia design projects through online collaboration, we also considered recommendations from Mishra and Koehler's study (2006), as cited in Çam and Koç (2019) by prioritizing discussions and dialogues mediated through various digital technology tools.

## Conclusion and future implications

This study demonstrates the feasibility of delivering a multimedia foundations course in South Africa's higher education system through technology-mediated collaboration. Our approach addresses the limitations of current instructional practices, where communication and interactions are often minimal. Implementing online collaborations involves increased interactions between teachers and learners, with instructors assuming roles as guides or mentors. However, this transition requires additional time, planning, and resources, which many instructors currently lack due to work pressures and administrative challenges. Our future work focuses on addressing these challenges and developing instructional guidelines to support instructors interested in designing collaborative online courses.

## References

- Çam, Ş., & Erdamar Koç, G. (2019). Technological Pedagogical Content Knowledge Practices in Higher Education: First Impressions of Preservice Teachers. *Technology, Knowledge and Learning*, 26(1), 123-153. DOI: 10.1007/s10758-019-09430-9.
- Hodges, C., Moore, S., Locke, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*, 27.
- Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.
- Lynch, S., Peters-Burton, E., Behrend, T., House, A., Ford, M., Spillane, N., Matray, S., Han, E., & Means, B. (2017). Understanding inclusive STEM high schools as opportunity structures for underrepresented students: Critical components. *Journal of Research in Science Teaching*, (55), 712-748.
- Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684>.
- Neo, M., Neo, K., & Lim, S. (2013). Designing a Web-Based Multimedia Learning Environment with Laurillard Conversational Framework: An Investigation on Instructional Relationships. *TOJET: The Turkish Online Journal of Educational Technology*, 12(3), 39-48. Online. Available: <https://files.eric.ed.gov/fulltext/EJ1016940.pdf>
- Ngambi, D., Brown, C., Bozalek, V., Gachago, D., & Wood, D. (2016). Technology enhanced teaching and learning in South African higher education - A rear-view of a 20-year journey. *British Journal of Educational Technology*, 47(5), 843-858. <https://doi.org>
- Schildkamp, K., Wopereis, I., Kat-De Jong, M., Peet, A., & Hoetjes, I. (2020). Building blocks of instructor professional development for innovative ICT use during a pandemic. *Journal of Professional Capital and Community*, 5(3/4), 281-293. <https://doi.org/10.1108/jpcc-06-2020-0034>

## Use of AI to Foster Greater Awareness and Reflexivity around Harmful Content in Online Discussion Tools

Kim MacKinnon, Jingyu Liu, Cresencia Fong, Jai Pandey  
kim.mackinnon@utschools.ca, jingyu.liu@utschools.ca, cresencia.fong@utschools.ca, jpandey@utschools.ca  
Eureka Research Institute at UTS (University of Toronto Schools)

**Abstract:** In this project, we explore the development of an AI-powered harmful content detector for an online discussion tool. We explored three possible large language processors (Moderation API, roBERTa, and ChatGPT) using a dataset of 5,747 publicly-accessible online posts. Results indicated that the pre-trained, fine-tuned roBERTa processor was the most accurate at detecting harmful content and distinguishing it from non-harmful content.

### Context and theoretical perspective

This study examined the development of an automated content moderation tool, particularly aimed at flagging posts containing harmful content within a locally developed online discussion platform. The larger aims of the project focus on the responsibility among K-12 schools to foster future-ready knowledge and skills among students (CMEC, 2016; OECD, 2018), such as effective communication and collaboration, and more recently AI-related global competencies (Dede, Etemadi, & Forshaw, 2021).

### Leveraging AI to reduce harm in online discussion

Beginning in May 2023, we had an opportunity to leverage AI in the design of a school-based threaded discussion tool that would offer students the chance to engage in ‘sober second thought’ before posting a message that is potentially harmful. Liu (2023) notes there are limited open source options of generative AI language models to detect harmful online posts, though, there are some content detection models that target specific types of harmful content, such as hate speech (Abro et. al, 2020), and cyberbullying (Reynolds et. al, 2011). However, one classifier known as the Moderation API by OpenAI provided one potential opportunity to move towards a more expansive model, allowing for flagging of: *hate, harassment, explicit sexual references, violence, and self-harm.*

Our rationale for integrating automated content detection into our school online discussion platform was threefold: 1) a fiduciary responsibility to ensure the safety of the learning environment (including online) for students, 2) the impracticality of manual daily monitoring of thousands of individual posts for potentially harmful content, and 3) student use of unfamiliar slang, shorthand, and insider references, which are continuously evolving. Furthermore, our intent was not simply to delete posts, or even to “reduce” (Gillespie, 2022) the noticeability of problematic posts, but to provide an opportunity for teachers to intervene to help foster development of important future-ready competencies, in this case digital rights and responsibilities.

We added some basic features to the discussion interface that would allow for flagging of potentially harmful content including: 1) a warning message that appears to the student prior to posting if the AI processor has detected harmful content, and includes the option to ‘edit’ or ‘post’ anyway, 2) the ability for teachers to quickly see posts that have been flagged (either manually or by the AI) as having harmful content, 3) the ability for other students to manually and anonymously flag a post as harmful, and include a written rationale that would be seen by the teacher, and 4) the ability for the teacher to see a version history of a post (including deletions).

### Research question

Our specific question was: *Which of three large language processors were most effective at supporting accuracy, precision and recall of harmful content?*

### Methodology

Three possible content moderation tools were compared for their sensitivity to the occurrence of harmful content detection: RoBERTa, ChatGPT, and the existing Moderation API. We first pre-trained a local instance of RoBERTa to detect harmful content. Publicly available posts from the social media application called Reddit were used for this purpose, including posts gathered from subreddits (e.g., on the topic mental health) to ensure there was a balanced dataset for training and testing that reflected all of the categories. Using the Moderation API we were able to efficiently create a coded list of harmful and non-harmful posts using the five categories mentioned above, which could then be used as a training data set. Once training was complete, all three detection tools were evaluated using a second test dataset.



## Data collection

A total of 45,974 posts were used to train the RoBERTa model. A smaller number of posts (5,747) was used as the test dataset with all three language processors. For each processor, the following recordings were taken: true positives, false positives, true negatives, and false negatives. Ideally, the best processor would be able to both maximize true returns and minimize false returns.

## Findings and discussion

Table 1 below shows the four results across the three processors that were examined:

**Table 1**

*Results of Each of the Three AI Processors (non-harmful=2890, harmful=2857, total=5747)*

	True Positives	False Positives	True Negatives	False Negatives
Moderation API	2058	799	2779	111
RoBERTa	2715	142	2664	226
ChatGPT	2381	476	2662	228

Overall, the Moderation API had an accuracy of 84%, RoBERTa had an accuracy of 94%, and ChatGPT had an accuracy of 88%. Although false negatives remained comparatively high, the fine-tuned RoBERTa processor was the standout language model for content moderation and harmful content detection, by maximizing true results while minimizing (overall) the false ones. Furthermore, given that RoBERTa is an open source processor and able to run on local hardware, it also stood as the most cost effective and capable for supporting privacy of school data.

## References

- Liu, J. (2023). Transforming learning through AI-enhanced online discussions: Fostering greater awareness and reflexivity around potentially harmful content. Final paper presented at ARIA Showcase, UofT.
- Abro, S. Shaikh, S., Hussain, Z., AliZafar, K., Khan, S., & Mujtaba, G. (2020). Automatic hate speech detection using machine learning: A comparative study. *International Journal of Advanced Computer Science and Applications*, 11(8).
- CMEC. (2016). *CMEC Pan-Canadian Global Competencies Descriptions*. Council of Ministers of Education, Canada. [http://www.ontariodirectors.ca/CODE-TLF/docs/tel/PanCanadian\\_Global\\_Competencies\\_Backgrounder\\_EN.PDF](http://www.ontariodirectors.ca/CODE-TLF/docs/tel/PanCanadian_Global_Competencies_Backgrounder_EN.PDF)
- Dede, C. Etemadi, A., & Forshaw, T. (2021). *Intelligence augmentation: Upskilling humans to complement AI*. The Next Level Lab at the Harvard Graduate School of Education. President and Fellows of Harvard College: Cambridge, MA.
- Gillespie, G. (2022). Do not recommend: Reduction as a form of content moderation. *Social Media & Society*, 8(3), <https://doi.org/10.1177/20563051221117552>
- OECD. (2018). *The Future of Education and Skills—Education 2030: The Future We Want*. Organization for Economic Cooperation and Development. [https://www.oecd.org/education/2030-project/about/documents/E2030%20Position%20Paper%20\(05.04.2018\).pdf](https://www.oecd.org/education/2030-project/about/documents/E2030%20Position%20Paper%20(05.04.2018).pdf)
- Reynolds, K., Kontostathis, A., & Edwards, L. (2011). Using machine learning to detect cyberbullying. In *10th International Conference on Machine Learning and Applications and Workshops*, volume 2, pages 241–244, 2011. doi: 10.1109/ICMLA.2011.152

## Acknowledgements

This research was supported by funding from the Eureka Research Institute at UTS, and the Mitacs Accelerator program.

## Iterative or Linear Design? Insights into Teacher Professional Development on Dialogic Teaching

Ying Zhang, The University of Hong Kong, bobzhang0829@connect.hku.hk

Yu Liu, Southwest University, morghlis@163.com

Zixiang Yu, The University of Edinburgh, andrewyu2019@gmail.com

Yu Li, The University of Hong Kong, ylioi@connect.ust.hk

Gaowei Chen, The University of Hong Kong, gwchen@hku.hk

**Abstract:** This study aimed to compare the effectiveness of a teacher professional development (PD) program on dialogic teaching designed in either an iterative or a linear mode. Two mathematics preservice teachers, Ms. Liu and Ms. Han, were selected as representative cases. Results showed that Ms. Liu scored much higher on coding tests, indicating more solid declarative knowledge of dialogic teaching. It implied that the iterative PD mode may help better consolidate the understanding of dialogic teaching.

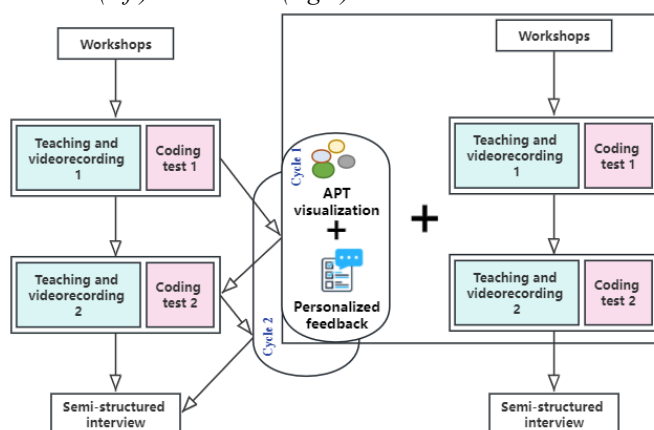
### Introduction

Research has shown that effective classroom interactions in mathematics lessons profoundly impact teaching and learning (e.g., Chen et al., 2020). Many frameworks, such as the Academically Productive Talk (APT, see Michaels et al., 2008), have been proposed to sharpen pre-service teachers' (PSTs) skills in dialogic teaching. However, many mathematics PSTs are trapped with limited chances to transfer productive dialogic teaching into practice systematically.

Considering the urgent need to design a suitable PD program to hone their dialogic teaching skills in authentic contexts, many studies highlight the need for developing and implementing an iterative PD mode (e.g., Moallem et al., 2015). Few studies have examined its counterpart, i.e., the linear mode, which refers to the non-iterative implementation of dialogic teaching during PD programs.

This study aimed to evaluate and compare the effectiveness of a PD program in either an iterative or a linear mode (see Figure 1). Specifically, it set out to address how the PD program in these two modes impacts mathematics PSTs' practical skills and declarative knowledge of dialogic teaching.

**Figure 1**  
*Iterative (left) and Linear (right) PD Modes*



### Methodology

This study adopted a QUAL–quan mixed method approach. Two mathematics PSTs – Ms. Liu (with iterative PD mode) and Ms. Han (with linear PD mode) – were selected as representative cases from an ongoing large-scale PD project. Both PSTs were highly motivated and showed no prior knowledge or experience in implementing dialogic teaching and the APT framework.

Following their attendance at the training workshops centered on dialogic teaching, both PSTs were asked to video–record two lessons and complete two coding tests (Tests 1 and 2), which entailed identifying APT talk moves (e.g., say more, revoice) in two real math lesson clips selected by the research team. The coding tests

were conducted to examine the declarative knowledge of dialogic teaching based on the APT framework. The percentage of APT talk moves used in the recorded classes and the performance of coding tests were calculated. For data triangulation, semi-structured interviews were conducted individually with two PSTs.

## Findings and discussion

Both PSTs demonstrated similar proficiency in practical skills for using APT talk moves during their two video-recorded lessons. Regarding the use of APT talk moves, both PSTs acknowledged the pedagogical benefits of dialogic teaching, e.g., activating students' thinking and making meaning in their own words. As Ms. Han said,

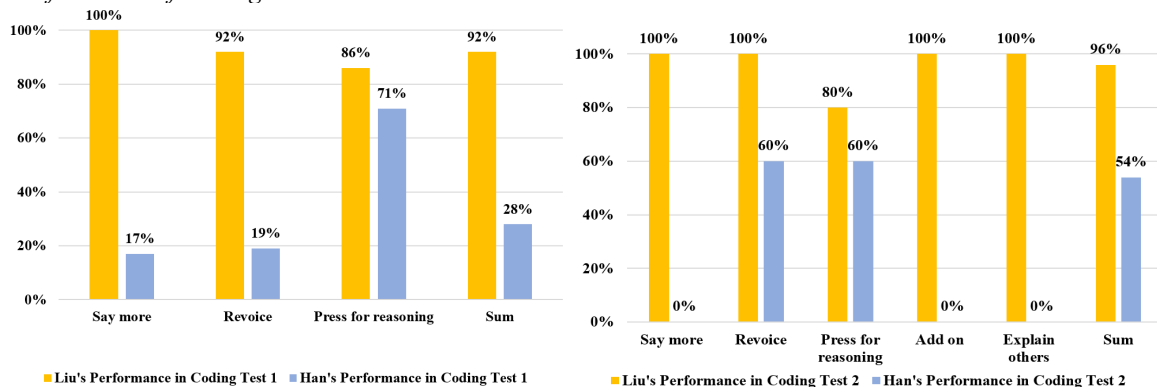
“Strategies like *say more* and *challenge* can activate their thinking, grow their mindset, and guide them to follow our lesson procedure. Otherwise, they are sitting in the classroom, but they are absent-minded.” (Ms. Han)

For the two coding tests, Ms. Liu scored much higher on both tests than Ms. Han (see Figure 2). This shows that Ms. Liu (with iterative PD mode) demonstrates more concrete declarative knowledge of the APT framework than Ms. Han (with linear PD mode). For Ms. Han, without giving her timely feedback on her strengths and difficulties of using APT talk moves yet simply asking her to finish the coding tests after recording the two lessons (see Figure 1), she completed the coding tests “mostly based on the memory (Ms. Han).” In comparison, when Ms. Liu was offered personalized feedback on her APT use in her first class teaching, she likely drew on the feedback to inform her future practice. As Ms. Liu said,

“When I saw the feedback on the first lesson, I felt that there weren't many APT talk moves. So, I consciously designed more for the second lesson.” (Ms. Liu)

It is argued that the iterative PD mode may help better consolidate the understanding of dialogic teaching for PSTs. However, comparing the two PD modes needs to be examined on a larger scale in future research.

**Figure 2**  
*Performance of Coding Tests 1 and 2*



## References

- Borko, H., Koellner, K., & Jacobs, J. (2014). Examining novice teacher leaders' facilitation of mathematics professional development. *The Journal of Mathematical Behavior*, 33, 149–167.
- Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of the Learning Sciences*, 29(4–5), 642–680.
- Michaels, S., O'Connor, M. C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized. *Studies in the Philosophy of Education*, 27(4), 283–297.
- Moallem, M. (2015). The impact of synchronous and asynchronous communication tools on learner self-regulation, social presence, immediacy, intimacy and satisfaction in collaborative online learning. *The Online Journal of Distance Education and e-Learning*, 3(3), 55–77.

## Acknowledgement

This work was supported by the ITC of the Government of the HKSAR (Grant No.: ITB/FBL/7026/20/P).

# Navigating Boundary Crossing in a Global Engineering Course: A Case Study

Ndubuisi Anuli, Asare Philip, Ajaz Mesha, Slotta Jim

Annuli.ndubuisi@mail.utoronto.ca, philip.asare@utoronto.ca, mesha.ajaz@mail.utoronto.ca, jslotta@gmail.com  
University of Toronto

**Abstract:** Industry and accreditation requirements for engineering graduates with global leadership competencies, such as global perspectives and multidisciplinary teamwork, have necessitated the need to support students to engage in complex collaborations that cross multiple boundaries (International Engineering Alliance, 2021). This paper reports on a case study of two design project teams from a globally situated engineering design course. It investigates how the boundaries that students encountered in globally situated design projects helped them to navigate conflict and make progress in problem-solving. Following a thematic analysis of students' data, two themes emerged around their knowledge boundary-crossing experiences and associated cultural dimensions namely: 1) unfamiliarity with design for global conditions, and 2) unfamiliarity with cultural and environmental impact.

## Introduction

With engineering world becoming increasingly global, engineering regulators and policymakers' have mandated that future graduates develop global leadership competencies, such as global perspectives and multidisciplinary teamwork (Engineers Canada, 2021). Hence, engineering educators have integrated complex collaboration projects, which allow students to work together with diverse partners across multiple boundaries including organizational, disciplinary, cultural, or geographical, into the engineering curriculum to help students build these global engineering leadership competencies (e.g., Ndubuisi et al, 2022). While these boundaries, seen as differences or unfamiliar practices (Akkerman & Bakker, 2011; Mankin, Cohen & Fitzgerald, 2004) can give rise to conflict (Jehn, 1997), they can also lead to transformational learning. This study examines complex collaboration boundaries to understand how they can support students' development of global competencies. The current paper advances our understanding of boundary crossing that occur within a design team, specifically: a) what boundaries were encountered in globally situated engineering design projects and, b) how can these boundaries enable students to make productive progress in global leadership skills?

## Theoretical perspectives

The study was guided by Vygotsky's (1978) social constructivist account of students' learning and development through socially co-constructed interactions, and Akkerman & Baker's (2011) *boundary crossing* concepts which allowed for the examination of the students' interactions and learning experiences across multiple boundaries within their design work. Boundary crossing refers to situations where individuals transition and interact across different sites or communities of practice (Akkerman & Baker, 2011).

## Method

This research is situated within a global engineering design (GED) course at a large North American University that enabled students to build engineering leadership and management skills within multidisciplinary engineering design team projects, while tackling real world design challenges with global significance. The course prioritized inclusive collaboration and transparent teamwork practices by engaging students with project management methods and reflections to support their design work. This paper utilized a case study methodology (Yin, 2009) and reports on two student design project teams, consisting of 4 members each, from the GED course. We examine two teams, in particular: Team 1 tackled the problem of water hyacinths - an invasive species of plant - overrunning the waterways, known as "klongs", in communities within Thailand. While Team 2 addressed the problem of reducing plastic wastes in a Ghanaian community.

Data were collected from students' team meeting recordings to provide insights into the teams' boundary-crossing interactions, conflict management and decision-making practices. Using thematic analysis (Braun & Clarke, 2012), the transcripts were deductively coded to Mankin, Cohen, and Fitzgerald's (2004) complex collaboration boundaries to capture boundary-crossing situations and emerging conflicts in the teams. Next, the coded data was collaboratively examined by two researchers to verify the coded categories and to determine patterns between the identified boundaries, associated conflicts, and the problem-solving approaches

applied by the students. Thereafter, the data were systematically examined for emerging themes within students' productive – and disruptive – exchanges, as well as any learning across the boundaries that they encountered.

## Findings and discussion

Our preliminary findings identified two themes around students' knowledge boundary-crossing experiences associated with the global design conditions: 1) unfamiliarity with the cultural dimensions, and 2) unfamiliarity with environmental and cultural impact.

### Unfamiliarity with the cultural dimensions of global conditions

Students in Team 1 struggled with the cultural dimension of a knowledge boundary that they encountered in the design work – the non-commonality of water hyacinths in the local Canadian environment. This difficulty led to a process conflict since they could not touch, feel nor test the plant to validate their proposed design solution as highlighted by a student: "...this feels like a really hard opportunity to work on given the timeframe. Yeah, like, without any access to, you know, water hyacinths to actually test to see if this works even slightly at all." To tackle this problem, the students applied mental modelling approaches to help them create an internal representation of the water hyacinth's behavior and visualize its response to mentally simulated tests.

### Unfamiliarity with the environmental and cultural impact of global conditions

Both Team 1 and Team 2 students were unfamiliar with the environmental impact of their proposed solution. For example, students in Team 1 experienced some cognitive conflict exploring modification approaches for an existing canal that can conserve the cultural integrity of the host environment: "I just had a question about ...environment analysis, ... is it compatible? Like, can it be built with minimal structural modification?"

Similarly, students in Team 2 encountered some cognitive conflict with the potential hazardous impact of their proposed solution to the host community: "Burning plastic definitely is an environmental concern." Thereafter, they brainstormed ideas for addressing the problem: "...if we're actually having to melt this stuff, we'll probably want to be conscious about open flame, which like, the wind to actually catch fire," and "I'm not entirely sure that we'll have that many issues with like the plastic itself burning and creating too many fumes, since we're gonna be melting at such a low temperature."

## Conclusion

This study is significant as it demonstrates students' leadership experiences, specifically the knowledge boundary-crossing situations that they encountered and the divergent thinking and mental modeling approaches they applied to advance their global design work. The findings can help educators understand strategies for incorporating global perspectives into their curriculum. This research contributes to both theory and practice of learning across boundaries within global engineering education and research on interdisciplinary collaboration.

## References

- Akkerman, S. F., & Bakker, A. (2011). Learning at the boundary: An introduction. *International Journal of Educational Research*, 50(1), 1–5.
- Braun, V., & Clarke, V. (2012). Thematic analysis. *American Psychological Association*.
- Engineers Canada. (2021). CEAB 30 by 30 Working Group Report. Accessed on March 31st 2024 at <https://engineerscanada.ca/sites/default/files/202111/Updated%20final%20CEAB%2030%20by%2030%20WG%20Report.pdf>
- Jehn, K. A. (1997). A Qualitative Analysis of Conflict Types and Dimensions in Organizational Groups. *Administrative Science Quarterly*, 42(3), 530.
- Mankin, D., Cohen, S., & Fitzgerald, S. P. (2004). Developing Complex Collaborations: Basic Principles to Guide Design and Implementation. In *Advances in Interdisciplinary Studies of Work Teams*, 10, 1–26. Emerald Group Publishing Limited.
- Ndubuisi, A., Marzi, E., Mohammed, D., Edun, O., Asare, P., & Slotta, J. (2022). Developing Global Competence in Global Virtual Team Projects: A Qualitative Exploration of Engineering Students' Experiences. *Journal of Studies in International Education*, 26(2), 259–278.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press.
- Yin, R. K. (2009). Case study research: Design and methods. *Sage*. 5.

# Supporting Dialogic Reflection through Visualising Classroom Talk in Teachers' and Students' Learning Communities

Pengjin Wang, Fan Chen, Deliang Wang, Gaowei Chen  
wangpj@connect.hku.hk, u3009428@connect.hku.hk, wdeliang@connect.hku.hk, gwchen@hku.hk  
Faculty of Education, The University of Hong Kong

**Abstract:** Dialogic reflection highlights the collaborative and data-led aspects of reflective practice. This study leverages visualisation technology to support teachers' and students' learning, focusing on fostering productive classroom talk. After participating for three months, two ICT teachers could better employ productive classroom talk while their students' subject knowledge of ICT improved. This study provides implications on the effective use of visualisation of classroom talk to support evidence-based, collaborative and dialogic reflection.

## Introduction

Dialogue is essential in human learning and development (Vygotsky, 1978), and reflection plays a fundamental role in higher mental functions, which are two defining characteristics of humans. Dialogic reflection combines these two functions, highlighting 'embracing a dialogic/collaborative view of reflection that allows potentially richer articulation and analysis' (Mann and Walsh, 2013, p. 303). To promote dialogic reflection among teachers and students, this study utilizes visualisation technology as a promising tool that transforms classroom talk into data for teachers' and students' reflection in a community of inquiry.

Dialogic reflection can focus on a wide array of elements and events within a learning community. One of the important aspects is classroom interaction during a lesson, as certain ways of organizing classroom talk are productive for learning (Michaels & O'Connor, 2015). Academically productive talk (APT) moves could be a valuable tool to orchestrate classroom talk. It has four goals, including (a) helping students share their own thoughts, (b) helping students listen carefully to one another, (c) helping students deepen their reasoning, and (d) helping students think with others (Michaels & O'Connor, 2015).

Visualisation of classroom talk indicates whether the classroom talk is productive. The Classroom Discourse Analyzer (CDA) is one of the representative tools, that transform classroom videos into visual data for analysis. CDA has proven effective in teachers' professional learning contexts, enhancing their use of productive classroom talk (see Chen, Chan et al., 2020 for details). In this study, the CDA was utilized to facilitate teachers' and students' reflection. This study investigates how visualisation-supported dialogic reflection (VSDR) influences students and teachers in a community, following the research question: To what extent does VSDR impact students' learning outcomes and teachers' use of productive classroom talk?

**Table 1**

*Low- and high-dose groups' performance in ICT subject knowledge and perceived teachers' APT*

	Dose	n	Pre	Post	t	df	p
			M (SD)	M (SD)			
ICT subject knowledge	Low	27	4.96 (1.02)	5.19 (0.92)	-0.844	26	.406
	High	26	4.77 (1.18)	5.65 (0.56)	-3.537	25	.002**
Perceived teachers' APT	Low	27	3.91 (.53)	4.29 (.62)	-2.826	26	.009**
	High	26	3.81 (.82)	4.17 (.68)	-2.258	25	.033*

\* $p < .05$ , \*\* $p < .01$ .

## Methods

The research was conducted in four 2<sup>nd</sup>-grade classes from an international primary school in Hong Kong, China. We conducted a three-month classroom intervention collaborating with two ICT teachers. To carry out the intervention, each teacher and his/her two classes form a group (i.e., two groups of teachers and students in total). As the school management team objected to using a control group, both groups received treatment, but with a higher dose (14 times of visualisation of classroom videos), and a lower dose (8 times of visualisation of classroom videos). Here, visualisation refers to a visual representation of the classroom talk in classroom videos. Each lesson video lasts for 30 minutes. After the visualisation transformation, teachers could choose to view the visualisation by themselves or select clips for students to view, supplemented with a discussion forum.

The pre- and post-tests of subject knowledge were designed by two teachers with the same level of difficulty and were administered before and after the implementation. Then, pre- and post-questionnaire were

used on student-perceived teacher APT in class, which contains two scales developed by Chen, Zhang and colleagues (2020) with six items in total, using a five-point Likert scale ranging from 1 (rarely) to 5 (frequently). It measures students' perception of teachers' use of APT moves during classroom talk, aligning with the goals of productive talk mentioned previously (Michaels & O'Connor, 2015).

## Results and discussion

First, to examine whether students' subject knowledge of ICT and their perceptions of teachers' APT were similar at the beginning of the intervention, an independent samples t-test was performed. The pre-test results revealed no significant difference in students' subject knowledge,  $t(51) = -.642, p = .524$ , and their perception of teachers' APT use,  $t(51) = -.563, p = .576$ , despite the low-dose group (27 students) attaining higher scores in subject knowledge tests and rating their teacher's use of APT higher (subject knowledge:  $M = 4.96, SD = 1.02$ ; teachers' APT:  $M = 3.91, SD = 0.52$ ) compared to the high-dose group (26 students) (subject knowledge:  $M = 7.77, SD = 1.18$ ; teachers' APT:  $M = 3.81, SD = .82$ ).

Second, a paired samples t-test was performed (see Table 1) to compare students' subject knowledge of ICT in pre-test and post-test within both the low-dose and high-dose groups. For the low-dose group, despite an increase in the mean scores in ICT subject knowledge between the pre-test ( $M = 4.96, SD = 1.02$ ) and post-test ( $M = 5.19, SD = 0.92$ ), the results were not statistically significant;  $t(26) = -0.844, p = .406$ . Conversely, the high-dose group exhibited a significant improvement in ICT subject knowledge from pre-test ( $M = 4.77, SD = 1.18$ ) to post-test ( $M = 5.65, SD = 0.56$ );  $t(25) = -3.537, p = .002$ . The result suggests that only students in the high-dose group had improved learning outcomes significantly over time.

Third, for students' perception of teachers' APT, students in both low-and high-dose groups differ significantly between pre-test (low-dose:  $M = 3.91, SD = .53$ ; high-dose:  $M = 3.81, SD = .82$ ) and post-test (low-dose:  $M = 4.29, SD = .62$ ; high-dose:  $M = 4.17, SD = .68$ ); low-dose:  $t(26) = -2.826, p = .009$ ; high-dose:  $t(25) = -2.258, p = .033$ . This suggests that students from both groups perceived a significant increase in their teachers' use of APT strategies over time.

To bridge the research gap, this study investigated whether and how visualisation as a technology supported dialogic reflection among students and teachers in a community. In terms of students' performance, both groups increased, but only the high-dose group's performance shows statistical significance. For the students perceived teachers' APT, students in both groups regarded significantly higher teachers' use of APT moves in the post-test. As a case study, we can see the potential of visualisation as a promising reflective tool that affords dialogic reflection for both teachers and students, since the experience is discursive, data-led and focuses on building a community (c.f. Mann & Walsh, 2013). Future research should consider collecting data on collaborative lesson planning sessions to reveal the dialogues during the lesson preparation for an enriched understanding.

## References

- Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of the Learning Sciences, 29*(4–5), 642–680. <https://doi.org/10.1080/10508406.2020.1783269>
- Chen, G., Zhang, J., Chan, C. K. K., Michaels, S., Resnick, L. B., & Huang, X. (2020). The link between student-perceived teacher talk and student enjoyment, anxiety and discursive engagement in the classroom. *British Educational Research Journal, 46*(3), 631–652. <https://doi.org/10.1002/berj.3600>
- Mann, S., & Walsh, S. (2013). RP or “RIP”: A critical perspective on reflective practice. *Applied Linguistics Review, 4*(2), 291–315. <https://doi.org/10.1515/applirev-2013-0013>
- Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. In Lauren B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through talk and dialogue* (pp. 347–362). AERA.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

## Acknowledgement

This work was supported by Hong Kong RGC, University Grants Committee (Grant No.: 17605221).

## Spatial Manipulative Note-Taking Tool for Small Group Face-to-Face Collaboration in Science Class

Litong Zeng, University of Illinois at Urbana Champaign, litongz2@illinois.edu

Xuesong Cang, University of Wisconsin Madison, xcang@wisc.edu

Afroza Sultana, Toronto Metropolitan University, afroza.sultana@torontomu.ca

Stacy Cernova, Toronto Metropolitan University, acernova@torontomu.ca

Megan Wang, Toronto Metropolitan University, meganl.wang@torontomu.ca

Dana L Gnesdilow, University of Wisconsin Madison, gnesdilow@wisc.edu

Sadhana Puntambekar, University of Wisconsin Madison, puntambekar@education.wisc.edu

Ali Mazalek, Toronto Metropolitan University, mazalek@torontomu.ca

Mike Tissenbaum, University of Illinois at Urbana Champaign, miketiss@illinois.edu

**Abstract:** This study explored the Idea Wall, a collaborative knowledge-building tool to support students' collaboration in small groups during a plant biology science curriculum. We examined the affordances and challenges of the Idea Wall and found the effective use of the tool's spatial organization capabilities by students, particularly the Yup Zone and the intermediary neutral spaces, for collaboratively organizing notes. But there's also a need for improvements in some features of the tool's design and instructional guidance.

### Introduction

Collaborative learning, emphasizing collective knowledge building through the exchange, negotiation, and expansion of ideas among participants, establishes a foundation for communal learning within the classroom (Tissenbaum & Slotta, 2019). This approach has been shown to foster deeper understanding and the development of shared expertise (Ong et al., 2020). Within technology-rich learning spaces, previous research has indicated the value of collaborative learning in encompassing a variety of modalities to enhance understanding and engagement (Dennen & Hoadley, 2013). However, it is important to carefully design and take into consideration the affordances of the features of the tool (Jeong & Hmelo-Silver, 2016). Understanding the relationship between modalities and affordances will help us create better collaborative learning environments that align with the learners' needs. To this end, this work builds on multiple prior research on the Idea Wall (described below), to answer the following research questions: 1) How does the current design of the Idea Wall enable students' small group face-to-face collaboration and collaborative knowledge building within a plant biology curriculum? 2) What improvements can be made to the Idea Wall to optimize collaborative interactions?

### Methods

The SimSnap learning environment was developed around three central technologies: Idea Wall, a digital plant simulation, and a digital notebook. For this study, our analysis only examines the small-group collaboration during Idea Wall activities. The Idea Wall is a collaborative spatial manipulative note tool that aims to facilitate collaborative knowledge co-construction, by allowing students to add ideas (in the form of notes) to a shared canvas, where they could spatially move them around to support the organization, negotiation, and build on of ideas. The canvas has distinct color-coded zones to enable collaborative negotiation: agreement (Yup - green), disagreement (No - red), and synthesis (Combine Zone - brown).

Participants were forty-three grade 8 students from two classes ( $n_1 = 24$ ;  $n_2 = 19$ ) and a teacher from a STEAM magnet school. After the study, 17 students who consented participated in a focus group interview.

There were five collaborative Idea Wall activities over 6 days, focused on the growing healthy plants. Each Idea Wall activity had 3-4 students per group. We screen-recorded the interactions of 4 groups (2 in each class) for 4 Idea Wall activities (IW2-IW4, 16 Idea Wall sessions total). Post-intervention, we conducted focus group interviews with three groups of students ( $n = 17$ ). Two researchers watched all recordings and interviews and collectively took notes, coded, and co-analyzed the data. We also examined all 4 groups final Idea Walls screenshots, and examined the collaborative activities of one group (1A-main-sunflower) and its jigsaw variation for IW4 (1a-jigsaw-sunflower). Coding Scheme: For the *sunflower* groups, we coded all their interactions within the Idea Wall, as well as their verbal interactions using a coding scheme adapted from Tissenbaum et al. (2017), resulting in the following five codes: 1) Making suggestions verbally about science content; 2) Making suggestions verbally about tool use; 3) Clarification; 4) Narration; 5) Maintaining joint attention.

### Findings



Across all activities each group submitted at least 2 notes in the Yup zone, with one group submitting 11 notes for IW2. An examination of the notes that did not make it to the Yup Zone revealed that two groups during IW2 placed notes in the No zone, and in each case, it was because they were off topic. When groups used the Combine Zone, these activities were classified as two distinct types: *Organizing* (tidying and re-organizing their notes); and *Rise Above*, which are ones where "students generate a deeper formulation of their understanding of the topic such as synthesizing key ideas together" (Ong et. al, 2020). Most of the students used the Combine Zone for *Organizing*, with many of those (7 of 16 groups) consisting of students cleaning up identical notes by combining them together.

In the interviews, one student stated that the Combine Zone helped them remove duplicate notes rather than cluttering up the Idea Wall. Students also used the Combine Zone to edit mistakes/mistyped ideas. Most students stated that they did not use the No zone, instead, just not putting the notes in the Yep zone.

1a-jigsaw-sunflower Idea Wall 4: There were 3 notes on the screen (one a combined note), and all got submitted to the Yup zone, and were all coded as relevant to the question. The No zone was not used by the group. In the Idea Wall, students seldom talked, and the talking that did occur was not about the collaborative notes.

1A-main-sunflower Idea Wall 5: In IW5, students wrote 7 notes in total but only submitted 2. They had 3 notes in the neutral zone and 2 notes in the Combined zone unsubmitted. No student interacted with or placed another student's notes into a zone. Despite limited interactions with other's notes, there were multiple discourse events that related to the notes. We did not find any instances in which students wrote about the content related to other students' discourse, which would have indicated a deeper level of synthesis. The students tended to focus only on writing down their findings from the previous group instead of talking about the differences.

## Discussion

The Idea Wall helped with the students' group coordination, enabling them to look at the prompt questions and the notes other members had written. In purely verbal collaboration, losing track of the conversation can make it difficult to reengage. In terms of spatially organizing notes, students understood and used the Yup Zone with no issues; however, some groups left valuable notes in the neutral zone and the Combine Zone. We were not sure if students understood that they could only submit the notes in the Yup Zone, in part, because in the current version, the submission button was placed at the bottom Idea Wall, which may have implied to the students that they were submitting the whole Idea Wall. Student engagement with the Idea Wall's No Zone was minimal and often off-topic. Focus groups revealed groups bypassed the No Zone, opting not to place notes in the Yup Zone instead. Some students repurposed the No Zone for deletion due to a lack of an actual delete feature. To this end, we anticipate future designs of the Idea Wall Yup Zone and No Zone to change in one of two ways: 1) require students to arrange all notes into either Yup or No Zone before they can submit, to induce more collaboration; or 2) remove the No Zone while adjusting the placement of the submission into the Yup Zone to make the neutral zone the de facto no zone. For the Combine Zone, students mainly used it to organize their notes as we did not provide deletion or editing functions. There is potential to improve students' ability to synthesize and create Rise Above notes. From our analysis, we believe more direct scaffolding on how to use the Combine Zone could support this.

Our analysis also showed that the Idea Wall could improve, facilitate, and support students' science knowledge building. Of particular interest were times when students submitted notes from previous Idea Wall discussions and tried to modify and discuss them. This points to the ability of the Idea Wall to support synthesizing ideas across successive activities. Students' verbal interactions were different for each group, with task-related joint attention verbal communications tending to lead to valuable insight and suggestions about science content, leading to important collaborative notes. However, there is a need to provide better scaffolding on how to effectively synthesize each other's ideas and engage in collaborative knowledge building.

## References

- Dennen, V. P., & Hoadley, C. (2013). Designing collaborative learning through computer support. The international handbook of collaborative learning, 389-402.
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?. *Edu Psychologist*, 51(2), 247-265.
- Ong, A., Lee Teo, C., Tan, S., & Song Kim, M. (2021). A knowledge building approach to primary science collaborative inquiry supported by learning analytics. *Education 3-13*, 49(3), 371-386.
- Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning*, 12, 35-64.
- Tissenbaum, M., & Slotta, J. D. (2019). Developing a smart classroom infrastructure to support real-time student collaboration and inquiry: a 4-year design study. *Instructional Science*, 47, 423-462.

## Tensions in Designing Technological Interactives in a Justice-Centered Middle School Biology Curricular Unit

Sugat Dabholkar, Tufts University, sugat.dabholkar@tufts.edu  
Ravit Golan Duncan, Rutgers University, rgduncan@gse.rutgers.edu  
Frieda Reichsman, Concord Consortium, freichsman@concord.org  
Rishi Krishnamoorthy, Pennsylvania State University, rkrish@psu.edu  
Edna Tan, University of North Carolina – Greensboro, e\_tan@uncg.edu

Burrell Smithen, Jaya Joshi  
bas321@scarletmail.rutgers.edu, jj785@scarletmail.rutgers.edu  
Rutgers University

**Abstract:** To design computer-supported collaborative learning environments that center equity, it is important to ensure that the integrated technological interactives are in alignment with the equity-related design considerations. We study the design of technological interactives in a justice-centered middle school biology curricular unit about stress. We present discuss tensions that surfaced as we attempted to foreground student agency, ethics of manipulating human systems even virtually, scientific accuracy of the represented phenomenon, and learnability using technological interactives.

### Introduction

In response to Uttamchandani, Bhimdiwala, and Hmelo-Silver's (2020) call to highlight CSCL's strengths and their potential to support equitable educational change by focusing on learner agency, shared meaning-making, and the disruption of social inequalities, in this poster we present a study about the design of technological interactives to center equity and support learner agency in collective knowledge building in classrooms. We conceptualize equity from a justice-orientation to support political struggles for making present the lives of those made missing by schooling and discipline-specific norms (Calabrese Barton and Tan, 2020) and processes of learning that empower learners to become authors of their own futures (Gutiérrez and Jurow, 2016).

Intentionally designed technological interactives embedded in learning environments have the potential to support critical conversations about topics such as race, and gender that are typically silenced in K-12 learning settings (e.g., Cortez, McKoy, & Lizárraga, 2022). We designed technological interactives in Bio4Community *Stressed Out!* unit, a justice-oriented middle school biology curricular unit about stress (Duncan et al., 2023). The curricular unit makes local and global injustices related to race, gender, class, and immigration visible through story-driven investigations and community ethnography. The unit highlights how systemic oppressions and the ensuing injustices create long-term stress and impact biological systems, and it engages students in proposing systemic solutions to address local injustices in their lives. The technological interactives were simulations of biological systems that students investigated to learn about the impact of chronic stress on biological systems.

### Research context and methods

This work is part of a larger grant-funded project to develop a middle-school biology unit about *stress* that we identified as a key health concern for the youth and their community (Duncan et al., 2023). In the Winter of 2023, we worked with two science teachers who co-taught the curricular unit in two middle school classrooms. The school is located in a sanctuary city in the NE-USA where 90.2% of the students in the district identify as Hispanic. To study the use of technological interactives, we collected seven screen recordings (4 to 12 minutes) of focal student groups (3 groups, 8 students) when they used technological interactives and screen recorded (18 to 20 minutes) student interviews about the use of interactives after they participated in the unit (5 groups, 13 students in total). We conducted a qualitative analysis of the screen recordings to identify use of these interactives in the context of the curriculum. The analysis involved viewing the videos by two researchers and identifying episodes that included successes (30 episodes) and concerns (20 episodes). We grouped the concerns four themes that we see as being potentially relevant for designing interactives in biology curricula.

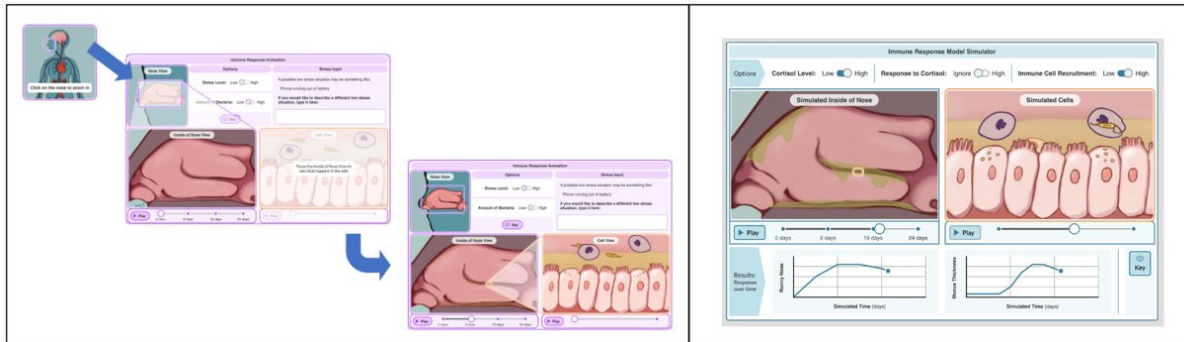
### Findings and discussion

The four categories of concern areas are as follows: 1) Connecting cellular level changes to whole person level effects, 2) Making comparisons between various states of the systems, 3) Understanding long-term regulation

issues with response to stress hormones, 4) Making sense of interlinked representations (Figure 1). We believe that these concerns arose because of two main tensions that were exacerbated given our social justice goals.

**Figure 1**

The animation sequence and the simulation in the Nose interactive.



*Student agency vs ethics of human research:* On one hand, in our designed we wanted to give more agency to students to investigate systems by manipulating various parameters but on the other hand, we did not want to do things that would violate axiological considerations of ethical research by supporting to take harmful actions such as increasing stress for a person even metaphorically using technology (Figure 1). There are troubling pasts and presents of manipulations of human bodies for research purposes (e.g., Washington, 2019)

*Scientific accuracy vs learnability - space and time:* To understand how long-term stress affects biological systems, we need to think about different scales of time and space. Plaque buildup because of a stressful lifestyle can take decades, whereas to observe the cellular processes involved in plaque formation one needs to have a time scale of milliseconds and visualize a space in micrometers. Because of scientific accuracy regarding time and space, we chose to create two views which were linked spatially and temporally. This design decision resulted in requiring students to make sense of these interlinked representations and make multiple comparisons at different time points and at different scales (tissue level and cell level).

The tensions that we discussed have implications for designing justice-oriented biology curricular units. It surfaces two important open questions for all biology curricular developers: How can we design technological interactives that allow whole person level understanding without allowing manipulations of body systems and social realities of the person? How can we be scientifically accurate about time and space scales and make visualizations easy to investigate and learn from?

## References

- Calabrese Barton, A., & Tan, E. (2020). Beyond Equity as Inclusion: A Framework of “Rightful Presence” for Guiding Justice-Oriented Studies in Teaching and Learning. *Educational Researcher*, 49(6), 433–440.
- Cortez, A., McKoy, A., & Lizárraga, J. R. (2022). The Future of Young Blacktivism: Aesthetics and Practices of Speculative Activism in Video Game Play. *Journal of futures studies*, 26(3).
- Duncan, R. G., Krishnamoorthy, R., Tan, E., Austin, T., Reichsman, F., Dabholkar, S., Worsley, T., Islam, A., Smithen, B., Joshi, J., & Asada, M. (2023). Designing science curricula that disrupt disciplinary boundaries towards sociopolitical change: A middle school life science unit. In Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), *Proceedings of the 17th International Conference of the Learning Sciences - ICLS 2023* (pp. 1138-1141). International Society of the Learning Sciences.
- Gutiérrez, K. D., & Jurow, A. S. (2016). Social design experiments: Toward equity by design. *Journal of the Learning Sciences*, 25(4), 565-598.
- Uttamchandani, S., Bhimdiwala, A., & Hmelo-Silver, C. E. (2020). Finding a place for equity in CSCL: Ambitious learning practices as a lever for sustained educational change. *International Journal of Computer-Supported Collaborative Learning*, 15, 373-382.
- Washington, H. A. (2019). *Medical apartheid*. Random House.

## Acknowledgments

This research is supported by the National Science Foundation under Grant No. 2000515. Any opinions, findings, or recommendations expressed are of the author(s) and do not necessarily reflect the views of NSF.

## Interaction Patterns and Structures in the R/Personalfinance Weekend Help and Victory Thread

Jose Sandoval-Llanos, Priya Sharma  
jfs@psu.edu, psharma@psu.edu  
The Pennsylvania State University

**Abstract:** This research explores the patterns and structures of sequenced interactions in the Weekend Help and Victory Thread (WHVT) of the r/personalfinance subreddit on Reddit. Utilizing descriptive and qualitative analysis, we investigated the roles of participants and their interaction patterns. Preliminary findings reveal limited collective knowledge construction, with a focus on individual learning. Noteworthy patterns include high dependence of top contributors to generate productive conversations and predominance of cyclical interactions. Ongoing analysis aims to uncover additional insights and patterns, contributing to our understanding of learning dynamics in online communities.

### Background

Social media platforms are popular sites for informal learning (Carpenter et al., 2022) providing learners with flexibility in when, where, and from whom to learn (Albers et al., 2016). These spaces have also been characterized as supporting collaborative knowledge construction where groups of individuals take up and build on each other's ideas to contribute to shared knowledge (Kimmerle et al., 2015). Our research seeks to understand how interaction patterns and sequences may lead to individual learning and collective knowledge construction in social media platforms, specifically focusing on the characteristics that trigger productive and engaging conversations. Interaction patterns and the role of initiators in online communities have been examined previously by Heo and Breuleux (2009), who identified four types of interaction patterns: cyclical, branched, chained, and complex. In this paper, we follow up on this classification of interaction patterns and we analyze the nature and impact of interactions between active participants of the Weekend Help and Victory Thread (WHVT) within the subreddit r/personalfinance. The research questions guiding this research are: What is the structure of participation in the WHVT? How do interactions facilitate individual learning and collective knowledge construction in the WHVT?

### Data collection and analysis

This study builds on ongoing work on the r/personalfinance subreddit on Reddit. The WHVT is an open forum where redditors post their questions, findings, or experiences related to personal finance, especially those who seek quick responses to their questions. Posts are organized in threads with an initial post and subsequent responses. Using online ethnography techniques (Bryman, 2012), we observed the WHVT during 2022. and selected 3 weeks with the highest number of posts for analysis, comprising 502 posts from 194 authors in 158 threads. Posts were extracted using Octoparse and analyzed in Excel to generate descriptive analysis at post and thread levels. We scrutinized these threads in detail to study how discussions unfolded. Elaborating on the findings presented by Heo and Breuleux (2009), we examined the roles of initiators and top contributors, and their interaction patterns.

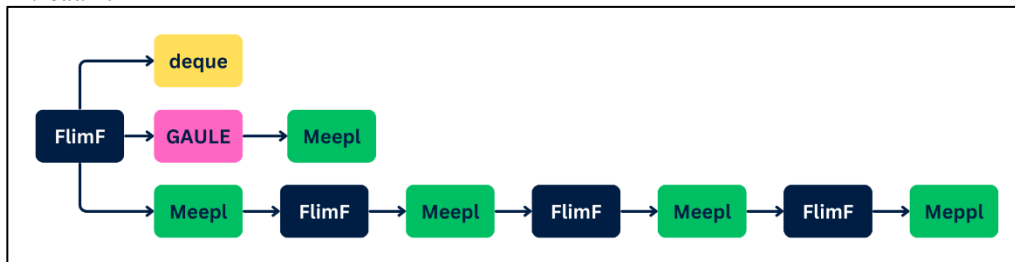
### Preliminary findings

Initiators started the interactions, but participation of top contributors was crucial to sustaining interactions. We analyzed the top ten participants with the highest number of posts during the three weeks. They accounted for a third of the total posts in this sample. While none initiated an interaction, they were responsible for two thirds (114 out of 190) of the first response to original posts, usually giving their opinion, sharing information, asking for more information from the initiators, or offering new perspectives to be considered. Conversely, more than half of the initiators (70 out of 135) failed to follow up on their conversations. Half of those who returned, followed up only once, usually in response to the initial response received.

When we analyzed the 158 threads, 19 had no response, 58 only had one response, and in 50 of them three or more participants interacted with the initial poster. We used Heo and Breuleux's (2009) categorization of interactions to further explore the 50 threads. We found that the most common pattern was cyclical interaction (exchange of knowledge between the initiator and an expert only) followed by branched (when different experts post separate responses to an initiator) and complex interactions (when responses show a combination of two or

more patterns). The least common type of interaction was the chained interaction where two or more experts build on each other's ideas with little or no presence of the initiator. Thread 29 (See Figure 1) evidenced all these types of interactions. *FlimF* was the initiator and deque, *GAULE*, and *Meepl* opened three branches posting individual non-related responses, although the branch opened by deque didn't have any response. Second, a chained interaction occurred when *Meepl* followed up on the branch opened by *GAULE*. The branch opened by *Meepl* showed a cyclical interaction pattern between *FlimF* and *Meepl*. We can characterize this thread as complex because of multiple interaction patterns. In our analysis of the WHVT, we found a predominance of cyclical interaction patterns between experts and initiators with few chained iterations between multiple expert respondents and initiators.

**Figure 1**  
Thread 29



## Discussion

We delved into the interactions between active participants within the WHVT in the *r/finance* subreddit to explore how participant interaction unfolds in this space. We observed some noteworthy features, such as the high dependence on the group of top contributors and the presence of cyclical interaction patterns. Interaction patterns suggest less focus on contributing to a shared knowledge base and an immediate interest in receiving answers to questions. While the involvement in the WHVT can be considered as a way of social learning, indicating a willingness to learn from and alongside others (Haythornthwaite et al., 2018), we have not observed a collective knowledge construction process, but rather an emphasis on individual learning (Kimmerle et al., 2015). This finding nuances the type of “learning from others” that may occur on social media platforms. We intend to further analyze lengthy threads to identify emerging new insights and patterns.

## References

- Albers, P., Pace, C. L., & Odo, D. M. (2016). From affinity and beyond: A study of online literacy conversations and communities. *Journal of Literacy Research, 48*(2), 221–250. <https://doi.org/10.1177/1086296X16659069>
- Bryman, A. (2012). *Social Research Methods* (Fourth ed.). Oxford University Press.
- Carpenter, J., Tani, T., Morrison, S., & Keane, J. (2022). Exploring the landscape of educator professional activity on Twitter: an analysis of 16 education-related Twitter hashtags. *Professional Development in Education, 48*(5), 784–805. <https://doi.org/10.1080/19415257.2020.1752287>
- Duan, J. J., Lu, L., & Xie, K. (2023). Examining knowledge construction in three social interactive learning environments: a comparison of knowledge networks, social networks, and social knowledge networks. *Interactive Learning Environments, 31*(6), 3914–3938. <https://doi.org/10.1080/10494820.2021.1944882>
- Haythornthwaite, C., Kumar, P., Gruzd, A., Gilbert, S., Esteve del Valle, M., & Paulin, D. (2018). Learning in the wild: coding for learning and practice on Reddit. *Learning, Media and Technology, 43*(3), 219–235. <https://doi.org/10.1080/17439884.2018.1498356>
- Heo, G. M., & Breuleux, A. (2009). Roles of initiators and interaction patterns: Exploring an informal online community at the interpersonal plane. *Computer Supported Collaborative Learning Practices, CSCL 2009 Conference Proceedings - 9th International Conference*, 13–17.
- Kimmerle, J., Moskaliuk, J., Oeberst, A., & Cress, U. (2015). Learning and Collective Knowledge Construction With Social Media: A Process-Oriented Perspective. *Educational Psychologist, 50*(2), 120–137. <https://doi.org/10.1080/00461520.2015.1036273>

# Initial Examination of Identity Development When Starting in an Online, Global, Collaborative, Learning Environment

Danielle P.Espino, Pepperdine University, danielle.espino@pepperdine.edu  
 Nastassia Jones, Southern University, nastassia.jones@sus.edu  
 Kristina Lux, Pepperdine University, kristina.lux@pepperdine.edu  
 Lu Liang, Pepperdine University, lu.liang@pepperdine.edu  
 Leah-Maria Gaskin, Southern University, leahmarie.gaskin@sus.edu  
 Eric Hamilton, Pepperdine University, eric.hamilton@pepperdine.edu  
 Seung Lee, Pepperdine University, seung.lee@pepperdine.edu

**Abstract:** This poster examines identity thinking from individual interviews of students as they begin participating in an online, collaborative, global learning environment. Transcripts of nine adolescent students in the US and Kenya of interviews taken in March 2023 and then September 2023 were examined using Epistemic Network Analysis to identify any changes in thinking amongst the students. Results indicated in March 2023, students indicated challenges related to their personality and importance of social interactions, then shifted to identifying challenges with STEM topics and being more collaborative in September 2023.

## Introduction

In the process of learning, practicing, experiencing, and reflecting, students develop a sense of belonging, recognition, and validation, constructing their identity in a social context (Vygotsky & Cole, 1978). Through collaborative engagement and leadership development activities, students expand their identity awareness and skills for academic and career success (Porfirio et al, 2023). Coleman & Davis (2020) suggest that asset-based pedagogy effectively increases STEM-based efficacy in technology and facilitates learning engagement and motivation. Flint & Jagers (2021) found asset-based teaching and learning not only increase learners' "sense of belonging, agency, and well-being" (p.260) but also raise students' level of excitement working with technology.

The Asset Based Learning Environments (ABLE) project aims to engage learners in an informal, online, international, project-based science, technology, engineering, and mathematics (STEM) learning and media making community. Underpinned by an asset-based approach, global synchronous online sessions between adolescents from multiple countries take place weekly during the school year. Prior to attending these sessions, participants engage in reflective pre-assessment interviews with planned post-assessment interviews a year later.

## Methods

The data analyzed in this study consists of individual assessment interviews taken of nine adolescent students from the US and Kenya in March 2023 and again in September 2023. Each interview was transcribed and reviewed separately by two researchers and then coded independently by two raters who followed a process of social moderation to reach final consensus on the coded data (Herrenkohl & Cornelius, 2013). The data was uploaded to the Epistemic Network Analysis web tool to create a visualization to support the analysis. Each utterance, separated by turn of talk between the interviewer and the student, represented the unit of analysis, and each interview defined as the conversations in which connections were limited with a window size of 4.

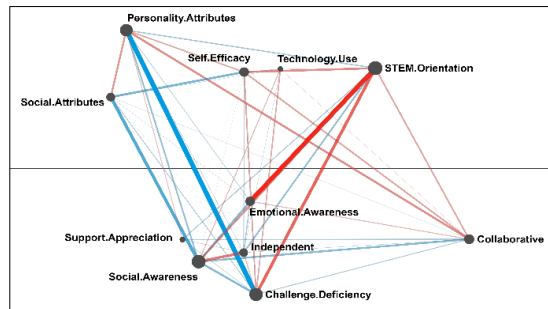
**Table 1**  
*Codebook of Constructs*

Construct	Definition
Emotional Awareness	Aware of emotional responses to situations
Social Awareness	Ability or inability to interact with each other
Technology Use	Actual use of technology tools within the project
STEM Orientation	Focus on science, technology, engineering, math as subjects
Collaborative	Narrower perspective on the act or orientation of collaboration
Independent	Relying on self rather than others in how to accomplish something
Personality Attributes	Thoughts and feelings as an individual person, personality and traits
Social Attributes	Identifying as part of a particular social construct grouping
Self-efficacy	Recognizing self as exceptionally confident or strong in something
Challenge Deficiency	Recognizing one is not good at or incapable of something
Support Appreciation	Acknowledging general help or support from a specific individual

## Results

The subtracted ENA network models examining discourse patterns from the nine student interviews conducted in March 2023 and September 2023 are seen in Figure 1. In March 2023, students had stronger connections between Personality Attributes and Challenge Deficiency, and also Social Attributes and Social Awareness. In September 2023 was a shift: connections between Collaborative and Personality Attributes and STEM Orientation were greater, and also between STEM Orientation and Emotional Awareness and Challenge Deficiency.

**Figure 1**  
*Subtracted Network of Student Interviews from March 2023 (blue) and September 2023 (red)*



## Discussion

The examination of the nine interviews gives some initial insight into identity thinking shifts from March 2023 to September 2023 after just a few months of involvement in the project community. In the March 2023 interviews, as seen by the strong connection between Personality Attributes and Challenge Deficiency, students focused on identifying personality attributes as a challenge. For example, a Kenyan student reflected, “*I get shy sometimes and then I forget most of the things I’m supposed to see...I get nervous.*” Students also emphasized a strong connection between Social Awareness and Social Attributes, sharing a lot about the importance of relationships in social groups, as reflected by another Kenyan student: “*If you’re my classmate, for example, and you come for me for help...I’ll help you...It’s actually really important because this together... Togetherness and unity.*”

In contrast, the September 2023 interviews have more emphasis on STEM Orientation, especially Emotional Awareness and Challenge Deficiency, indicating a more openness to discussing their struggles with STEM topics, as well as how it makes them feel. For example, one US student was able to reflect on their challenge explicitly, in a way that was not previously clear in March 2023: “*I might not be that strong in quickly doing things in my brain. Like math, sometimes I have to write it down.*” There was also a shift towards a more collaborative orientation, as reflected by a US student: “*I like doing, like hands on things. Like last year, my friends and I did a maglev project where we made magnetic levitation cars. And so I like doing things like that.*”

This shift from students reflecting on challenges related to their personality and generally talking about the importance of interacting with others in March 2023 to being more open on their challenges related to STEM as well as becoming more collaborative in their orientation in September 2023 took place in an unexpectedly short amount of time. With formal post-assessment interviews at the end of the US academic year, these initial results are able to use ENA to show progression shifts in thinking from just a few months of project involvement.

## References

- Coleman, S.T., & Davis, J. (2020). Using Asset-Based Pedagogy to Facilitate STEM Learning, Engagement, and Motivation for Black Middle School Boys. *Journal of African American Males in Education*, 11(2), 76–94.
- Flint, A. S., & Jagers, W. (2021). You matter here: The impact of asset-based pedagogies on learning. *Theory Into Practice*, 60(3), 254–264.
- Herrenkohl, L. R., & Cornelius, L. (2013). Investigating elementary students’ scientific and historical argumentation. *Journal of the Learning Sciences*, 22(3), 413-461.
- Porfirio, L., Chavarin, A., Linzy, M., Lopez, H., & Montiel, A. (2023). Students Take the Lead: Promoting Diversity and Inclusion in Adult Education through Ambassador Training. *COABE Journal: The Resource for Adult Education*, 12(1), 52–60.
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

## Acknowledgments

The authors gratefully acknowledge funding support from the US National Science Foundation (#2215613) for the work this paper reports. Views appearing in this paper do not reflect those of the funding agency.

## Making Space for Critical Action: Re-visioning Computational Thinking

Preeti Raman, Toronto Metropolitan University, praman@torontomu.ca  
Rubaina Khan, University of Toronto, rubaina.khan@mail.utoronto.ca  
Meera Sharma, University of Toronto, mrinalini.sharma@mail.utoronto.ca  
Renato Carvalho, University of Toronto, renato.carvalho@utoronto.ca  
Chandan Dasgupta, University of Twente, c.dasgupta@utwente.nl  
Aakriti Mahajan, University of Toronto, aakriti.mahajan@mail.utoronto.ca  
James Slotta, University of Toronto, jim.slotta@utoronto.ca

**Abstract:** While school makerspaces promise to inspire and excite, the challenge of meaningfully integrating them into schools remains. Guided by a philosophy of praxis that stresses the need for education to interweave theory, action, and reflection to advance positive social change in our communities (Freire, 1970), this paper reports on the co-design of a school space called the Critical Action Learning Lab (CALL) for inclusive making to support computational thinking and critical action through curriculum-informed learning.

### Introduction

Schools around the world are racing to invest in makerspaces as a way to support design thinking and 21st century competencies. However, they seldom promote the deep forms of learning envisioned by makerspace advocates (e.g., Halverson & Peppler, 2018), wherein students explore domain content while engaging in critical inquiry, collaboration, design and computation. In the developing South, schools are experiencing new forms of cultural imperialism as they attempt to emulate classic western designs, with an even poorer fit to their curriculum and cultures. The challenge remains, to democratize makerspaces and enable a generation of maker curricula by schools (Blikstein, 2018). One approach that has gathered attention is that of critical making, in which students are given voice and classrooms are democratized through empowering forms of design, storytelling and fabrication. Such activities can support students' development of identity, and serve as magnets into STEM and related disciplines for marginalized students.

This paper reports on our design of a Critical Action Learning Lab (CALL) for schools in Bengaluru, India, working in close collaboration with teachers and school leaders to create critical making lessons that fit within the school culture and curriculum. These lessons, in turn, help determine the form and function of an inclusive space in which students participate in making, unmaking, remaking and reimagining in various forms and materials – including historic and culturally relevant means – focusing on what is important to them. This paper will present the design process for the CALL space, curriculum, and teacher community, with a particular focus on the intersection of computational thinking with critical action curriculum design using the Critical Action Learning Exchange (CALE) framework (Carvalho et al., 2021). We adopt a critical framing of Computational Thinking, where learning occurs at the intersection of emerging understandings of how realities are shaped and what strategies/actions are needed to amplify the voices of the marginalized and oppressed (Kafai & Proctor, 2021). Engaging in design-based participatory research, we explored various affordances and functions of a Critical Action Learning Lab (CALL) for middle and high school students, focussing our inquiry on the following research questions: How can we design a culturally relevant making space for critical action learning? In what ways does computational thinking intersect with critical pedagogy and action regarding global issues within a local cultural context?

### Background and theoretical framework

Computational Thinking (CT) is often understood as a particular form of problem solving and reasoning in which the learner addresses open-ended problems to formulate the problem in such a way that its solutions can be represented as algorithms that can be worked through by Computer Scientists (Slotta et al., 2020). CT develops competencies that will serve students in learning across disciplines and throughout their lives. For this project, we envisage that our teachers engage in computational thinking in pursuit of critical action. In our context, opportunities for such engagement are created while the teachers focus on collectively designing a school makerspace that is sensitive to the available school resources. The primary purpose of such a space was affording critical action through making.

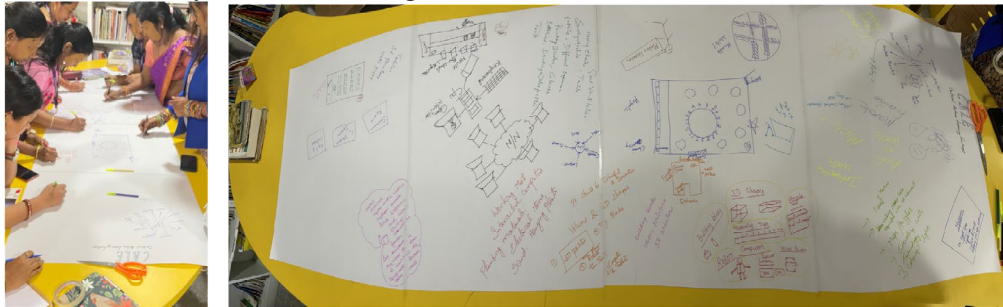


## Methods

Data for this paper includes pictures and drawings of the visualized space, audio and video transcription of the sessions, design artifacts created during the online and in-person sessions including curricula, and field notes. We engaged in both online and in-person PCD sessions with the teachers (Figure 1) to understand existing practices and spaces for learning within the teachers' cultural context. Following a few online sessions on Zoom, three researchers from our team visited the school to engage in four in-person co-design sessions over four days with educators.

### Figure

*Collaborative Artifact Creation during PCD In-Person Session 2*



## Initial analysis and findings

Our reflexive thematic analysis of data from audio and video recordings from the workshops involved pattern recognition within the data and the identification of themes through “careful listening/watching and re-listening of the data (Guest et al., 2021). We developed an apriori codebook based on research questions. The codes that are the focus of the first iteration are: “experiences that shape learning”, “criticality in curriculum”, “CT skills” and “recommendations for makerspaces”. In subsequent iterations of coding, we allowed for additional codes to emerge. These include: “21<sup>st</sup> century competencies,” “knowledge and skills” “culture and inclusivity” “STEM pathways” and “experiences in the space”. Initial analysis suggests an alignment of certain CT processes with particular elements of critical action, which can help inform our understanding of how CT occurs within critical action learning.

## Conclusion

In recent years, many learning scientists have called upon our community to re-vision school makerspaces: Blikstein (2018, 2020) calls for local generation of culturally-relevant maker curricula by schools; Kafai (2021) encourages us to find practices and possible trajectories of how students can deepen their CT skills. Through this work, we begin to tackle the complex challenge of designing a “making” space for critical action, aiming to empower students and teachers to engage in meaningful making and transformative action to drive positive change in their communities.

## References

- Blikstein, P. (2018). Maker movement in education: History and prospects. *Handbook of Technology Education*, 419, 437.
- Blikstein, P. (2020). 10 CHEESEMAKING EMANCIPATION. *Designing constructionist futures: The art, theory, and practice of learning designs*, 115.
- Carvalho, R., Raman, P., Boldyreva, E., Ndubuisi, A., Burron, G., Zhang, X., Ghasempour, E., Slotta, J. (2021). Designing a Global Community of Critical Action Educators. Proceedings of the International Society of the Learning Sciences (ISLS) Annual Meeting 2021, Bochum, Germany: ISLS
- Freire, P. (1970). *Pedagogy of the oppressed*. Continuum.
- Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. sage publications.
- Halverson, E., & Peppler, K. (2018). The maker movement and learning. In *International handbook of the learning sciences* (pp. 285–294). Routledge.
- Kafai, Y. B., & Proctor, C. (2022). A reevaluation of computational thinking in K–12 education: Moving toward computational literacies. *Educational Researcher*, 51(2), 1(pp. 46-151).
- Slotta, J.D., Chao J. & Tissenbaum, M. (2020). Fostering Computational Thinking and Design Thinking in the IB. Report prepared for the International Baccalaureate.

# Fine-Tuning Large Language Models for Data Augmentation to Detect At-Risk Students in Online Learning Communities

Hongming Li, Anthony F. Botelho  
hli3@ufl.edu, abotelho@coe.ufl.edu  
University of Florida

**Abstract:** We introduce a working approach that combines the method of fine-tuning large language models (LLMs) to create augmented data for the regression predictive models aimed at detecting at-risk students in online learning communities. This approach has the potential to leverage scarce data to improve urgency detection, and it can also present the role of artificial intelligence in enhancing the resilience of educational communities and ensuring timely interventions within online learning settings.

## Introduction

Online learning communities, epitomized by Massive Open Online Courses (MOOCs), have become pivotal in fostering accessible education. Yet, identifying and supporting at-risk students in these environments remains a complex challenge (Romero et al., 2010; Marbouti et al., 2016). Prior research has applied machine learning to predict student performance and dropout in MOOCs using features like forum posts, assignment submissions, and quiz scores (Kloft et al., 2014). The use of text as a data source for urgency detection in educational settings adds to this body of work and is related to sentiment analysis (Nasukawa & Yi, 2003).

Large Language Models (LLMs), such as GPT-3, have been extensively utilized for natural language processing tasks (Brown et al., 2020). However, their application in educational settings, particularly for the detection of at-risk students, is less explored. The use of LLMs for data augmentation, as proposed in this study, intersects with ongoing research into overcoming data scarcity in machine learning (Shorten et al., 2019). XGBoost, an advanced gradient-boosting machine learning algorithm, has been successfully applied in various domains, including educational data mining (Chen et al., 2013). In the context of resilience and responsiveness of educational communities, AI has played a role in adaptive systems that personalize the educational experience based on student needs (Pechenizkiy et al., 2009).

Building upon these foundations, our work extends the field by employing fine-tuned LLMs for data augmentation to enhance the predictive modeling of at-risk students. This innovative approach prompts the following research questions: *RQ1*: How can fine-tuned LLMs be utilized to generate synthetic data that improves the accuracy of urgency detection in MOOCs? *RQ2*: To what extent does training the XGBoost model with the augmented dataset generated by fine-tuned LLMs enhance the model's performance in identifying at-risk students compared to training with the original dataset?

## Methodology and findings

The cornerstone of our methodology is the Stanford MOOCPosts dataset (Stanford University, 2014), containing 29,604 anonymized posts from public online courses, with a subset of 3,505 entries being used for this study. These entries have been manually scored for urgency on a scale from 1 (least urgent) to 7 (most urgent) to reflect the need for immediate attention from the educational institution. We developed a model using the XGBoost regression algorithm and applied the following: *Preprocessing*: Utilize standardized NLP procedures to normalize, tokenize, remove stop words, and perform stemming on text data to ensure uniformity and reduce noise. *Feature Extraction*: TF-IDF Vectorization was applied to convert the processed text into a feature matrix, which captures the importance of terms within the documents. *Model Training*: The XGBoost model was trained on the original dataset, with data split into training (70%), validation (10%), and testing (20%) sets.

For fine-tuning, we used GPT-3.5turbo and considered the open-source LLaMA2 model as an alternative due to its cost-effectiveness for future applications. The fine-tuning process involves adjusting the models based on our urgency-scored dataset, using prompt engineering to guide the LLMs. We transform the initial data into a format suitable for LLM processing, as exemplified by the script provided, which restructures data into a JSON-lines (jsonl) format that defines the role and content for system-user-assistant interactions. This clarifies the task and expected output for the model, allowing it to generate posts that mimic the style and urgency of the students' original content. After the initial training, additional synthetic data is created using the fine-tuned LLM to simulate the distribution of urgency in student posts, doubling the size of data. The models' performance was evaluated using the test set (20%) from the original dataset to establish a baseline.

The analysis of our model's performance, as detailed in Table 1, indicates a modest enhancement in predictive accuracy following the introduction of augmented data generated through fine-tuned large language models. While the decrease in Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for the augmented dataset model signals a positive trend, the improvements are small.

This incremental progress underscores the nuanced nature of urgency detection within educational texts and suggests that further refinement of the data generation process and predictive modeling techniques is necessary. The modest gains prompt more profound consideration of the complexities in interpreting student discourse and point to additional layers of sophistication in our Fine-tuned LLM.

**Table 1**

*Model Performance Before and After Using Fine-Tuned LLM Generation for Augmented Data*

	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error
Original Dataset	1.370	1.170	0.911
Augmented Dataset	1.299	1.140	0.877

This study introduces an innovative AI-based approach for identifying at-risk students in MOOCs by fine-tuning LLMs to generate synthetic data for enhancing predictive models. The preliminary XGBoost model, trained on a dataset with human-annotated urgency scores, provides insights for more sophisticated AI applications in online learning. While initial results are promising, as evidenced by modest improvements in MSE, RMSE, and MAE (Table 1), further refinement is needed. The incremental gains highlight the complexities of detecting urgency in educational discourse (Nasukawa & Yi, 2003) and align with research on overcoming data scarcity in machine learning (Shorten et al., 2019).

This research underscores the critical role of data quality and volume in training effective AI systems for educational applications. It also provides insights into the scalability of AI solutions, demonstrating how open-source models can be leveraged to achieve outcomes comparable to their resource-intensive counterparts. Our findings contribute to the ongoing discourse on the practical challenges and ethical implications of implementing AI in education, including cost, data privacy, and the use of AI-generated content. Should AI-augmented data prove effective, it would underscore the viability of LLMs in educational technology.

## References

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
- Chen, Y., Jia, Z., Mercola, D., & Xie, X. (2013). A Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index. *Computational and Mathematical Methods in Medicine*, 2013, e873595. <https://doi.org/10.1155/2013/873595>
- Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Proc. of the EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 60-65).
- Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. P. C. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1-15.
- Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing Favorability Using Natural Language Processing. *Proceedings of the International Conference on Knowledge Capture - K-CAP '03*. <https://doi.org/10.1145/945645.945658>
- Pechenizkiy, M., Trcka, N., Vasilyeva, E., & van der Aalst, W. (2009). Process mining online assessment data. In *Proc. of the 2nd Int'l Conference on EDM* (pp. 279-288).
- Romero, C., Ventura, S., & Garcia, E. (2010). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.
- Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2019). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular pharmaceutics*, 16(7), 2776-2790.
- Stanford University. (2014). The Stanford MOOCPosts Data Set. Stanford.edu. <https://datastage.stanford.edu/StanfordMoocPosts/#procedures>

## Acknowledgments

We would like to thank Dr. Jinnie Shin from the College of Education at the UF, NSF (e.g., 2331379, 1903304, 1822830, 1724889), as well as IES (R305B230007), and MathNet.

## Author Index

- Aakriti Mahajan, 439  
 Adéla Plechatá, 355  
 Afroza Sultana, 431  
 Ajaz Mesha, 427  
 Alex Acquah, 315  
 Ali Mazalek, 431  
 Alwyn Vwen Yen Lee, 241  
 Alyssa Wise, 331  
 Amanda Barany, 315  
 Amanda Thompson, 367  
 Amelie Vrijdags, 365  
 Andi Scarola, 315  
 Andrea Figueroa, 59  
 Andrea Vaughan, 343  
 Andreas Wildner, 395  
 Andria Gentry, 351  
 Angie Chen, 91  
 Anika Radkowsch, 403  
 Anna Weidenbacher, 221  
 Annabel Stoler, 315  
 Annelies Raes, 359, 365  
 Anthony F. Botelho, 441  
 Ariel Han, 233  
 Asare Philip, 427  
 Ashley Boone, 19  
 Askari Mehdi, 367  
 Ayano Ohsaki, 217  
 Bahar Shahrokhian, 301  
 Bas de Leng, 193  
 Batuhan Sayis, 163  
 Ben Zwickl, 181  
 Bernease Herman, 59  
 Bertrand Schneider, 177, 323  
 Blessed A. Mhungu, 421  
 Bodong Chen, 229, 331  
 Bookyung Shin, 269  
 Bradford Mott, 43, 185, 253  
 Burrell Smithen, 433  
 Cai-Ting Wen, 107  
 Caitlin Snyder, 107  
 Caleb Vatal, 107  
 Calixto Gutiérrez-Braojos, 413  
 Carlos Rodríguez-Domínguez, 413  
 Carol K. K. Chan, 209, 229, 385, 399  
 Carol M. Forsyth, 27  
 Cecilia Aragon, 59  
 Chandan Dasgupta, 193, 439  
 Changzhao Wang, 273  
 Charis Clevenger, 297  
 Chelsea Brown, 277  
 Chen Feng, 43, 253  
 Chen-Chung Liu, 107  
 Chencheng Le, 265  
 Chenyu Hou, 381  
 Chew Lee Teo, 229, 241  
 Chris Andrews, 331  
 Chris Palaguachi, 285  
 Chrissa Manoli, 193  
 Christian Cammarota, 181  
 Christian Schunn, 315  
 Christina Stiso, 315, 349  
 Christina Wekerle, 99  
 Chun Yen Chang-Sundin, 241  
 Chunlin Lei, 397  
 Cindy E. Hmelo-Silver, 43, 185, 253, 315, 349, 409  
 Clark A. Chinn, 315, 349, 377  
 Constanze Richters, 395, 403  
 Corey Schimpf, 367  
 Cresencia Fong, 423  
 Crina Damsa, 261  
 Cynthia D'Angelo, 285  
 Daeun Hong, 43, 185, 253  
 Dalila Dragnić-Cindrić, 409  
 Dan Zhu, 397  
 Dana Gnesdilow, 197, 431  
 Daniel Bodemer, 237  
 Daniel Hickey, 331  
 Daniel Sánchez, 225  
 Daniela Linda, 413  
 Danielle Murphy, 315, 349  
 Danielle P. Espino, 437  
 David Karger, 375  
 David Zhou, 301  
 Davinia Hernández-Leo, 163, 357, 373, 401, 417  
 Deliang Wang, 429  
 Destiny Williams-Dobosz, 189  
 Dick Ngambi, 421  
 Dina Soliman, 155  
 Edith Braun, 389  
 Edna Tan, 433  
 Edwin Chng, 177  
 Elizabeth Dyer, 285  
 Emily Kerzabi, 27  
 Emily Watts, 297  
 Emma Mercier, 411  
 Eowyn P. O'Dwyer, 377  
 Eric Hamilton, 437  
 Erum Afzal, 389  
 Eve Manz, 315  
 Evelyn Pollins, 343  
 Faiaz Rahman, 341  
 Fan Chen, 429  
 Fan Wu, 345  
 Fati Wu, 51  
 Feng Chen, 185  
 Fien Depaepe, 359, 365  
 Frank Fischer, 245, 395, 403



Frieda Reichsman, 433  
Fumihiko Mitsunaga, 371  
Gaowei Chen, 385, 425, 429  
Gaoxia Zhu, 265, 381  
Gautam Biswas, 107, 323  
Gaëlle Guigon, 387  
Grace Xing, 309  
Grant Chartrand, 331  
Guangji Yuan, 229, 241  
Guido Makransky, 355  
Guo Su, 75, 265  
Ha Nguyen, 3, 139  
Haesol Bae, 253  
Haiyang Xin, 345  
Hannie Gijlers, 193  
Hanxiang Du, 381  
Haowei Li, 407  
Harm Biemans, 115  
Heather Toomey Zimmerman, 281  
Hillary Swanson, 139  
Hiroki Oura, 371  
Hironori Egi, 383  
Hiroshi Kato, 371  
Hong Shui, 331  
Hongming Li, 441  
Htay Min Khaung, 377  
Huiying Cai, 393  
Héctor Florido, 417  
Indrani Dey, 297  
Ingo Kollar, 99, 213, 221  
Iris Saban, 353  
Ishan Sodhi, 91  
Ishari Amarasinghe, 163, 401  
Iza Marfisi-Schottman, 123  
J. Bryan Henderson, 315  
Jai Pandey, 423  
James D. Slotta, 427, 439  
James Lester, 43, 185, 253, 409  
Jaya Joshi, 433  
Jean Heutte, 365  
Jeffrey B. Bush, 277, 297  
Jenna Matthews, 139  
Jennifer Kahn, 273  
Jessica Andrews-Todd, 27, 205  
Jessica Roberts, 19  
Jessica Tan, 75  
Jiajun Wu, 147  
Jialiang LIU, 385  
Jianhua Zhao, 209, 385, 399  
Jianwei Zhang, 131, 241  
Jiaqi Xu, 363  
Jimmy Zambrano R., 361  
Jina Kang, 35  
Jing Wang, 83  
Jingyu Liu, 423  
Jinjun Xiong, 309  
Jinzhi Zhou, 315, 349  
Joey Huang, 293  
Johanna Ott, 221  
John Guzmán, 361  
Jonathan J. Foo, 405  
Jose Sandoval-Llanos, 435  
Joseph E. Michaelis, 289  
Joshua Danish, 315, 349, 391  
Joshua Katz, 411  
Juan Zheng, 381  
Julia Eberle, 419  
Julia Fecke, 389  
Julia Hornstein, 213, 221  
Jumana Almahmoud, 375  
Jun Oshima, 67  
Junjie Zhang, 397  
Justice Walker, 315  
Justin Hodgson, 331  
Kaiju Kangas, 415  
Kamali Sripathi, 375  
Kamila Misiejuk, 261  
Kana Matsueda, 67  
Kasiyah Junus, 369  
Katharina M. Bach, 395  
Kathy L. Schuh, 341  
Kati Sormunen, 415  
Khadija El Aadmi-Laamech, 357  
Kim MacKinnon, 423  
Kira Baker-Doyle, 343  
Krista Glazewski, 43, 185, 253, 409  
Kristina Lux, 437  
Kurt Vanlehn, 301  
Kylie Peppler, 233  
Laura Brandl, 403  
Lauren Albert, 349  
Leah-Maria Gaskin, 437  
Lenuta Alboaic, 389  
Leqi Li, 281  
Leslie Rupert Herrenkohl, 315  
Li-Jen Wang, 173  
Lia Sadita, 369  
Linda Smith, 367  
Linling Lan, 345  
Linmeng Lu, 393  
Linor L. Hadar, 353  
Lintang Matahari Hasani, 369  
Liru Hu, 147  
Lishan Zheng, 265  
Litong Zeng, 197, 431  
Lixiang Gao, 245  
Lu Liang, 437  
Lubna Hakami, 163  
Marc Beardsley, 401  
Marc Facciotti, 375  
Marcelo Worsley, 323  
Martin Bordewieck, 419  
Martin Greisel, 99, 213, 221  
Martin R. Fischer, 395, 403



- Masataka Kaneko, 383  
Mathieu Muratet, 387  
Mathieu Vermeulen, 387  
Matin Yarmand, 407  
Matthew Dunham, 181  
Matthias Stadler, 245, 403  
Meera Sharma, 439  
Megan Humburg, 409  
Megan Wang, 431  
Meghan Krawczyk, 249  
Mei-Hwa F. Chen, 241  
Mei-Yee Mavis Ho, 75  
Mengxi Zhou, 391  
Mengxiao Zhu, 83, 205  
Mi Song Kim, 269  
Michael Foster, 181  
Michele Igo, 375  
Michelle Perry, 189  
Mike Tissenbaum, 197, 431  
Mike Verostek, 181  
Mohammadreza Farrokhnia, 115  
Mohammed Saqr, 261  
Molly Goldstein, 411  
Morgan Vickery, 391  
Myat Min Swe, 377  
Nadir Weibel, 407  
Nastassia Jones, 437  
Nathaniel W. Cradit, 249  
Ndubuisi Anuli, 427  
Nessrine Machaka, 285  
Nga Hoang, 297  
Nicole M. Hutchins, 107  
Nigel Bosch, 189  
Nikol Rummel, 323  
Nisha Devasia, 59  
Nitasha Mathayas, 391  
Noëlle Bender, 355  
Oktay Ülker, 237  
Olga Chernikova, 245, 395  
Omid Noroozi, 115, 315  
Ong Kian Keong Aloysius, 241  
Patricia Santos, 357  
Paul Hatch, 289  
Paula Rodríguez-Chirino, 413  
Pengjin Wang, 429  
Pingting Chen, 331  
Preeti Raman, 439  
Priya Sharma, 435  
Qiannan Niu, 345  
Qianru Lyu, 75, 315  
Qiao Jin, 91  
Qiuyu Lin, 315, 349  
Qizhi Xu, 83, 205  
Ralf Schmidmaier, 403  
Ralitsa D. Debra, 421  
Ravit Golan Duncan, 315, 349, 433  
Remi Kalir, 331  
Renato Carvalho, 439  
René Lobo-Quintero, 373  
Richard Davis, 323  
Rishi Krishnamoorthy, 433  
Roberto Martinez-Maldonado, 323  
Rogers Kaliisa, 261  
Ru Wang, 407  
Rubaina Khan, 439  
Rukmini Manasa Avadhanam, 331  
Ryo Yoshikawa, 371  
Sadhana Puntambekar, 197, 431  
Sameh Said-Metwaly, 365  
Sanaz Ahmadzadeh Siyahrood, 309  
Sarah Evans, 59  
Sarai Coba-Rodriguez, 343  
Sariah Lopez-Fierro, 3  
Sayed Mohsin Reza, 315  
Sebastian Simon, 123  
Seiji Sekine, 377  
Selena Steinberg, 391  
Seonyoung Keum, 269  
Seth Corrigan, 201, 233  
Seung Lee, 253, 437  
Shan Li, 381  
Shenshen Han, 233  
Shigen Shimojo, 11  
Shihui Feng, 209, 399  
Shotaro Naganuma, 67  
Shulong Yan, 257  
Shurui Gao, 51  
Siem Buseyne, 359, 365  
Siew Cheng Aileen Chai, 75, 315  
Sini Davies, 415  
Somi Park, 269  
Sonsoles López-Pernas, 261  
Stacy Cernova, 431  
Sugat Dabholkar, 433  
Susan M. Land, 281  
Svetlana Yarosh, 91  
Sébastien George, 123  
Takaki Kondo, 371  
Takeo Noda, 383  
Takuya Kitamoto, 383  
Tatsuyuki Kawamura, 11  
Teng-Yao Cheng, 173  
Thibault Carron, 387  
Thierry Danquigny, 365  
Thomas Brush, 253  
Tianlong Zhong, 241, 381  
Tianshu Wang, 43, 185, 253  
Tony E. Wong, 181  
Toshinobu Kasai, 217  
Toshio Mochizuki, 377  
Tsubasa Minematsu, 67  
Tsukasa Hirashima, 369  
Tugce Özbek, 99  
Valdemar Stenberdt, 355



Wakako Fushikida, 371  
Wei-Peng Teo, 75, 405  
Weijia Chen, 51  
Wenli Chen, 75, 265, 315, 405  
William R. Penuel, 297  
X. Christine Wang, 309  
Xiaobo Liu, 347  
Xiaojie Zhang, 209, 399  
Xiaoqing Gu, 393  
Xiaoran Li, 379  
Xiaoshan Huang, 381  
Xiaotian Zou, 43, 185, 253  
Xiaoyan Chu, 363  
Xin Wang, 205  
Xinran Zhu, 331  
Xinyi Li, 315  
Xueqi Feng, 155  
Xuesong Cang, 197, 431  
Xuesong Zhai, 363  
Yanping Xu, 281  
Yanqing Sun, 131  
Yeonji Jung, 331  
Ying Zhang, 425  
Ying-Tien Wu, 173  
Yiqiu Zhou, 35  
Young Hoan Cho, 269  
Youngsun Choo, 269  
Yu Li, 315, 349, 423, 425  
Yu Liu, 423, 425  
Yu Xia, 257  
Yu Zhang, 347  
Yugo Hayashi, 11  
Yusuke Hayashi, 369  
Yuxin Chen, 305  
Yvette Baggen, 115  
Zachary Ryan, 315, 349  
Zheng Chu, 241  
Zhuoxin Xiao, 345  
Zihan Guo, 51  
Ziwei Dou, 83  
Zixiang Yu, 425  
Zongting Ge, 83