4-2023

An Exploration of Computational Text Analysis of Co-Design Discourse in a Research-Practice Partnership

Mei Tan
Stanford University, mxtan@stanford.edu

Victor R. Lee
Stanford University, vrlee@stanford.edu

Recommended Citation

Tan, M., & Lee, V. R. (2023). An Exploration of Computational Text Analysis of Co-Design Discourse in a Research-Practice Partnership. Paper presented at the 2023 Annual Meeting of the American Educational Research Association, Chicago, IL.

An Exploration of Computational Text Analysis of Co-Design Discourse in a Research-Practice Partnership

Mei Tan & Victor R. Lee

Stanford University

{mxtan; vrlee} @stanford.edu

Paper presented at the 2023 Annual Meeting of the American Educational Research Association, Chicago, IL

To cite, please use

Tan, M., & Lee, V. R. (2023). *An Exploration of Computational Text Analysis of Co-Design Discourse in a Research-Practice Partnership*. Paper presented at the 2023 Annual Meeting of the American Educational Research Association, Chicago, IL.

Acknowledgments

This work was supported in part by funding from the US National Science Foundation under Grant No. 2031382 and 2031404. The opinions expressed herein are those of the authors and do not necessarily reflect those of the National Science Foundation. The authors thank the participants for their willingness to participate in this research study.

Abstract

In combination with contextualized human interpretation, computational text analysis offers a quantitative approach to interrogating the nature of participation and social positioning in discourse. Using meeting transcript data from the development of a co-design research-practice partnership, we examine the roles and forms of participation that contribute to an effective collaboration between a multileveled school system and researcher partners. We apply computational methods to explore the language of co-design and multi-stakeholder perspectives in support of educational improvement science efforts and our theoretical understanding of partnership roles. Results indicate participation patterns align with documented roles in co-design partnerships and highlight the space dedicated to process reflection, context sharing, and logistical coordination.

Keywords: *statistical natural language processing; co-design; computational text analysis; research-practice partnership*

Objectives

Collaboration and multi-stakeholder involvement to produce systems change are common features of many educational improvement science efforts (Lewis, 2015), although their detection and measurement remain an active area of research. The present study explores the use of computational text analysis methods as a means for detecting collaboration and multistakeholder participation across a multileveled school system and researcher partners. We hypothesize that computational analyses may reflect notable aspects of participation and social positioning in these meetings. Once identified, they become objects of reflection and motivators of change to serve project goals, such as more equitable contribution in group conversations. For this paper, we illustrate what computational analyses applied to meeting co-design transcript data can show us about the collaboration and examine how some of the interactional dynamics change over time.

Theoretical Framework

A consequential move among education researchers in recent years has been to pursue school improvement efforts by cultivating long-term mutualistic research-practice partnerships (RPPs; Coburn & Penuel, 2016). Partnerships involve direct and regular engagement between researchers and practitioners to identify and respond to persistent problems of practice through the combined power of research and practice-based wisdom. One form of partnership research gaining popularity is collaborative design, or “co-design” RPPs. In a co-design, researchers and teachers may work together to design new district-wide curriculum materials to align with new state standards or professional learning models using new design frameworks emerging in the research literature (Severance et al., 2016). Early foundations of the co-design RPP model can be found in design-based research, where the design and enactment of an intervention are

viewed as objects of research intended to inform practice (The Design-Based Research Collective 2003).

The roles and forms of participation that contribute to the development and maintenance of effective RPPs are still being interrogated. Practice partners may not be familiar with nor consistently recognize the unique nature of the partnership interaction with research partners (Lee et al., 2019). Even the division between researcher and practitioner roles may oversimplify complex organizational structures, where professional roles and titles impact modes of engagement differently across contexts (Farrell et al., 2019; Lee et al., 2014; Penuel 2017). We view co-design RPPs as boundary work between multiple communities (Penuel et al., 2015), each with their own structures and norms. Further, we believe that the ways in which we interact through the use of language are reflective of more than the semantics of what is being said. Rather, language serves to position, identify, maintain, and contribute to structures of power and influence in addition to serving many other functions (e.g., Clark, 1996; Gee, 2013). On this foundation, this work examines the extent to which computational methods of examining language in discourse can support our theoretical understanding of co-design roles and participation patterns.

Data

The data used in this work come from a co-design RPP funded by the National Science Foundation creating a math-integrated computer science curriculum in a rural-serving school district that can be implemented by teachers and paraprofessionals in parallel. In the partner district, the decision was made to have paraprofessional educators (titled “Computer Lab Specialists” or “CLSs”), who already lead computer lab instruction related to computer literacy and digital safety, be the primary leaders of new elementary computer science instruction. This RPP was established during the remote work periods of the COVID-19 pandemic, and included

an intentional research component where the interactional dynamics of the RPP were their own object of study in response to a call for more research on how RPPs are structured and unfold over time (Coburn & Penuel, 2016).

For the duration covered by this analysis, the partnership involved seven practitioners, including two district personnel, two fifth-grade teachers and three computer lab specialists. The partnership additionally assembled four primary researchers (i.e., principal investigators) and five graduate student research assistants from two universities, and three guest researchers from separate institutions forming an advisory board. Between June 2021 and June 2022, 54 meetings took place via Zoom and were recorded and automatically transcribed. Five were design meetings involving teachers and CLSs to engage in co-design. The others were project-team planning meetings that involved district and university personnel, much of which involved "designing for co-design" sessions (Lee et al, 2022).

Table 1: *Comparison of design and project meeting transcript duration and utterance lengths*

	n	utterances	mean duration	min duration	max duration
Design Team Meetings	5	1856	58 min	35 min	119 min
Project Team Meetings	49	12467	69 min	22 min	167 min

The transcripts contain 14,323 utterances following the progress of the partnership (Table 1). Because automated transcription can contain some inconsistencies, preliminary data cleaning involved regularizing recorded speaker names, mapping speaker names to participant roles and consolidating consecutive utterances from the same speaker. In less than 1% of utterances, often consisting of short fragments, speaker names were automatically recorded as "unknown speaker", and these utterances were reviewed and removed in analyses involving aggregation by speaker role. Though analyses leveraged the entire dataset, this report focuses

on interpreting the results from the 5 design team meetings in which math teachers and computer lab specialists participated in lesson co-design.

Methods

Participation Metrics

To understand role-based participation, baseline computations extracted simple speech features referenced in studies of dominance in multiparty dialogue, including: the number of speaking turns, floor grabs, interruptions, and durational statistics, aggregated by participant role for each meeting. Part-of-speech tagging facilitated the assessment of psycholinguistic features relating to studies of social rank and influence in language, including pronoun and function word use.

Topic Modeling

To identify focal themes in the partnership, we employ topic modeling, an unsupervised algorithm that clustering words and utterances to detect topics in texts. This study relies on statistical natural language processing techniques, using algorithms based on counting the patterns in which words occur to avoid the assumptions that more complex models introduce. Processed utterances are used as inputs in a Latent Dirichlet Allocation topic model, a Bayesian generative model that estimates the proportion of language devoted to different topics based on the co-occurence of words across utterances (Blei 2012). This established technique in finding topics in conversation analysis is used in a variety of settings, including classification of classroom conversation as academic content and classroom management (Liu 2020). The results treat each utterance as a weighted mixture of topics, where each topic is defined by a weighted list of terms representing the topic vocabulary.

While determining the optimal number of topics to be extracted for greatest topical coherence is a computational task, the labeling and interpretation of topics is a qualitative task

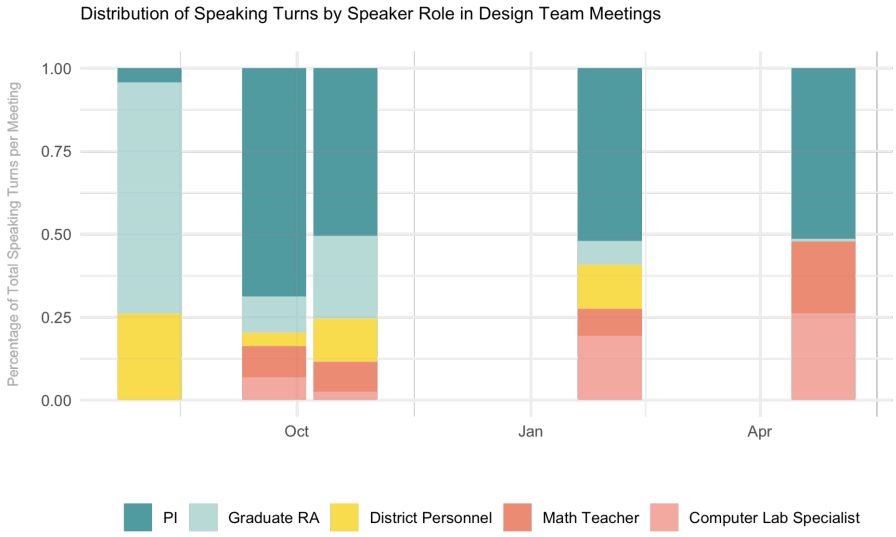
by way of partnership participant interviews. This exploratory interpretation of topics is not definitive but offers a signal that can be leveraged to assess how partnership roles and participation relate to topics of discussion. We model the distribution of topic loading across speaking turns and durations, as well as rate of contribution by participant role.

Results

Who is talking?

Prior studies of dominance in multi-party conversations have noted discriminatory features such as the number of speaking turns, the number of words, and the number of floor grabs by speakers (Rienks 2006). Dominance has also been linked with speaking time, with greater explanatory power in role-based relationships (Schmidt-Mast 2002, Hung 2007). Drawing upon these findings, speech metrics indicate increased participation from math teachers and CLSs in later design sessions. While speaking time is dominated by researchers in the first design sessions, participation shifted to an aggregated near-equal distribution of speaking turns between researchers and practitioners by the last session (Figure 1).

Figure 1: *Distribution of speaking turns by participant role in design meetings over time*



Upon review of transcripts, it was noted that research team members increased participation from practice partners by explicitly calling on practitioners by name and inviting their perspectives. This was a deliberate effort to publicly cede the floor and provide an uninterrupted speaking turn to practitioners. This strategy was taken up more in later co-design sessions, and over time, the number of named references in utterances increased three-fold. While less than 5% of utterances in the earlier design sessions reference participants by name, computer lab specialists were referenced in nearly 20% of utterances in the latter design sessions (Figure 2).

Figure 2: *Number of times participants of each role are referenced by name in design meetings*

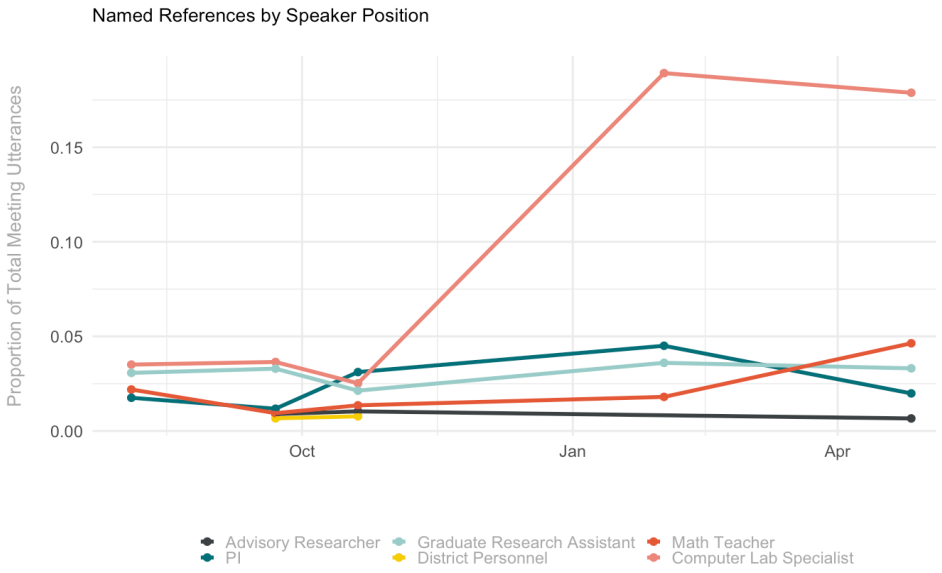
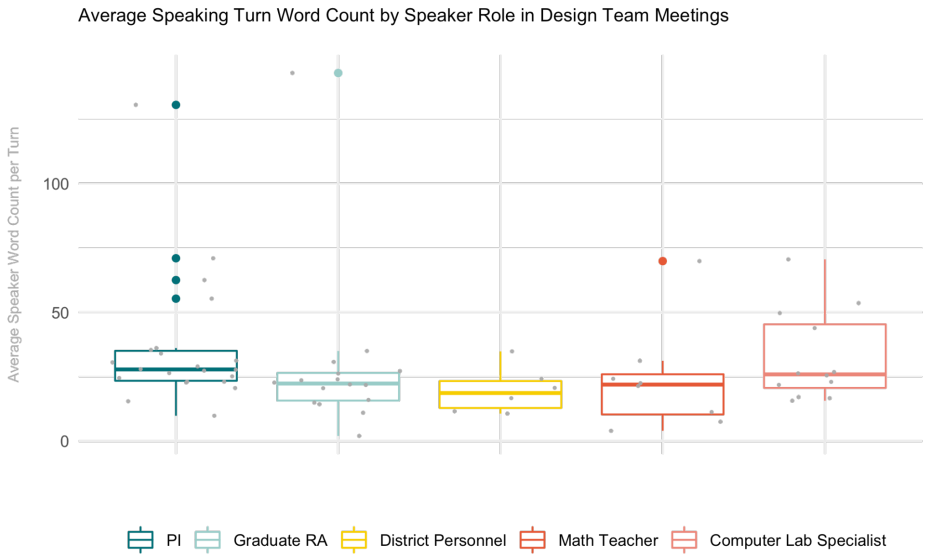


Figure 3: *Distribution of average utterance word counts by participant role in design meetings*



Duration and word count statistics indicate a tendency for researchers (PIs) to monologue (Figure 3). A closer examination of the transcripts showed that researchers sought to act as responsive sensemakers in these longer utterances, stepping back to summarize or emphasize what was being said by the group. A contrasting participation pattern is observed for district personnel (e.g., school district coordinators), who contributed the shortest utterances with decreased participation over time.

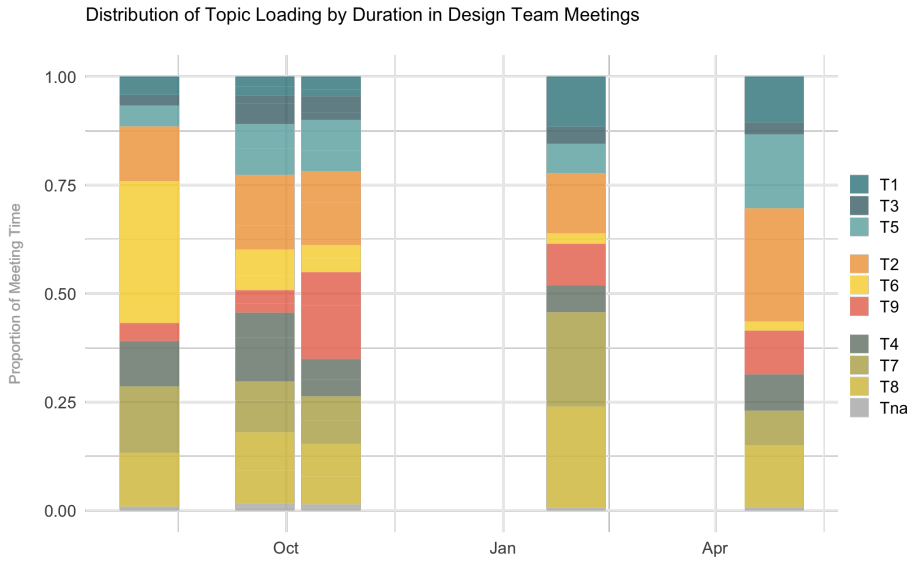
What is the group talking about?

We evaluate the focal themes in the partnership through topic modeling. The most coherent topic model identified 9 topics across all data points, and the keywords for these topics were reviewed and labeled by one of the meeting attendees. Three topics describe aspects of research design, three topics describe aspects of lesson design, and the remaining topics describe the sharing of context and management of logistics (Figure 4).

Figure 4: *Identified topics and interpreted labels*

T1	Design and implementation of surveys to collect and discuss feedback from students and teachers.
T3	Collecting, managing and analyzing data in the research design process.
T5	Reviewing and evaluating the research design, co-design experience and partnership participation.
T2	Integration of mathematics and flexibility in coordinating standards alignment with topical connections in computer science lesson design.
T6	Presentation and discussion of computer science curriculum, scratch and block coding mechanics.
T9	Lesson construction, review and discussion of lesson implementation.
T4	Discussion of the context surrounding the project, partnership, district and participants.
T7	Information access and management, notes, emails, files and remote conferencing.
T8	Meeting scheduling, attendance and logistics.
Tna	No topic loading given.

Figure 5: *Distribution of speaking time by topic in design meetings*



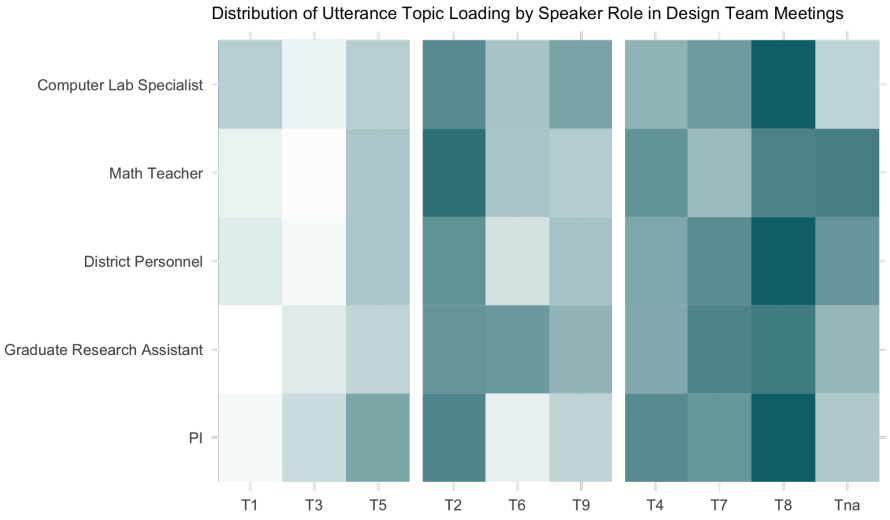
We observed the distribution of time dedicated to each topic across design sessions (Figure 5). Time devoted to the themes of collecting and discussing student feedback and reviewing the research co-design process increased in the later sessions, while the details of

lesson construction and implementation saw the highest engagement in early October. The largest portion of speaking time is dedicated to logistics, scheduling and context-setting, such that in January these topics comprised over 50% of session time. The prioritization of these topics reflect a need to establish norms in working relationships during the formation of a new partnership, and future work will explore the role of social and organizational talk in different types of partnerships and varying timespans. Thus, even with co-design as the focus, substantial talk time was still logistical in nature.

Who talks about which topic, and how?

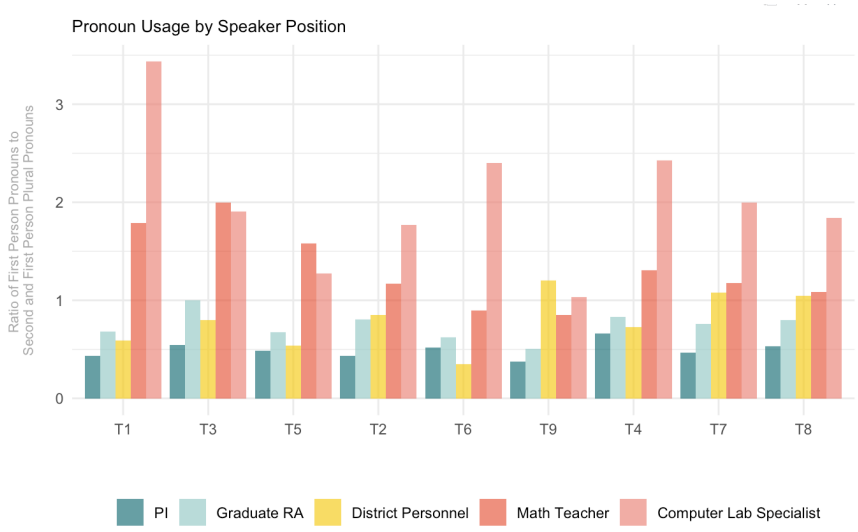
We correlate the rate of contribution for each participant role with identified topics to assess whether particular roles tended to speak on one topic more than another (Figure 6). Expectedly, topics related to research design have lower rates of contribution across participants in design sessions. PIs lead conversation reviewing the design experience (T5) and summarizing takeaways, while CLSs report student feedback and provide commentary on the ways surveys were administered and received (T1).

Figure 6: *Distribution of speaking time by topic for each participant role in design meetings*



All participants engaged actively in conversation related to the integration of math in the design of computer science lessons (T2), though math teachers and primary researchers dedicated higher portions of their speaking time to this topic. CLSs contributed at a higher rate to implementation details and lesson construction (T9), while graduate student research assistants were the primary speakers around the scratch computer programming environment (T6), with engagement from math teachers and CLSs.

Figure 7: *Comparison of rates of first-person pronoun usage by participant roles*



In all topics, computer lab specialists and math teachers used a higher rate of first-person pronouns (Figure 7). Prior work in psycholinguistic analysis has associated using more second and first-person plural pronouns with the other-focused nature of higher social rank, while using more first-person pronouns is associated with the self-focused nature of lower social rank (Kacewicz, 2013). However, this metric equally reflects the speech acts associated with documented roles participants play in partnership. In the designer pilot and validator roles,

practitioners provide input from practice by adapting design to local context and giving feedback from implementation (Sjolund, 2021). Closer examination of the transcripts saw practitioners contribute their observations and first-hand experiences to the discussion.

Significance

The goal of using computational methods is not to supplant existing powerful methods of qualitative human analysis (that are also underway), but to demonstrate additional ways we can use interactions to understand RPPs and co-design. In applying these methods to quantify co-design role dynamics in the domain of a research-practice partnership, we present a new, computationally-powered approach to interrogating the nature of participation and positions in an RPP. Such approaches can help provide further insight into how RPP interactions, and especially co-designed, are organized and enacted. This may become a valuable source of formative feedback for co-design teams and could potentially be used as one additional resource to help assess the effectiveness of RPP collaborations (Henrick et al., 2017).

References

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Coburn, C. E., & Penuel, W. R. (2016, 2016/01/01). Research–Practice Partnerships in Education. *Educational Researcher, 45*(1), 48-54. https://doi.org/10.3102/0013189X16631750

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Farrell, C. C., Harrison, C., & Coburn, C. E. (2019, 2019/04/01). "What the Hell Is This, and Who the Hell Are You?" Role and Identity Negotiation in Research-Practice Partnerships. *AERA Open, 5*(2), 2332858419849595. https://doi.org/10.1177/2332858419849595

Gee, J. P. (2013). An introduction to discourse analysis: Theory and method. Routledge.

Henrick, E. C., Cobb, P., Penuel, W. R., Jackson, K., & Clark, T. (2017). *Assessing Research-Practice Partnerships: Five Dimensions of Effectiveness*. Retrieved from New York, NY:

Hung, H., Jayagopi, D., Yeo, C., Friedland, G., Ba, S., Odobez, J. M., Ramchandran, K., Mirghafori, N., & Gatica-Perez, D. (2007). Using audio and video features to classify the most dominant person in a group meeting. *ACM Multimedia*, 835–838.

Kacewicz E., Pennebaker J. W., Davis M., Jeon M., Graesser A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2):125-143.

Lee, V. R., Clarke-Midura, J., Shumway, J., & Recker, M. (2022). “Design for Co-Design” in a Computer Science Curriculum Research-Practice Partnership. In C. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences - ICLS 2022* (pp. 1049-1052). Hiroshima, Japan: ISLS.

Lee, V. R., Leary, H. M., Sellers, L., & Recker, M. (2014). The role of school district science coordinators in the district-wide appropriation of an online resource discovery and sharing tool for teachers. *Journal of Science Education and Technology, 23*(3), 309-323. doi:10.1007/s10956-013-9465-5

Lee, V. R., Recker, M., & Rogowski, A. (2019). Researchers or Service Providers? A case of renegotiating expectations in a research-practice partnership. In T. Ruecker & V. Svihla (Eds.), *Navigating Challenges in Qualitative Educational Research: Research, Interrupted* (pp. 80-94). New York, NY: Routledge.

Lewis, C. (2015, 2015/01/01). What Is Improvement Science? Do We Need It in Education? *Educational Researcher, 44*(1), 54-61. https://doi.org/10.3102/0013189X15570388

Liu, J., & Cohen, J. (2021). Measuring Teaching Practices at Scale: A Novel Application of Text-as-Data Methods. *Educational Evaluation and Policy Analysis*, 43(4), 587–614. https://doi.org/10.3102/01623737211009267

Rienks, R., Zhang, D., Gatica-Perez, D., & Post, W. (2006). Detection and application of influence rankings in small group meetings. Proceedings of the 8th international conference on Multimodal interfaces. ACM Press, 257–264.

Schmid Mast, M. (2002). Dominance as expressed and inferred through speaking time. *Human Communication Research*, 3, 420–450.

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2016, 2016/10/01). Organizing for Teacher Agency in Curricular Co-Design. *Journal of the learning sciences, 25*(4), 531-564. https://doi.org/10.1080/10508406.2016.1207541

Sjolund, S., Lindvall, J., Larsson, M., Ryve, A. (2021). Mapping roles in research-practice partnerships: a systematic literature review. *Educational Researcher*.