

AN ABSTRACT OF THE DISSERTATION OF

Bill J. Brooks for the degree of Doctor of Philosophy in Chemical Engineering
presented on December 12, 2013.

Title: Technology to Promote Concept-based, Active Learning and Enable Education
Research

Abstract approved: _____
Milo D. Koretsky

This dissertation centers on development of two web-based, in class learning tools, the AIChE Concept Warehouse (CW) and the Web-based Interactive Science and Engineering Learning Tool (WISE), and describes the investigation of student learning through their use. The intent of these tools is to promote deeper conceptual understanding in students through engaging them in active learning pedagogies in the classroom. While the tools provide many useful features, this dissertation will focus primarily on a line of research into activities afforded by the technology's ability to facilitate the presentation of concept questions to individual students in class and the subsequent collection of student responses in the form of multiple-choice answer selections and associated short written reflections justifying that answer choice.

The dissertation presents four studies that investigate aspects of student learning using these tools in the active learning classroom. The first study investigates a structured pedagogy called peer instruction. In peer instruction students first answer a question individually, then form small groups where they discuss their answers and explain their reasoning, and finally re-answer the same question individually. The effect of displaying a histogram of responses to the class of their individual answer

selections before group discussion is investigated. It found that students improve their understanding when engaging in peer instruction even if they answered correctly at first and that students tend to change to the popular answer irrespective of the histogram. It also found that even though the histogram does not appear to influence correct answer selections, it did influence self-reported confidence ratings.

The second study uses a crossover experimental design to investigate the influence of prompting students to write an explanation justifying their answer choice when they respond individually. Results show improved thinking with written explanations. It argues that improved thinking is prompted by the students' construction of logical arguments to justify their answer choices.

The third and fourth studies further explore written explanations. The third illustrates the novel use of word clouds as a quick analysis method for large quantities of written explanations, and describes the integration of that feature into the CW. The fourth study characterizes the amount of time that it takes for students to answer questions. The results show that response time is affected by writing an explanation, whether students chose the correct multiple choice answer or not, and the question difficulty.

Finally, some preliminary data are presented on developing a psychometric measure, based on modified Item Response Theory (IRT), to provide an automated mechanism of question quality control in the CW.

©Copyright by Bill J. Brooks
December 12, 2013
All Rights Reserved

Technology to Promote Concept-based, Active Learning and Enable Education
Research

by
Bill J. Brooks

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented December 12, 2013
Commencement June 2014

Doctor of Philosophy dissertation of Bill J. Brooks presented on December 12, 2013.

APPROVED:

Major Professor, representing Chemical Engineering

Head of the School of Chemical, Biological & Environmental Engineering

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Bill J. Brooks, Author

ACKNOWLEDGEMENTS

I should probably start by acknowledging God, or whatever you want to call him, her, it, or they – that universal spirit that animates and binds all things in existence. While I am not sure what to attribute my indomitable will and persistence to, I am confident in attributing my abundant patience to God. That patience has given me an interesting and invaluable perspective that has been a boon over the course of the latter half of my life, even if it is frequently impossible to articulate properly.

My friends and family have been a fount of support throughout my over-extended academic career. Grandpa was rather supportive a time I will only vaguely allude to when I needed support the most. I should acknowledge my friends Garrett and Craig for not giving me too much of a hard time about being an absentee friend so frequently and not giving me too much of a hard time about when I'll finally graduate. The same goes for my *little* brother, Brian.

I appreciate my advisor, Dr. Milo Koretsky, who gave me support and guidance in my research and studies. It is a testament to his patience and understanding that he has so frequently accommodated or forgiven my considerable set of idiosyncrasies. I know awkwardness and frustration has to have been a consistent theme in his life whenever I've been involved.

I want to thank all of my current and former laboratory members, office mates, other peers, and folks I've otherwise worked with. I especially want to express gratitude to Debbi Gilbuena. She's become like a sister over the course of graduate school.

I want to thank the faculty members both here at OSU and at OIT that I fondly remember from courses and other interactions throughout both my graduate and undergraduate studies. I would name them, but the list would be lengthy and I am afraid I might miss some as I write this. The list of departments they are or were associated with is lengthy enough on its own. They hail from such disparate fields as computer engineering, electronics, math, physics, chemistry, science and math education, and chemical engineering.

I would be remiss if I didn't explicitly thank the COE IT folks, Paul Montagne, Jordan Jones, Keith Price and Steve Cleveland. Steve's contribution is especially relevant. Steve has helped solve problems he may never know existed because I solved them while writing an email seeking his miraculous help. There is a particular irony in that considering Chapter 3 of this dissertation.

Finally, I would like to thank all of those other people not explicitly referenced but without whom this dissertation would not be possible.

Thank You!

CONTRIBUTION OF AUTHORS

In Chapters 2, Dr. Milo Koretsky was involved with the study design, writing, and interpretation of data. Dr. Debra Gilbuena coded a subset of the data to provide inter-rater reliability and some funky figure formatting. In Chapter 3, Dr. Milo Koretsky and Dr. Adam Higgins contributed to the study design, writing, and interpretation of data. Rachel White and Alec Bowen were involved with writing; they were extensively involved with emergent code development and subsequent coding of written response data. In Chapter 4, both Dr. Debra Gilbuena and Dr. Milo Koretsky contributed to the writing and interpretation of the data, some of which came from the study in Chapter 3. Dr. Stephen Krause contributed to the writing as well as word cloud data and interpretation for the muddiest point reflections. In Chapter 5, Dr. Milo Koretsky and Dr. Dedra Demaree were involved with writing and interpretation of the data. In Appendix 2, Dr. Debra Gilbuena was involved in all aspects of the paper. In Appendix 3, Dr. Milo Koretsky and Dr. Adam Higgins contributed to data collection, writing and analysis; Dr. Marita Barth and Paula Weis were involved with data collection. Samuel Mihelic contributed to applying the IRT model to the results from the chemistry concept questions and significantly modified and documented my original algorithm for fitting the IRT model.

TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	1
1.1 Background.....	1
1.2 Theoretical Framework	3
1.3 Software Development	5
1.3.1 Web-based Interactive Science and Engineering (WISE) Learning Tool	6
1.3.2 American Institute of Chemical Engineering (AIChE) Concept Warehouse	8
1.4 Dissertation Summary	13
2. The Influence of Group Discussion on Students' Responses and Confidence during Peer Instruction	14
2.1 Abstract.....	15
2.2 Teaching and Learning with Peer Instruction	16
2.3 Methods	20
2.4 Results	25
2.4.1 Multiple-Choice Responses	25
2.4.2 Change in Multiple-Choice Response after Group Discussion	26
2.4.3 Change in Codes of Written Explanations after Group Discussion.....	29
2.4.4 Student Confidence before and after Group Discussion	32
2.4.4 Confirmation of Consensuality-Correctness Trend	34
2.5 Discussion.....	35
2.6 Conclusion.....	40
2.7 Supporting Information	40
2.8 Acknowledgments	41
2.9 References	41
3. Promoting Scientific Reasoning through Written Explanations to Multiple-Choice Concept Questions.....	45
3.1 Abstract.....	46
3.2 Introduction	47
3.3 Background.....	48
3.4 Experimental Methods.....	55
3.4.1 Participants & Setting	55
3.4.2 Study Design	56
3.4.3 Data Analysis	60

TABLE OF CONTENTS (Continued)

	<u>Page</u>
3.5 Results	63
3.5.1 Quantitative Analysis of Multiple-Choice Responses	63
3.5.2 Qualitative Analysis of Isomorphic Concept Questions	65
3.6 Discussion.....	70
3.7 Conclusions	76
3.8 Acknowledgements	77
3.9 References	77
4. Using Word Clouds to Analyze Students' Short Written Responses	82
4.1 Abstract.....	83
4.2 Introduction	83
4.3 Background.....	85
4.4 AIChE Concept Warehouse	87
4.5 Word Clouds in ConcepTest Assessment.....	88
4.5.1 Methods.....	89
4.5.2 Results	91
4.5.3 So what can we learn from these word clouds?	94
4.6 Word Clouds to Examine Muddiest Point Reflections.....	96
4.6.1 Methods.....	96
4.6.2 Results	97
4.6.3 So what can we and students learn from the Muddiest Point Reflection word cloud?	98
4.7 Word Clouds for other Short Written Exercises.....	100
4.8 Planned Word Cloud Improvements for the AIChE Concept Warehouse	101
4.9 Conclusions & Implications	101
4.10 Acknowledgements	102
4.11 References	102
5. Student Response Times to In-class Thermodynamics Concept Questions	106
5.1 Abstract.....	107
5.2 Introduction	107
5.3 Methods	109

TABLE OF CONTENTS (Continued)

	<u>Page</u>
5.4 Results	112
5.5 Discussion.....	116
5.6 Conclusions & Implications	118
5.7 Acknowledgements	118
5.8 References	118
6. Conclusion.....	121
Bibliography.....	127
APPENDICES.....	137
Appendix A – The Influence of Group Discussion on Students’ Responses and Confidence during Peer Instruction Supporting Information.....	138
A.1 Coding Scheme	139
A.2 Conceptual questions used in this study	140
A.3. ANOVA of Student Confidence	148
Appendix B – Preliminary Development of the AIChE Concept Warehouse	149
B.1 Introduction	150
B.2 Design for Diffusion.....	154
B.3 The AIChE Concept Warehouse – Design & Development	161
B.4 The AIChE Concept Warehouse – Beta Testing.....	171
B.5 Community Development Activities.....	173
B.6 Future Plans	175
B.7 Summary & Implications for the Future	176
B.8 Acknowledgments	177
B.9 References	177
Appendix C – Item Response Theory Analysis of Multiple-choice Concept Questions using the AIChE Concept Warehouse.....	180
C.1 Introduction	181
C.2 Background & Theoretical Framework.....	182
C.3 Methods	184
C.4 Results & Discussion	193
C.5 Conclusion and Implications for the Future	196
C.6 References	198

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 1.1 Screenshot of the WISE student interface showing a sample question and some of the data types that can be collected (e.g. multiple-choice answer and written explanation).....	7
Figure 1.2 Screenshot of a WISE data view with sample data for the question depicted in Figure 1.1.	8
Figure 1.3 Screenshot of a student's (incorrect) answer from the questions tab from the student interface of the CW.....	9
Figure 1.4 Screenshot of the AIChE Concept Warehouse instructor conceptests search page.	11
Figure 1.5 Screenshot of sample detailed results from the CW. Note: the web-page below this image contains all of the written explanations and answer selections in addition to the rest of the word clouds.	12
Figure 2.1. Percentage of students who changed from the correct multiple-choice answer to an incorrect answer (% from correct) and the percentage change from an incorrect answer to the correct answer (% to correct). (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion (*p<0.05).	27
Figure 2.2. Percentage of students who changed from the popular multiple-choice answer to the unpopular answer (% from consensus) and the percentage change from an unpopular answer to the popular answer (% to consensus). (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion (*p<0.05).	29
Figure 2.3. Number of students per exercise whose explanations improved (+ code Δ) or got worse (-code Δ). (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion (*p < 0.05).	30
Figure 2.4. Average number of students per exercise whose explanations improved (+ code Δ) or got worse (code Δ) organized by their multiple-choice answer change. (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion: combination of the results for the consensually correct and the consensually wrong exercises (*p < 0.05).	32

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure 2.5. Confidence vs correctness, for consensually wrong (CW) and consensually correct (CC) exercises. (A) Before group discussion (pre), with intermediate results shown. (B) Before group discussion (pre), without intermediate results shown. (C) After group discussion (post), with intermediate results shown. (D) After group discussion (post), without intermediate results shown.	34
Figure 3.1. Expanding Piston question as it was presented in the treatment condition of Cohort B.	50
Figure 3.2. Balloon Rising as it was presented in the treatment condition of Cohort A.	58
Figure 3.3 Histograms of (a) the number of questions binned by aggregate correct answer percentage across students and (b) number of students binned by aggregate correct answer percentage for each student across questions.	63
Figure 3.4 Percentage of questions where the treatment condition was (white) and was not (shaded) statistically different from the comparison.	64
Figure 3.5 Percentage of correct answers for the treatment condition and comparison for the Expanding Piston and Balloon Rising concept questions.	67
Figure 3.6. Code distribution for the explanations written by the treatment condition for (a) Expanding Piston and (b) Balloon Rising.	68
Figure 4.1. Sample question (piston question) as it was presented to students who wrote explanations.	90
Figure 4.2. Sample question (balloon question) as it was presented to students who wrote explanations.	91
Figure 4.3. Muddiest point reflection as it was delivered to students.	97
Figure 5.1. Sample concept question.	111
Figure 5.2. Average response time vs. question difficulty when students did not write an explanation for (a) correct answers ($P < 0.001$) and (b) incorrect answers.	115
Figure 5.3. Average response time vs. question difficulty when students wrote an explanation for (a) correct answers ($P < 0.001$) and (b) incorrect answers.	115

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 2.1. Cohort A Answer Summary for Students Who Saw Results Displayed before Group Discussion.....	26
Table 2.2. Cohort B Answer Summary for Students Who Not See Results Displayed before Group Discussion.....	26
Table 2.3. Answer Summary for Questions in Which the Most Popular Answer Was Correct, yet No Answer Had a Majority	35
Table 3.1. Experimental design matrix	57
Table 3.2. Sample student explanations for Expanding Piston (EP) and Balloon Rising (BR) by code level.....	62
Table 4.1. Multiple-choice answer options, word clouds, and representative explanations for the concept question depicted in Figure 4.1 (piston question). ...	93
Table 4.2. Multiple-choice answer options, word clouds, and representative explanations for the concept question depicted in Figure 4.2 (balloon question)..	94
Table 4.3. Word cloud and sample quotes for the sample Muddiest Point Reflection.	98
Table 5.1. Average time taken (in seconds) for correct and incorrect answers, with and without written explanations. N indicates the number of responses in that condition.	113

LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
Figure A.1. Throttling Valve exercise pre discussion.....	140
Figure A.2. Throttling Valve exercise post discussion	141
Figure A.3. Equilibrium exercise	142
Figure A.4. Spray Can exercise.....	143
Figure A.5. Mixing exercise.....	144
Figure A.6. Adiabatic Air exercise.....	145
Figure A.7. Solid-Liquid Melt exercise	146
Figure A.8. Chem Reaction exercise.....	147
Figure B.1. Number of questions added to the AIChE Concept Warehouse during the initial development period.....	155
Figure B.2. Screenshot of a student’s answer from the questions tab from the student interface	160
Figure B.3. Advantages of a web-based user interface and a commercial database *O’Reilly40, **Bulger et al.41	162
Figure B.4. Step 2: Storyboard for the question search section with keyword search and filtering options highlighted, November 2010.	165
Figure B.5. Step 3: Implementing the question search section of AIChE Concept Warehouse prototype V1, February 2011.	166
Figure B.6. Step 4: Revision after design team testing of the question search section of AIChE Concept Warehouse prototype V2, August 2011.	168
Figure B.7. Step 5 (in progress): Revision after initial external testing of the question search section of AIChE Concept Warehouse prototype V3, December 2011....	170
Figure C.1. Question “Outlet vs. Inlet Enthalpy in Isothermic Endothermic Reaction” as presented to the treatment group	188
Figure C.2. Sample questions to highlight the effect of model parameters on probability prediction. A) dashed line (a=2; b=-0.6; c=0) B) solid line (a=-2; b=0.8; c=0)	190

LIST OF APPENDIX FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
Figure C.3. Representative question (CW question ID: 2213) response curve from the pre-test. The blue line is the IRT model fit, predicting the probability students will select a correct answer as a function of their ‘aptitude’. The red stars represent the average score for students in their aptitude group. The number at the top (3.01109) gives the Chi-square value representing the fit (poor) between the model prediction and the data.....	194
Figure C.4. Representative question (CW question ID: 2210) response curve from the post-test. The blue line is the IRT model fit, predicting the probability students will select a correct answer as a function of their ‘aptitude’. The blue stars represent the average score for students in their aptitude group. The number at the top (0.394234) gives the Chi-square value representing the fit (good) between the model prediction and the data.	195
Figure C.5. Flagged question (CW question ID: 2230) response curve from the post-test. The blue line is the IRT model fit, predicting the probability students will select a correct answer as a function of their ‘aptitude’. The blue stars represent the average score for students in their aptitude group. The number at the top (1.45784) gives the Chi-square value representing the fit (bad) between the model prediction and the data.....	196

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
Table A.1. Coding scheme for conceptual questions.	139
Table A.2. ANOVA: Confidence as a function of Correctness, Consensuality, Pre/Post Peer Interaction, and whether or not a bar graph of intermediate responses was displayed.....	148
Table B.1. Attributes of an innovation described by Rogers	154
Table B.2. Summary of instructor interface functions organized by sub menu sections within each tab	156
Table B.3. Summary of AIChE Concept Warehouse Student functions organized by tab and sub menu sections	159
Table C.1. Experimental design matrix.....	186

Technology to Promote Concept-based, Active Learning and Enable Education Research

1. Introduction

The research in this dissertation focuses on characterizing student responses to concept questions using in-class active learning pedagogies. This chapter provides some background for the studies that are presented in the following chapters. First, the motivation and theoretical framework are presented. Two web-based software tools, the Web-based Interactive Science and Engineering (WISE) Learning Tool and the AIChE Concept Warehouse (CW), are described next. Finally, an outline of the manuscripts that comprise the remainder of the dissertation is presented.

The two software tools were developed as a central part of this project. They enable delivery of concept questions and facilitate data collection and analysis. They were designed to be used in class and promote students' construction of conceptual understanding through active learning. They allow students to select an answer choice to a multiple choice concept question and then write a reflection on why they chose that answer. The WISE learning tool was developed first, essentially as an in-house tool for use at Oregon State University. The CW was then developed to extend the impact this type of tool and make it available to faculty and students across the United States and internationally.

1.1 Background

Science, Technology, Engineering and Mathematics educators have been concerned about the gap between rote problem solving ability and conceptual understanding for decades (Nurrenbern & Pickering, 1987; Sawrey, 1990). In spite of

many attempts at reform, education is still heavily lecture based and emphasizes routine problem solving. This routine problem solving focus can reinforce rote memorization over conceptual understanding (Elby, 1999; Watters & Watters, 2007). In one study, physics students achieved a 70% success rate on numerical problems and only 10% for conceptual problems for the same concept (McDermott, 2001). In another study, chemistry students scored 95% on numerical problems and only 38% on conceptual questions (Haláková & Prokša, 2007). This lack of conceptual understanding hinders the student's ability to solve new problems based on the same concept because they lack the functional understanding to use their knowledge in new situations (Bransford, 2007).

It has been shown that student-centered, active learning environments are more effective than traditional lectures at fostering conceptual understanding (Hake, 1998; Prince, 2004; Poulis, Massen, Robens, & Gilbert, 1998). In 2010, Dancy and Henderson studied the pedagogical practices and instructional change of physics faculty. They found a number of impediments to adoption of more effective pedagogies such as active learning. The largest impediment for faculty to adopt these practices was unelaborated time; the second largest was specifically time to learn about and implement the necessary changes. Quality concept questions are difficult to construct and their development can consume a large amount of time (Crouch, Watkins, Fagen, & Mazur, 2007; Fagen, Crouch, & Mazur, 2002). One objective of the projects in this dissertation is to reduce the time and effort needed for engineering instructors to implement active learning. Therefore, the research focuses on

techniques that are more amenable to integration into engineering learning environments as they currently exist.

Peer Instruction (Mazur, 1997) is a structured questioning process that is intended to actively engage all of the students in a class. In this technique, a multiple-choice ‘ConcepTest’ or ‘clicker’ question is posed to the class. Students first answer the question individually. Depending on the aggregate response, students can be encouraged to discuss their answers in small groups and then individually give a final answer. This sequence is then typically followed by a class-wide discussion. Through this formative assessment and feedback, the instructor can dynamically adjust the pace and coverage of lecture to match student learning. Peer Instruction (PI) is a relatively simple modification to lecture that prompts students to actively engage in their own learning, to think critically about the material during class, and to learn from and teach each other. However, PI is not always practiced as prescribed (Dancy and Henderson, 2010; Turpen and Finkelstein, 2009). There is a variation in use that can influence its effectiveness and this variation needs additional study. Next, I describe the theoretical lens from which much of this work is pursued.

1.2 Theoretical Framework

In investigating the effectiveness of concept-based, active learning in the classroom, it is useful to consider both the constructivist and sociocultural perspectives of student learning. Constructivism is a way to explain why educational activities need to be learner centered and active; students need to construct mental schema to learn and, as for that construction, activity is necessary. From a

constructivist view, students arrive in the classroom with prior knowledge and preconceptions about how the world works. The construction of knowledge is viewed to be the result of a student's use of existing knowledge to make sense of new experiences. The sociocultural perspective shows how this construction is mediated through the situated nature of the classroom and interaction with peers.

Effective teaching must engage preconceptions and leads to both the modification of concepts and the reorganization of knowledge structures (Bransford, 2000). Draper, Cargill and Cutts (2002) suggest that during these types of activities, students learn because they are required to engage in a level of cognitive processing or reprocessing of the lecture material. Posner, Strike, Hewson and Gertzog (1982) posit a model where assimilation or accommodation can be used to process information that challenges preconceptions, assuming the new, conflicting information is not rejected as an anomaly. In this context, assimilation is the use or modification of current schema to interpret the new information; accommodation is the addition of new schema to replace or supplement the old. From Clark's (2006) point of view, students "hold multiple conceptual elements and ideas at various levels of connection, contradiction and organization" and gains in understanding occur by the process of restructuring their ideas. This viewpoint is reinforced by the novice/expert literature which shows that experts differ from novices not only in the extent of their domain specific knowledge but also in the organization of that knowledge (Chi, Peltovich & Glaser, 1981).

The sociocultural perspective views learning as a process of transforming participation in valued sociocultural activities (Lave & Wenger, 1991). Learning is socially mediated and intimately influenced by the culture and activities in which the learning is situated. Penuel, Abrahamson and Roschelle (2006) suggest that the sociocultural perspective can place into context key elements of student learning using active, technology-based pedagogies where the nature of classroom interactions fundamentally change. They identify the key role of tools like language in mediating cultural activities. From this perspective, engagement in the talking and writing of science coincides with participation in a community of practice that develops expertise. Specifically, learning in engineering requires dialog about the concepts. This dialog is encouraged both through the pedagogy (e.g., discussion portion of PI) and the technology (e.g., written explanations of multiple choice concept questions). While the former has been studied extensively, the role of the latter has not.

1.3 Software Development

To facilitate active learning pedagogies such as PI, a variety of hardware and software tools have been developed. While polling via a show of hands or flashcards may be appropriate for Peer Instruction as originally developed by Mazur, technology like “clickers” can afford potentially beneficial options (e.g. written explanations) and data logging. Clickers have been commercially produced and marketed by a variety of companies (e.g. Turning Point Technologies, iClicker, Qwizdom, etc.). Ubiquitous presenter is an example of an enhanced tool that allows the collaborative sharing and annotating of power point slides in class. While it enables an added dimension of

dialog, Ubiquitous Presenter software requires instructors to have a tablet PC and, ideally, the students as well. The specific tools developed for this project, the Web-based Interactive Science and Engineering (WISE) Learning Tool and the AIChE Concept Warehouse (CW), are described next.

1.3.1 Web-based Interactive Science and Engineering (WISE) Learning Tool

The WISE learning tool (Koretsky & Brooks, 2008) was initially developed to leverage the College of Engineering's wireless laptop initiative to enable and promote active learning pedagogies. The wireless laptop initiative mandated engineering students to own a Wi-Fi capable laptop that they can bring to class. WISE permitted the use of hardware and infrastructure that students already possessed to get similar activity and interaction as with clickers but collect richer data like Ubiquitous Presenter, without the additional hardware requirements of either. Top Hat Monocle (<http://www.tophatmonocle.com/>) and Lecture Tools (<http://www.lecturetools.com/>) are similar web-based tools that have been commercially produced in the last several years, after WISE was developed; a large advantage of these tools is that they are compatible with most web-enabled devices (e.g., laptops and cellphones).

WISE provides important data collection features such as the ability to automate collection of short answer explanations and self-reported confidence. These elements of WISE make it particularly useful for instant summarization of student responses and convenient collection of the results for analysis. Figure 1.1 provides a screenshot of the WISE student interface. Figure 1.2 shows summary data available for class discussion after students have responded to the question from Figure 1.1.



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

[Virtual Hand Raise](#)

[Exercises](#)

[Conceptualization](#)

[Math Primer](#)

[Real World](#)

[Examples](#)

[Links](#)

[My Statistics](#)

[Homepage](#)

Logged in as:

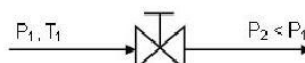
Milo ([logout](#))

Contact point:

[Milo Koretsky](#)

An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and the pressure drops through the valve to a lower value, P_2 .

Assuming the valve is well insulated and inlet and outlet pipes connected to the valve are the same diameter, what is the relationship of the outlet temperature T_2 to the inlet temperature T_1 ?



- $T_2 > T_1$
 $T_2 < T_1$
 $T_2 = T_1$
 Can't answer until I know what gas is flowing

Please explain your reasoning in the space provided below.

Submit

Figure 1.1 Screenshot of the WISE student interface showing a sample question and some of the data types that can be collected (e.g. multiple-choice answer and written explanation).

One element that limits the use of WISE in a broader context is that it was designed such that each instructor has their own bank of questions that could not easily be shared. The sharing and organization of, and searching for, questions are improvements in the following software development activity, the AIChE Concept Warehouse.



Web-based Interactive Science and Engineering Learning Tool

Grade Sheet

[Exercise Library](#)
[Class Activities](#)
[Grade Book](#)
[View Grade Sheet](#)
[Manage Students](#)
[View Statistics](#)
[Administration](#)
[Help](#)

Logged in as:
 Milo ([logout](#))

Contact point:
[Milo Koretsky](#)

An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and the pressure drops through the valve to a lower value, P_2 .

Assuming the valve is well insulated and inlet and outlet pipes connected to the valve are the same diameter, what is the relationship of the outlet temperature T_2 to the inlet temperature T_1 ?

Please explain your reasoning in the space provided below the answer choices.



1 $T_2 = T_1$		0
2 $T_2 < T_1$		23
3 $T_2 > T_1$		9
4 Can't answer until I know what gas is flowing		0

Source: Ron Miller

Figure 1.2 Screenshot of a WISE data view with sample data for the question depicted in Figure 1.1.

1.3.2 American Institute of Chemical Engineering (AIChE) Concept Warehouse

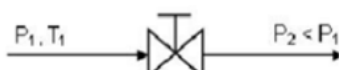
The AIChE Concept Warehouse (CW) was developed with the goal of creating a community of learning within the discipline of chemical engineering (ChE) focused on active, concept-based instruction. In the same way as WISE, it leverages commonly available hardware and compatibility of web-enabled devices to lower the barrier to adoption. Figure 1.3 shows how a student sees the question on his/her device (the same question as Figure 1.1).

Class : Thermodynamics ▾

Time remaining: 00:03:10

An **ideal gas** flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and the pressure drops through the valve to a lower value, P_2 .

Assuming the valve is well insulated and inlet and outlet pipes connected to the valve are the same diameter, what is the relationship of the outlet temperature T_2 to the inlet temperature T_1 ?



- $T_2 > T_1$
 $T_2 = T_1$
 $T_2 < T_1$
 Can't tell without knowing the type of gas.

Please explain your answer in the box below.

since well insulated and no work is being done, deltaU has to equal zero. This means that the energy lost from the drop in pressure would lead to an increase in temperature.

Please rate how confident you are with your answer.

- substantially moderately neutral moderately substantially
 unsure unsure confident confident

Submit

Figure 1.3 Screenshot of a student's (incorrect) answer from the questions tab from the student interface of the CW.

The CW also decreases another large barrier to adoption by housing questions pertinent to courses throughout the core curriculum. The initial development paper (Brooks et al., 2012, see Appendix B) presents a description of the status and available

features at that time. It also reports on initial deployment, community building activities, and future plans for the AIChE Concept Warehouse.

The AIChE Concept Warehouse project addresses several limits of WISE to enable more general use. Additional features have been added as well, including concept inventories, reflection exercises, and interactive virtual laboratories. The questions are organized by category: class, misconception (defined here as a miscategorization of topic, e.g., temperature vs. heat), topic, subtopic, and surface feature. Surface features as we define them are superficial attributes that typically distract students; two problems may require the same concept to solve but different surface features. Searching can be accomplished two ways. First, the categorizations can be selected and associated questions will be presented. The second is a text based search. Text is searchable like Google.com, both in options and speed. Figure 1.4 shows an example screenshot of a text search for “open system” for the categories “first law of thermodynamics (topic) and “open valve” (surface feature).

AIChE Education Division **CONCEPT WAREHOUSE** Brooks | [logout](#) | [help](#)

HOME CONCEPTTESTS CONCEPT INVENTORIES INSTRUCTIONAL TOOLS CLASSES PROFILE SUPPORT

ABOUT SEARCH **MANAGE TESTS** STATISTICS NEW QUESTION

Class :

Classes [Advanced Search](#)

Misconceptions Showing 1-5 of 8 results (in 1.036106s) Results per page:

Topics

1st Law of Thermodynamics

Subtopics

Surface Features


Open valve (adiabatic)

My Questions

Unused (self)

Type

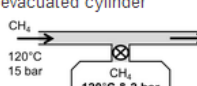
ID:1 Modified Throttling Valve



An **ideal gas** flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and...

Contributor: Bill Brooks Class: Energy Balances
 Author: Ron Miller Thermodynamics
 Type: Multiple Choice
 Rating: ★★★★★

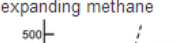
ID:77 Flow work on a partially evacuated cylinder



Methane at 15 bar and 120°C flows into a container that has methane at 120°C and 2 bar when the valve is opened. When the pressure in the container reaches 15 bar, the temperature of the methane in the container is _____ 120°C.

Contributor: John Falconer Class: Thermodynamics
 Type: Multiple Choice
 Rating: ★★★★★

ID:109 Following the isenthalp of expanding methane



Methane undergoes a rapid expansion through an insulated valve. The starting pressure is 5,000 psia and the starting temperature is 475°R. The temperature _____.

Contributor: John Falconer Class: Thermodynamics

Figure 1.4 Screenshot of the AIChE Concept Warehouse instructor conceptests search page.

One focus of the studies presented in this dissertation is the effect of written explanations on students' answer selection and student thinking. The CW has a word cloud function that is useful in this context. The CW automatically generates word clouds from written response data. Normally, they are aggregated into a single cloud

but, for multiple choice concept questions, they are also categorized by answer option (see Figure 1.5).

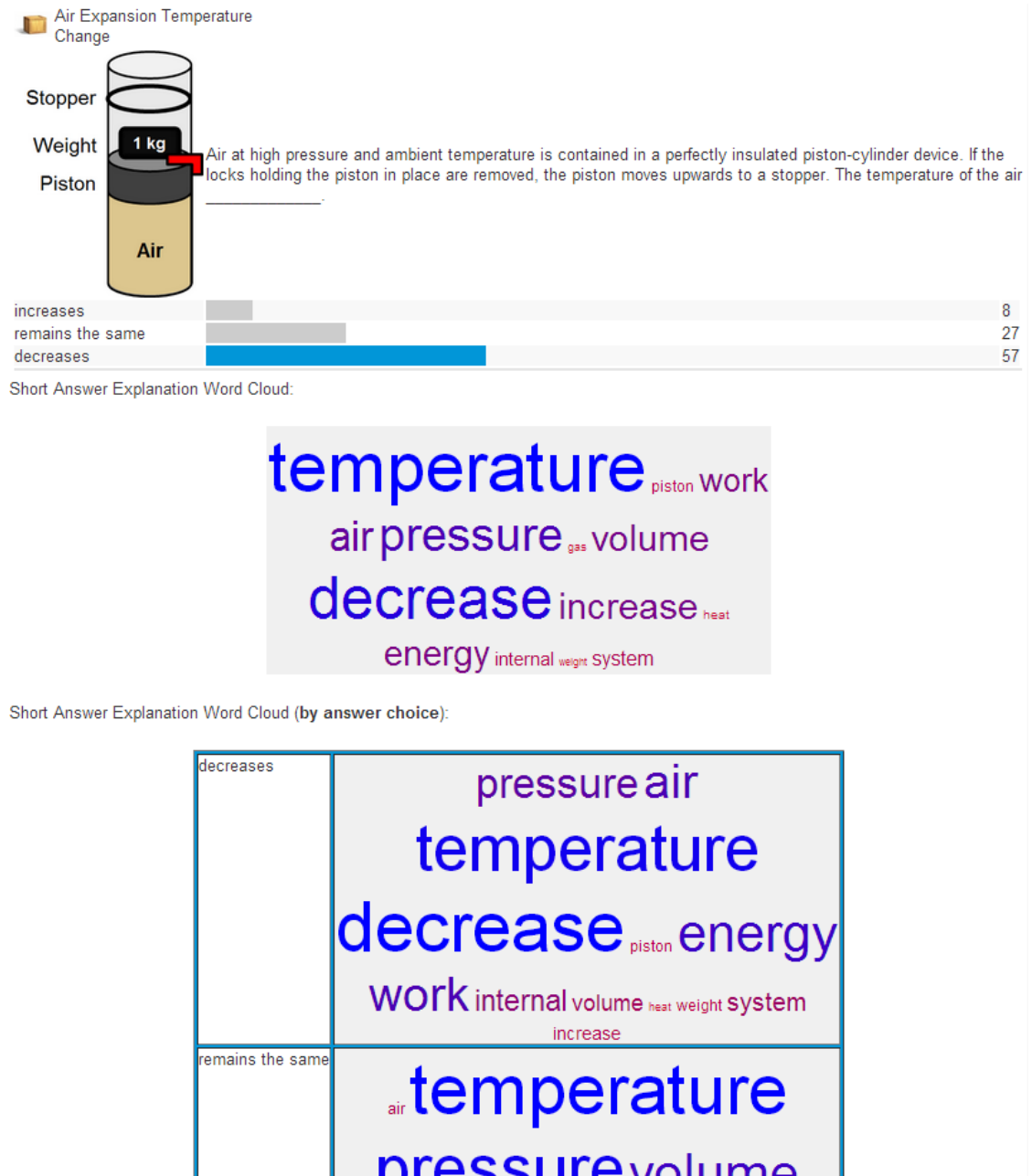


Figure 1.5 Screenshot of sample detailed results from the CW. Note: the web-page below this image contains all of the written explanations and answer selections in addition to the rest of the word clouds.

1.4 Dissertation Summary

This dissertation is organized in a manuscript format as follows. Chapter 2 presents a study that examines the effects of showing an “intermediate” distribution of answers to the class, i.e., after they have answered a multiple-choice concept question for the first time, but before they assemble into groups. Chapter 3 describes a study that used a split design to examine differences in multiple choice responses when a short answer explanation is requested as compared to when one is not. Chapter 4 presents a description of how word clouds have been integrated into the CW and how they can be used to quickly analyze written responses even when class sizes are large. Chapter 5 reports on a study to characterize response times to multiple-choice concept questions when written reflections were requested. Finally, a report of preliminary findings related to the application of Item Response Theory to concept questions where students write short answer explanations is included as an appendix.

2. The Influence of Group Discussion on Students' Responses and Confidence during Peer Instruction

Bill J. Brooks & Milo D. Koretsky

*School of Chemical, Biological, and Environmental Engineering

Oregon State University

Corvallis, OR, USA 97331-2702

(541)737-4591

Journal of Chemical Education

<http://pubs.acs.org/doi/abs/10.1021/ed101066x>

November 2011, Volume 88, Issue 11, pages 1477-1484

2.1 Abstract

Peer instruction is an active-learning pedagogy in which students answer short, conceptually based questions that are interspersed during instruction. A key element is the group discussion that occurs among students between their initial and final answers. This study analyzes student responses during a modified form of peer instruction in two chemical thermodynamics classes. It seeks to understand the changes in student thinking that result from group discussion. In addition to selecting multiple-choice answers, students are asked to write a short explanation of their choice and rate their confidence in their answer. The written explanations have been coded based on the depth and accuracy of the concepts applied. Two cohorts are studied. For Cohort A, initial results of the entire class are shown to the entire class before group discussion, whereas for Cohort B, they are not. After group discussion, the majority of students select the consensus answer, whether it is right or wrong, and whether they see intermediate results. However, display of intermediate results substantially impacts student confidence. A statistically significant number of students who initially had correct answers and did not change had a higher value code assigned to their explanations after group discussion.

2.2 Teaching and Learning with Peer Instruction

Chemistry educators have been concerned about the discrepancy between conceptual understanding and problem solving by routine for decades (Nurrenbern & Pickering, 1987; Sawrey, 1990). Indeed, many science classes use lecture-based instructional delivery and emphasize routine problem-solving skills. This type of instruction can reinforce rote learning rather than conceptual understanding (Elby, 1999; Watters & Watters, 2007). In one study, chemistry students demonstrated a 95% success rate on numerical problems as compared to a 38% rate on similar conceptual problems (Haláková & Prokša, 2007). Similarly, a study of physics students showed a 70% success rate on a numerical problem but only a 10% success rate on a conceptual problem on the same concept (McDermott, 2001). It has been shown that this lack of conceptual understanding severely restricts students' ability to solve different types of problems based on the same concepts because they do not have the functional understanding to use their knowledge in new situations (Bransford, Brown & Cocking, 2000). Active, student-centered learning environments are more effective than traditional lecture based methods at promoting conceptual understanding (Hake, 1998; Poulis, Massen, Robens & Gilbert, 1998; Prince, 2004). In this study, the ability of students to learn from their peers is examined as part of one active learning technique, peer instruction.

Peer instruction is a structured questioning process that actively involves all students in the class (Mazur, 1997; Mazur, 2009; Smith et al, 2009). In this technique, a multiple-choice "conceptest" or "clicker" question is presented to the class. The

class first answers the question individually. Depending on the aggregate response, students can be encouraged to discuss the answer in small groups and then individually submit a final answer. This sequence is then typically followed by a class-wide discussion. In this way, the instructor can dynamically adjust the pace and extent of coverage to match student learning. Peer instruction prompts students to actively engage in their own learning, to think critically about the material during class, and to learn from and teach each other.

Researchers have sought to establish the effectiveness of peer instruction. Most commonly, multiple-choice pre- and posttests, typically administered at the start and end of the term (often using reliable and valid concept inventories), are used to compare learning gains in courses taught using peer instruction with more traditional pedagogies (Van Dijk, Van Der Berg & Van Keulen, 2001; Crouch & Mazur, 2001; Kalman, Milner-Bolotin & Antimirova, 2010). In one study of 30 introductory physics classes across 11 universities, average normalized gains (the ratio of the gain achieved to the maximum gain possible) of 0.39 were measured (Crouch, Watkins, Fagan & Mazur, 2007). While these studies offer compelling evidence that peer instruction results in improved learning, they integrate the effects over the entire term, and do not specifically elucidate the changes students undergo in a specific instance. A few researchers have specifically compared students' answers to conceptual, multiple-choice questions before and after group discussion (Smith et al, 2009; Singh, 2005), and infer that group discussion is beneficial by examining differences in multiple-

choice answers. Both studies found cases in which the correct answer was identified as a result of group discussion when all the individuals were initially wrong.

In this study, we seek to more extensively examine changes in student thinking that result directly from group discussion. One approach could be to follow specific groups using protocol analysis; however, this approach is labor intensive, and it would be difficult to cover the entire cohort. In this study, we take another approach.

Reflective written explanations are collected as students answer questions individually before group discussion, and then as they individually answer after group discussion, tracking each member within each group. The observation of changes from members within specific groups is intended to identify specifically the effect of student-student interaction on the use of language and conceptual development. We can consider an analogy to a chemical process. Instead of inferring a mechanism from measuring the response of output to changes in inputs, we are measuring properties during the process itself. Admittedly, the act of measuring the process can change the process itself. Thus, differences promoted by this additional reflection activity should be kept in mind when interpreting the results.

The peer instruction method formulated by Mazur (1997) does not specify the instructor to display an intermediate bar graph that shows a histogram of individual student answers to the entire class before group discussion. However, in many instances, the delivery of peer instruction in the classroom is described explicitly with the display of such intermediate results (Van Dijk, Van Der Berg & Van Keulen, 2001; Kalman, Milner-Bolotin & Antimirova, 2010; Caldwell, 2007; Kennedy &

Cutts, 2005). Instructor uncertainty about the benefits of displaying the intermediate result is illustrated by one case in which the intermediate results were sometimes shown to the class and were not at other times, although no systematic method was used to decide (Simon, Kohanfars, Lee, Tamayo & Cutts, 2010). One might argue that the intermediate display of results allows students to monitor their understanding in relation to the entire class, and, thereby, provides them reassurance even when they are wrong; on the other hand, students' desire to conform with the group may prompt them to adopt the consensus, even when they believe the group is wrong. One intent of the study presented in this paper is to ascertain how the display of the initial individual multiple-choice response distribution to the class affects the answer choice, written reflections, and confidence of students after group discussion.

Students' self-confidence in their knowledge can significantly affect how they interact and perform in the classroom (Bowen & Roth, 1999). Confidence is an important factor in their ability to monitor their own thinking, which is one aspect of metacognition (Schraw, Dunkle, Bendixen & Roedel, 1995). Koriat (2008) argues that confidence judgments are based on inferences from cues that can either constitute knowledge or reside in the feedback from task performance. The impact of peer instruction and other similar pedagogies has been reported to positively impact student confidence. Student interviews (Dufresne, Gerace, Leonard, Mestre & Wenk, 1996) and surveys (Nayak & Erinjeri, 2008) indicate that students gain confidence by being able to see where their answers fit within the responses of the other students in the class. However, association of student confidence in the correctness of their answers

directly to specific questions is rare, perhaps in part, because many classroom response systems do not have that feature (Barber & Njus, 2007). A few researchers have reported the result of confidence to specific questions in similar types of questions on concept inventories (Allen, 2006; Lasry & Aulls, 2007). The study described in this paper measures students' self-reported confidence directly following the cognitive act of answering the question. We seek to see how student confidence might influence the content of group discussion, and, reflexively, how the group discussion changes confidence.

This study seeks to relate the reflective cognitions of students before and after group discussion to their multiple-choice responses during peer instruction and to their confidence in their answers. Specifically, the research questions are:

- Does group discussion during peer instruction help students develop a more explicit understanding of difficult concepts in chemical thermodynamics?
- How does the display of the initial individual multiple-choice response distributions to the class affect the answer choice and confidence of students after group discussion?

2.3 Methods

This study spans two years in the second term of a junior-level undergraduate chemical thermodynamics course at a large public university. Each year 64 students participated in the study. The research was approved by the Institutional Review Board and participants signed informed consent forms. The course is a required course for chemical engineers and taken as an elective by a small number of biological and

environmental engineers. Therefore, each cohort has had similar programmatic experiences across the two-and-a-half previous years of the curriculum. However, it is not possible to characterize the equivalence of the two cohorts in detail, and results of this study should be interpreted with that in mind.

The Web-based interactive science and engineering (WISE) learning tool is used to collect student responses, such as the multiple-choice answers, written explanations, and confidence ratings used in this study (Koretsky & Brooks, 2008). WISE is enabled through a wireless laptop initiative, which mandates that every student own a laptop computer. In the class studied in this paper, WISE was used once a week in the 2-h recitations that the entire class attended. This study reports results when the peer instruction technique was used; however, other types of technology-driven active pedagogies were used during these recitation sections as well.

Conceptual exercises were assigned using WISE. Students were first asked to answer individually, without consulting their neighbors or the instructor. For the questions analyzed in this study, students chose a multiple-choice answer, wrote a short answer explanation, and reported their confidence using a Likert scale of 1–5 where 1 is “substantially unsure,” and 5 is “substantially confident”. In the first year (Cohort A), a distribution of the responses was displayed in bar graph form after the question was initially answered; in the second year (Cohort B), this “intermediate” bar graph was omitted. After answering individually, the class was encouraged to self-select groups of two or three students to discuss their answers. During this group discussion, the instructor did not interact with the groups directly except to answer

general questions in front of the entire class. Therefore, the responses were co-constructed within the student group. After group discussion, the exercise was assigned again and students again responded individually. In addition to the responses above, students identified the members of their peer-instruction group. In all cases, a class-wide discussion followed in which the instructor sought to provide closure and to model ways of thinking, forms of speech, and proper terminology. While we believe this last step is a critical step in student learning, it is not addressed in the current study.

Students received full credit for participating, but also received extra-credit for correct answers. It has been shown grading incentive can affect student discourse in peer instruction, and it has been suggested that discussion is suppressed in a “high stakes” environment (James, 2006; Willoughby & Gustafson, 2009). This approach is intended to both encourage student discussion by decreasing the stakes, yet still provide incentive for students to respond correctly.

This study primarily reports analysis of a subset of five question pairs (10 exercises), which are labeled Throttling Valve, Spray Can, Mixing, Equilibrium, and Adiabatic Air. Two of the questions were modified from domain-specific concept inventories (Miller, Streveler, Olds, Nelson & Geist, 2005; Midkiff, Litzinger, & Evans, 2001), validated and reliable multiple-choice instruments that explore students’ conceptual understanding in a given subject area. One of the questions was modified from work that developed conceptests in chemical thermodynamics (Falconer, 2007), and two questions were developed by the authors. A variety of different conceptual

learning activities was used in this class, and questions or tasks vary from year to year. The five question pairs were chosen because they were delivered using the modified peer instruction method described above and were used in both years of the study and could, therefore, be compared. For Cohort B, two additional question pairs developed by the authors, Solid Liquid Melt and Chemical Reaction, were used to provide cases in which the proportion of correct answers was in between those of the five question pairs above. The content of the questions, as displayed to the class on their laptops, is presented in the Supporting Information (Appendix A).

In two cases, the second response set was not collected. For Cohort A, Adiabatic Air did not include this part because the majority of the class chose the correct answer, just like Cohort B, and like Cohort B, chose the correct answer for the wrong reason. The fact that they chose the correct answer for the wrong reason only became apparent during the analysis after class. For Cohort B, group discussion and follow-up were not used with Mixing. Instead, a different technique was used. Students were given different explanations from the previous year and asked to judge which one most effectively used evidence-based reasoning to make its case.

The average time for students to respond to a question initially was 7 min with a standard deviation of 2 min. The average time for discussion between responses was 7 min with a standard deviation of 7 min. The student response time after group discussion was 4 min with a standard deviation of 2 min. The time spent on each conceptual question is longer than is typical for peer instruction (Mazur, 1997), largely because students are asked to provide short-answer written explanations of why they

selected an answer. This method prompts the students to be reflective; they are encouraged to think about their reasons for an answer. It also provides insight into their thought processes and how those processes change with peer instruction. This difference should be kept in mind when considering the results presented in this paper.

The mixed methodological basis of this research is grounded in a phenomenological perspective of ascertaining how student multiple-choice and short-answer explanations reflect conceptual understanding, and how that understanding changes as a result of group discussion. Coding the free-response, short-answer explanations involved open coding, a process used to infer categories of meaning, using a technique similar to that of Newcomer and Steif (2008) in their analysis of written explanations to a concept question in statics. The process involves proposing a code, coding individually, comparing among the coders, modifying the code, and repeating until convergence. Three researchers, including a thermodynamics textbook author, participated in this process. The validity of the codes was further verified by comparing specific questions' codes for consistency to misconceptions on the same concept reported in the literature (Barker & Millar, 2000; Clark, 2006; Loverude, Kautz & Heron, 2002; Meltzer, 2004; Rozier & Viennot, 1991; Sanger & Badger, 2001). A hierarchical coding scheme was created for each exercise that incorporates and ranks the important concepts and misconceptions. The specific codes are included in the Supporting Information (Appendix A).

Two graduate student researchers coded the written explanations. Both researchers have undergraduate degrees in chemical engineering, are former

thermodynamics teaching assistants, and are pursuing Ph.D. degrees focusing on science and engineering education. The interrater reliability using the Cohen's κ statistic is 0.80, indicating reasonable agreement. Less than 5% of the coded responses had a discrepancy greater than one code value. The statistical significance of change in multiple-choice responses or coded quality of response is reported using a nonparametric sign test. The statistical significance of factors that affect self-reported confidence is analyzed using multifactor analysis of variance (ANOVA).

2.4 Results

2.4.1 Multiple-Choice Responses

Tables 2.1 and 2.2 show a summary of the distribution of multiple-choice responses for the primary exercise pairs for the two cohorts. It includes the percentage correct and the percentage of students who chose each incorrect response. The exercises labeled "pre" are based on the initial individual student responses and those labeled "post" are individual responses after group discussion. As discussed earlier, group discussion did not follow the initial response for two of the exercises, so only "pre" data are shown. For Throttling Valve and Equilibrium, most students in both cohorts initially chose the same incorrect answer, which is labeled "Wrong A, %"; for Spray Can, the majority of Cohort A initially chose the correct answer but the majority of Cohort B initially chose the same wrong answer. For both Adiabatic Air (Cohort B) and Mixing (Cohort A), the majority initially chose the correct multiple-choice answer. The sets with a most popular wrong answer are labeled "consensually wrong" and the ones with a most popular correct answer are labeled "consensually correct."

Table 2.1. Cohort A Answer Summary for Students Who Saw Results Displayed before Group Discussion

Answer Disposition	Exercise	Correct Answer, %	Wrong A, %	Wrong B, %	Wrong C, %
Consensually Wrong	Throttling valve - pre	13	61	25	1
	Throttling valve - post	10	81	9	0
	Equilibrium - pre	10	48	38	5
	Equilibrium - post	7	60	33	0
Consensually Correct	Adiabatic Air - pre	89	7	3	1
	Spray Can - pre	56	31	9	4
	Spray Can - post	76	24	0	0
	Mixing - pre	63	31	4	1
	Mixing - post	94	5	2	0

Table 2.2. Cohort B Answer Summary for Students Who Did *Not* See Results Displayed before Group Discussion

Answer Disposition	Exercise	Correct Answer, %	Wrong A, %	Wrong B, %	Wrong C, %
Consensually Wrong	Throttling Valve - pre	13	67	18	2
	Throttling Valve - post	27	71	2	0
	Equilibrium - pre	3	48	47	2
	Equilibrium - post	2	74	23	2
	Spray Can - pre	45	50	3	2
	Spray Can - post	20	79	2	0
Consensually Correct	Mixing - pre	82	10	5	3
	Adiabatic Air - pre	88	7	3	2
	Adiabatic Air - post	98	2	0	0

2.4.2 Change in Multiple-Choice Response after Group Discussion

Changes in multiple-choice responses as a result of group discussion for each cohort are presented in Figures 2.1 and 2.2. In Figure 2.1, the percentage of those students who changed their answers and changed from correct to incorrect (labeled “from correct”) or from incorrect to correct (labeled “to correct”) for each exercise pair is shown. Sign tests show statistical significance in the difference in number

between “to correct” and “from correct” for the Throttling Valve (Cohort B, $n = 13$, $x = 2$, $p = 0.012$), Spray Can (Cohort A, $n = 20$, $x = 3$, $p = 0.002$; Cohort B, $n = 18$, $x = 2$, $p = 0.001$), Mixing (Cohort A, $n = 25$, $x = 3$, $p = 0.000$), and Adiabatic Air (Cohort B, $n = 6$, $x = 0$, $p = 0.016$) exercises. In the statistical analysis, n is the total number that were changed after subtracting the number of pairs that did not change; x is the number that increased or the number that decreased, whichever is smaller; and p is the confidence level that the null hypothesis (the group discussion had no effect) can be rejected. In all cases when the answer was consensually correct, a statistically significant number of students changed their answer from incorrect to correct. When the answer was consensually wrong, there is no clear pattern.

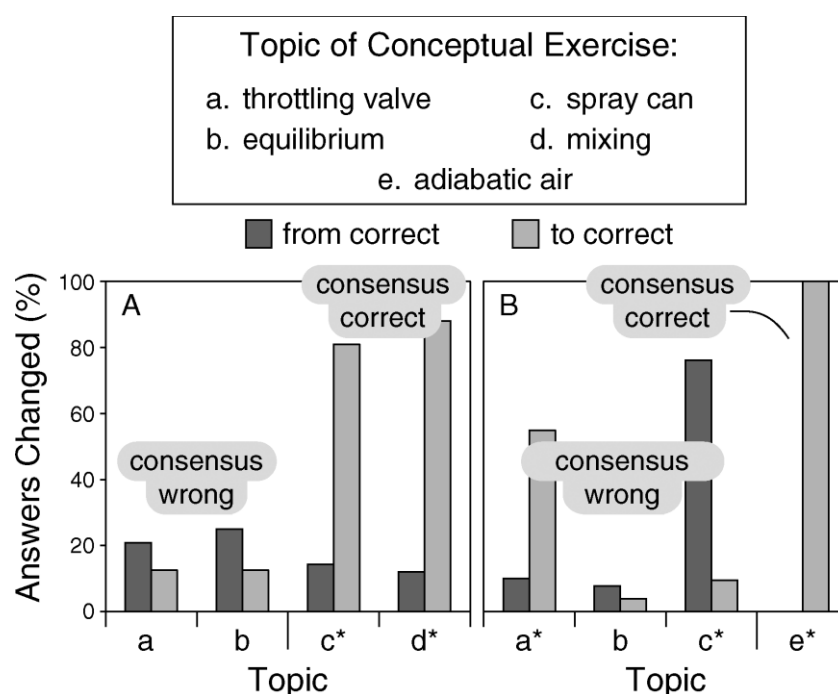


Figure 2.1. Percentage of students who changed from the correct multiple-choice answer to an incorrect answer (% from correct) and the percentage change from an incorrect answer to the correct answer (% to correct). (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion (* $p < 0.05$).

Similarly, Figure 2.2 shows the changes from the consensus answer (labeled “from consensus”) and to the consensus multiple-choice answer (labeled “to consensus”). In all cases, more students changed “to consensus” than “from consensus”. Sign tests show significance for changes to the consensus answer for three of the consensually wrong exercises—Throttling Valve (Cohort A, $n = 21$, $x = 3$, $p = 0.001$), Equilibrium (Cohort B, $n = 25$, $x = 5$, $p = 0.002$), and Spray Can (Cohort B, $n = 20$, $x = 2$, $p = 0.001$)—and for the consensually correct exercises, Spray Can (Cohort A, $n = 20$, $x = 3$, $p = 0.002$), Mixing (Cohort A, $n = 25$, $x = 3$, $p = 0.000$), and Adiabatic Air (Cohort B, $n = 6$, $x = 0$, $p = 0.016$). While the changes to Equilibrium for Cohort A were not statistically significant to 95% confidence (Cohort A, $n = 14$, $x = 4$, $p = 0.090$), over twice the number of students switched “to consensus” than “from consensus”, even though the “to consensus” answer was incorrect. Essentially, students tend to change to the consensus answer after group discussion, regardless of whether that answer is correct or not or whether the intermediate results were displayed to the class.

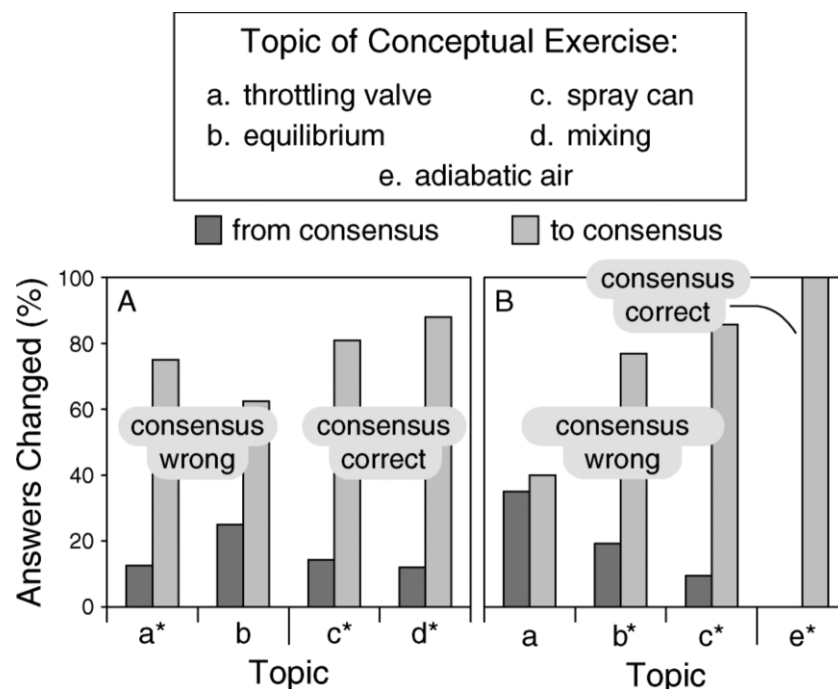


Figure 2.2. Percentage of students who changed from the popular multiple-choice answer to the unpopular answer (% from consensus) and the percentage change from an unpopular answer to the popular answer (% to consensus). (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion (* $p < 0.05$).

2.4.3 Change in Codes of Written Explanations after Group Discussion

Figure 2.3 shows the number of written explanations by exercise pair that improved (+ code Δ) or declined (- code Δ) in code value after group discussion. All exercises except Spray Can for Cohort B had more students improve than decline. All of the consensually correct exercises gave consistent, statistically significant gains: Spray Can (Cohort A, $n = 25$, $x = 2$, $p = 0.000$), Mixing (Cohort A, $n = 28$, $x = 9$, $p = 0.044$), and Adiabatic Air (Cohort B, $n = 16$, $x = 2$, $p = 0.003$). Cohort B's Spray Can (Cohort B, $n = 18$, $x = 5$, $p = 0.049$), a consensually wrong exercise, had a statistically significant decline, while Equilibrium had a statistically significant increase (Cohort

B, $n = 22$, $x = 4$, $p = 0.003$). The explanation-coding scheme is expected to correlate with correctness of multiple-choice responses, because students with richer conceptual explanations are more likely to get the problem correct. After group discussion on the consensually wrong exercises,—Throttling Valve, Equilibrium, and Spray Can (Cohort B)—explanations for 28 of the 36 correct multiple-choice responses were coded at the highest level (a value of 4).

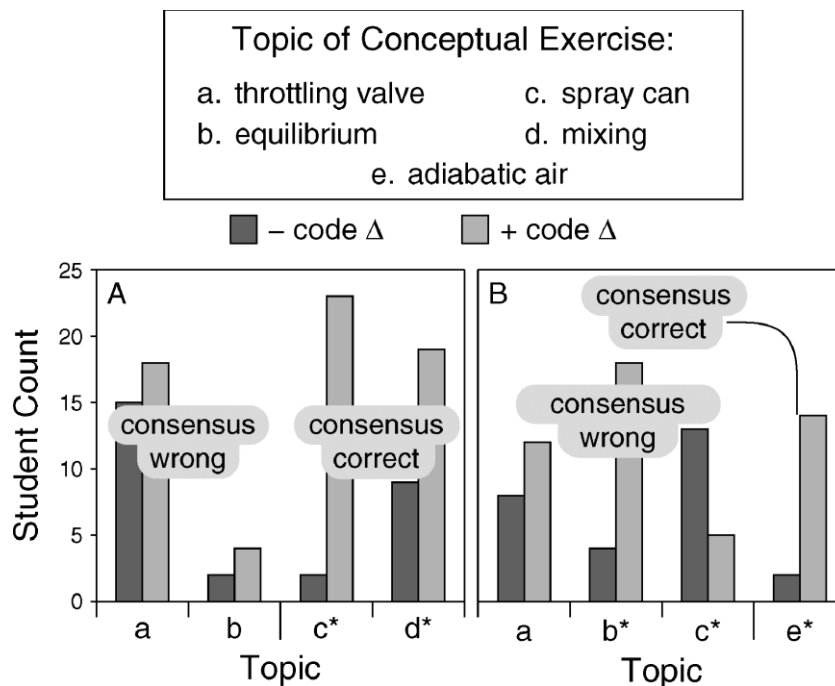


Figure 2.3. Number of students per exercise whose explanations improved (+ code Δ) or got worse (-code Δ). (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion (* $p < 0.05$).

Figure 2.4 aggregates the coded written explanation of all question pairs for each cohort and organizes them by the correctness of the multiple-choice answers, before and after group discussion. Therefore, the categories become the number of code differences of students who: selected the correct multiple-choice answer both

before and after group discussion (labeled “correct twice”), selected the correct multiple-choice answer initially then incorrect answer after group discussion (labeled “correct to wrong”), selected the wrong answer initially and correct answer after (labeled “wrong to correct”), and had incorrect answer choices both before and after (labeled “wrong twice”). Significance can be shown for the differences in code changes (+ code Δ , - code Δ) in three of the four categories for both cohorts. Correct twice (Cohort A, $n = 27, x = 7, p = 0.010$; Cohort B, $n = 18, x = 4, p = 0.015$) and incorrect to correct (Cohort A, $n = 22, x = 0, p = 0.000$; Cohort B, $n = 11, x = 0, p = 0.001$) had more gains in code ratings (+ code Δ). Correct to incorrect (Cohort A, $n = 9, x = 0, p = 0.002$; Cohort B, $n = 30, x = 11, p = 0.001$) had a decline in code rating (- code Δ). While the incorrect twice was not statistically significant when each cohort was considered separately, when the data from both years are aggregated, it is also significant ($n = 65, x = 24, p = 0.023$).

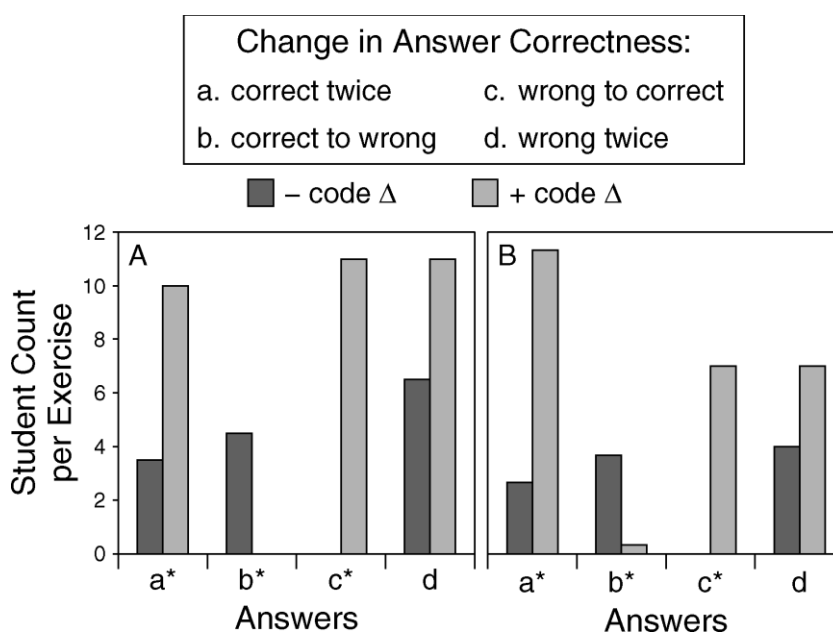


Figure 2.4. Average number of students per exercise whose explanations improved (+ code Δ) or got worse (code Δ) organized by their multiple-choice answer change. (A) The cohort when the results were displayed before group discussion. (B) The cohort when the results were not displayed before group discussion: combination of the results for the consensually correct and the consensually wrong exercises (* $p < 0.05$).

2.4.4 Student Confidence before and after Group Discussion

Figure 2.5 shows the average reported confidence for incorrect and correct multiple-choice answers for each cohort, grouped by exercise category of consensually correct (CC) and consensually wrong (CW). Statistical significance is determined by multifactor ANOVA. Panels A and B in Figure 5 show results before group discussion. In both cases, there is no significant correlation between confidence and either correctness or consensuality. Panels C and D in Figure 5 show results after group discussion. The results for Cohort A to whom the intermediate results were shown (Figure 5C) show a correlation between confidence and consensuality, as

$F(1,169) = 16.49$, $MSE = 12.36$, $p = 0.000$. Those students who selected the consensus answer were more confident after group discussion, whether it was correct or not. Conversely, those students who had the *correct* nonconsensus answer and saw the intermediate results were the least confident of all cases in Figure 2.5. A multifactor ANOVA analysis was performed considering main effects and interactions for the entire data set (see the Supporting Information in Appendix A). Consensuality, correctness, and pre/post group discussion show an influence on reported confidence. Two significant interactions, the display of an intermediate bar graph with pre/post group discussion and consensuality with correctness, influence student confidence as well.

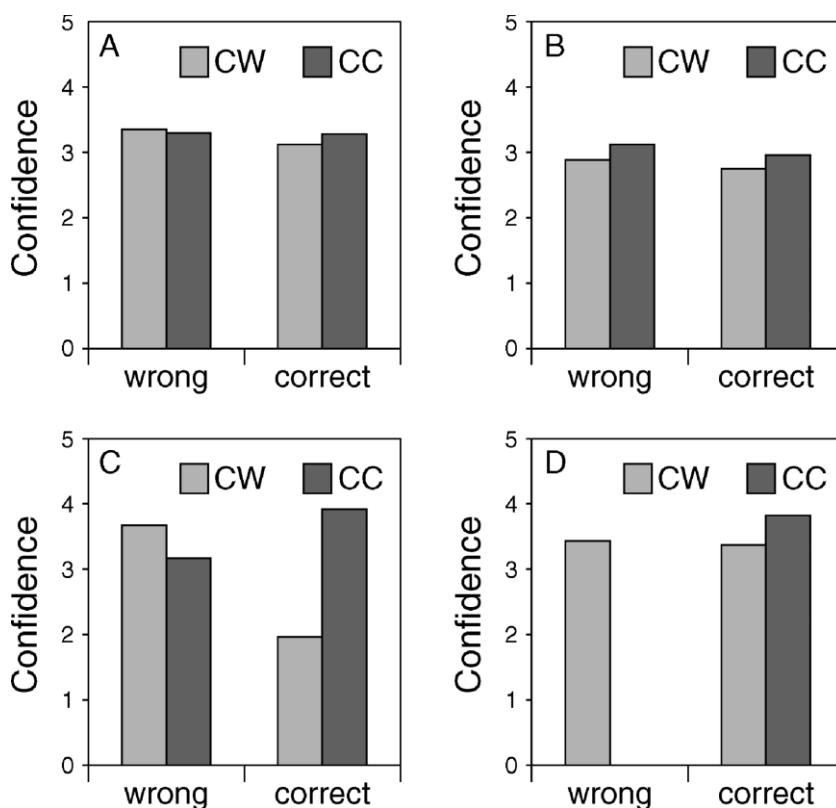


Figure 2.5. Confidence vs correctness, for consensually wrong (CW) and consensually correct (CC) exercises. (A) Before group discussion (pre), with intermediate results shown. (B) Before group discussion (pre), without intermediate results shown. (C) After group discussion (post), with intermediate results shown. (D) After group discussion (post), without intermediate results shown.

2.4.4 Confirmation of Consensuality-Correctness Trend

For Cohort B (no intermediate bar graph), results from two additional questions are reported, labeled Solid-Liquid Melt and Chemical Reaction. Table 2.3 shows a summary of the distribution of multiple-choice responses. Neither question had an initial majority, but the correct answer in each case was the most popular. Sign tests show statistical significance in the difference in number between “to correct” and “from correct” and to changes to the consensus answer for both exercises: Solid-

Liquid Melt (Cohort B, $n = 18$, $x = 2$, $p = 0.001$), and Chemical Reaction (Cohort B, $n = 31$, $x = 2$, $p = 0.000$).

Table 2.3. Answer Summary for Questions in Which the Most Popular Answer Was Correct, yet No Answer Had a Majority

Exercise	Correct %	Wrong A%	Wrong B%	Wrong C%
Solid-Liquid Melt, Pre	39	30	18	14
Solid-Liquid Melt, Post	64	24	7	5
Chemical Reaction, Pre ^a	36	34	12	7
Chemical Reaction, Post	85	13	2	0

^aThese numbers do not add up to 100% because there were additional minor wrong D and E answers selected for this question.

2.5 Discussion

A statistically significant number of students who originally had the correct multiple-choice answer (Figure 2.4, correct twice, Cohorts A and B) had a higher value code assigned to their explanations after group discussion, and therefore demonstrated more explicit understanding of difficult concepts in chemical thermodynamics. This result suggests that, in the manner studied here, the group discussion during peer instruction can help students develop richer explanations, even though they originally selected the correct multiple-choice answer. Such a result is anticipated by Chi et al. (1994) who show that the active process of explaining encourages students to integrate new knowledge with existing knowledge and leads to richer conceptual understanding. Similarly, a statistically significant number of students who had the wrong initial answer improved in the coded-value of their response (wrong to correct, Cohorts A and B; wrong twice, aggregate). Only students who changed the correct answer to the wrong answer show lower value codes. This

latter type of change is more commonly associated with consensually wrong answers, whether or not the intermediate results are displayed to the class (Figures 2.1 and 2.2).

One interpretation of these results is that when the answer is consensually wrong, the gains realized during group discussion by peer instruction may be limited. Therefore, it may be better to use peer instruction with consensually correct exercises because students are more likely to adopt the correct answer after group discussion. However, inspection of Figure 2.3 shows that, even in some cases that are consensually wrong, like for Equilibrium, the explanations more often get better even though the number of students choosing the right answer goes down. These two measures give conflicting data on learning. Results of studies in the literature that solely examine the correctness of the multiple-choice response need to be interpreted with this in mind. Additionally, it should be remembered that our study only probes an intermediate state in the peer-instruction process. The group discussion and re-answer is typically followed by class-wide discussion. It may be that this final stage of instruction is especially important for consensually wrong exercises.

The two questions shown in Table 2.3, Solid-Liquid Melt and Chemical Reaction, show large gains in the percentage of students who chose correctly. While care must be taken to generalize from these limited results, there appears to be a threshold where the class tends toward the correct answer. In constructing conceptual questions, and in providing context before the peer-instruction activity, it appears that the closer an instructor can come to such a threshold, but not cross it, the greater the student learning.

In comparing the pedagogical techniques of peer instruction and class-wide discussion, Nicol and Boyle (2003) speculate that peer instruction is more effective because students are more likely to encounter conceptual conflict in smaller groups, and therefore are more likely to actively construct mental models. Conversely, it seems worthwhile to examine what conditions cause students to acquiesce and adapt the popular interpretation in peer instruction.

The tendency for students to select the most popular answer after group discussion could have several simultaneous root causes. Early studies in social psychology show that subjects commonly switch answers to the views of the majority or of experts, even when they believe it is wrong (Asch, 1955). Perez et al. (2010) hypothesize that the display of a bar graph of student responses during peer instruction leads students to an unconscious desire to conform, and that this bias toward the consensus answer lowers the quality of group discussion. While such an explanation would clearly apply to the case where a bar graph of intermediate results is shown as for Cohort A, Cohort B demonstrated an approximately equivalent tendency to switch to the consensus answer even though they were not shown the intermediate results (Figure 2.2). Bargh (2006) has shown that more subtle stimuli in the environment can activate content in memory and make related constructs more accessible. Other factors, such as the local majorities and the social dynamics within a discussion group, can also contribute. Such factors can activate common misconceptions of the students. It follows that effective environments for peer instruction are those that recognize such stimuli and are designed to more fully engage students in the cognitive conflict elicited

by a well-designed question. More research is needed to specify the design of these settings, both in the context of the sociocultural aspects of the class and in the misconceptions associated with the specific domain content.

When comparing Cohort A, who saw a bar graph of the intermediate results of the entire class before group discussion, and the Cohort B, who did not see intermediate results, no clear differences are observed in the patterns by which answers changed or in the changes in quality of students' written reflections. Conversely, a distinctive difference between the two cohorts is observed in their self-reported confidence after group discussion (Figure 2.5). Students who see the intermediate results are significantly more confident if they agree with the majority of their peers and chose the consensus answer than those who disagree, regardless of whether it is correct or wrong.

The consensuality principle proposed by Koriat (2008) states that metacognitive judgments such as confidence are influenced by factors that make it compelling to the majority and these judgments are independent of correctness. For the case of consensually wrong responses, he states "metacognitive judgments will be counterdiagnostic of the correctness of the answer" (p. 954). In this study, clearly the class-wide display of the initial multiple-choice responses for Cohort A impacts student confidence. Koriat (2008) distinguishes between *information-based* and *experience-based* metacognitive judgments. Information-based judgments result from domain-specific knowledge and direct evidence-based reasoning, while experience-based judgments "rely on heuristics that make use of internal, mnemonic cues" (p.

955). We conjecture that student answers characterized by experience-based judgments are less likely to be correct and more likely to contain misconceptions.

Such a framework is consistent with the novice–expert literature in which experts tend to be slower at problem-scoping than novices (Chi, Feltovich & Glaser, 1981; Steif, Lobue, Fay & Kara, 2010). One objective of pedagogies such as peer instruction is to help students develop from experience-based judgments to information-based reasoning.

In this study, students who chose the correct, but nonconsensus answers are the least sure about the correctness of their response (Figure 2.5). However, this latter set is mostly comprised of students who have strong conceptual knowledge and stick by their correct answer to these difficult questions over the popular answer, despite their lack of confidence (78% of the corresponding explanations were coded at the highest level). It is proposed that these students are generally making information-based judgments that are logically grounded and less susceptible of misconceptions.

Additionally, Hattie and Timperley (2007) argue that the most desirable circumstance for students to be receptive to feedback is when they have high confidence but are incorrect. It is unclear whether the display of the intermediate bar graph during peer instruction benefits learning. On one hand, students' desire to conform to the group might lower their engagement and discourage information-based interpretations. On the other hand, display of intermediate results can provide opportunities for enhanced learning if instructors properly address these issues in class-wide discussion following the second round of answers. Specifically, through directed, context-related coaching,

the instructor can shift student responses from experience-based judgment to information-based reasoning. It would be useful to study student cognition after class-wide discussion, especially for consensually wrong exercises in which the instructors deliberately try to encourage information-based thinking.

2.6 Conclusion

This study examined changes in student thinking that result directly from group discussion during a modified form of peer instruction by collecting reflective written explanations and ratings of self-confidence in addition to multiple-choice responses. Students tended to pick the consensually popular answer, whether it was correct or incorrect, and whether an intermediate result was shown to the entire class. However, students commonly improved the quality of their written answer as a result of group discussion. On the other hand, showing students the intermediate results significantly affects their confidence. The results from this study suggest that the most effective questions have a high number of incorrect responses, but a correct answer that is most common. Instructors using this technique are also encouraged to aid students to transition from experience-based judgment to information-based reasoning.

2.7 Supporting Information

Coding scheme for analysis of conceptual exercises; conceptual exercises used in this study; results from the ANOVA of student confidence. This material is available in Appendix A and via the Internet at <http://pubs.acs.org>.

2.8 Acknowledgments

The authors acknowledge financial support from a LL Stewart Faculty Scholar Award and a Teaching and Learning Grant from the Center for Teaching and Learning. We are grateful to Debra Gilbuena for her contributions to the coding methodology and to Michael Prince, Ronald Miller, and John Falconer for feedback on the paper.

2.9 References

- Allen, K. (2006). The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics. *Available at SSRN 2130143*.
- Asch, S. E. (1955). Opinions and social pressure. *Readings about the social animal*, 17-26.
- Barber, M., & Njus, D. (2007). Clicker evolution: seeking intelligent design. *CBE-Life Sciences Education*, 6(1), 1-8.
- Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36(2), 147-168.
- Barker, V., & Millar, R. (2000). Students' reasoning about basic chemical thermodynamics and chemical bonding: what changes occur during a context-based post-16 chemistry course? *International Journal of Science Education*, 22(11), 1171-1200.
- Bowen, G. M., Roth, W. M., & McGinn, M. K. (1999). Interpretations of graphs by university biology students and practicing scientists: Toward a social practice view of scientific representation practices. *Journal of Research in Science Teaching*, 36(9), 1020-1043.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*: National Academy Press Washington, DC.
- Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. *CBE-Life Sciences Education*, 6(1), 9-20.
- Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3), 439-477.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science*, 5(2), 121-152.
- Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, 24(4), 467-563.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69, 970.

- Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics*, 1(1), 40-95.
- Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of computing in higher education*, 7(2), 3-47.
- Elby, A. (1999). Another reason that physics students learn by rote. *American Journal of Physics*, 67, S52.
- Falconer, J. L. (2007). Concepttests for a Chemical Engineering Thermodynamics Course. *Chemical Engineering Education*, 41(2), 107-114.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66, 64.
- Haláková, Z., & Prokša, M. (2007). Two kinds of conceptual problems in chemistry teaching. *Journal of chemical education*, 84(1), 172.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. doi: 10.3102/003465430298487
- James, M. C. (2006). The effect of grading incentive on student discourse in peer instruction. *American journal of physics*, 74, 689.
- Kalman, C. S., Milner-Bolotin, M., & Antimirova, T. (2010). Comparison of the effectiveness of collaborative groups and peer instruction in a large introductory physics course for science majors. *Canadian Journal of Physics*, 88(5), 325-332.
- Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning*, 21(4), 260-268.
- Koretsky, M., & Brooks, B. (2008). *A Web-based Interactive Science and Engineering Learning Tool that Promotes Concept-Based Instruction*. Paper presented at the Proceedings of the Annual Conference of the American Society for Engineering Education.
- Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 945.
- Lasry, N., & Aulls, M. W. (2007). The effect of multiple internal representations on context-rich instruction. *American Journal of Physics*, 75, 1030.
- Loverude, M. E., Kautz, C. H., & Heron, P. R. (2002). Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas. *American Journal of Physics*, 70, 137.
- Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ 07458: Prentice Hall.
- Mazur, E. (2009). Farewell, lecture. *Science*, 323(5910), 50-51.
- McDermott, L. C. (2001). Oersted Medal Lecture 2001: "Physics Education Research—The Key to Student Learning". *American Journal of Physics*, 69, 1127.

- Meltzer, D. E. (2004). Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course. *American Journal of Physics*, 72, 1432.
- Midkiff, K. C., Litzinger, T. A., & Evans, D. (2001). *Development of engineering thermodynamics concept inventory instruments*. Paper presented at the Frontiers in Education Conference, 2001. 31st Annual.
- Miller, R. L., Streveler, R. A., Nelson, M. A., Geist, M. R., & Olds, B. M. (2005). *Concept inventories meet cognitive psychology: Using beta testing as a mechanism for identifying engineering student misconceptions*. Paper presented at the Proceedings of the American Society for Engineering Education Annual Conference.
- Nayak, L., & Erinjeri, J. P. (2008). Audience response systems in medical student education benefit learners and presenters. *Academic Radiology*, 15(3), 383-389.
- Newcomer, J. L., & Steif, P. S. (2008). Student thinking about static equilibrium: Insights from written explanations to a concept question. *Journal of Engineering Education*, 97(4), 481-490.
- Nicol, D. J., & Boyle, J. T. (2003). Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education*, 28(4), 457-473.
- Nurrenbern, S. C., & Pickering, M. (1987). Concept learning versus problem solving: Is there a difference? *Journal of Chemical Education*, 64(6), 508.
- Perez, K. E., Strauss, E. A., Downey, N., Galbraith, A., Jeanne, R., & Cooper, S. (2010). Does displaying the class results affect student discussion during peer instruction? *CBE-Life Sciences Education*, 9(2), 133-140.
- Poulis, J., Massen, C., Robens, E., & Gilbert, M. (1998). Physics lecturing with audience paced feedback. *American Journal of Physics*, 66, 439.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, 93(3), 223-231.
- Rozier, S., & Viennot, L. (1991). Students' reasonings in thermodynamics. *International Journal of Science Education*, 13(2), 159-170.
- Sanger, M. J., & Badger, S. M. (2001). Using computer-based visualization strategies to improve students' understanding of molecular polarity and miscibility. *Journal of Chemical Education*, 78(10), 1412.
- Sawrey, B. A. (1990). Concept learning versus problem solving: Revisited. *Journal of Chemical Education*, 67(3), 253.
- Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, 87(3), 433.
- Simon, B., Kohanfars, M., Lee, J., Tamayo, K., & Cutts, Q. (2010). *Experience report: Peer instruction in introductory computing*. Paper presented at the Proceedings of the 41st ACM technical symposium on Computer science education.
- Singh, C. (2005). Impact of peer interaction on conceptual test performance. *American journal of physics*, 73, 446.

- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122-124.
- Steif, P. S., Lobue, J. M., Kara, L. B., & Fay, A. L. (2010). Improving problem solving performance by inducing talk about salient problem features. *Journal of Engineering Education*, 99(2), 135-142.
- Van Dijk, L., Van Der Berg, G., & Van Keulen, H. (2001). Interactive lectures in engineering education. *European Journal of Engineering Education*, 26(1), 15-28.
- Watters, D. J., & Watters, J. J. (2007). Approaches to learning by students in the biological sciences: Implications for teaching. *International Journal of Science Education*, 29(1), 19-43.
- Willoughby, S. D., & Gustafson, E. (2009). Technology talks: Clickers and grading incentive in the large lecture hall. *American Journal of Physics*, 77, 180.

3. Promoting Scientific Reasoning through Written Explanations to Multiple-Choice Concept Questions

Bill J. Brooks, Rachel M. White, Alec S. Bowen, Adam Z. Higgins & Milo D.
Koretsky

School of Chemical, Biological, and Environmental Engineering

Oregon State University

Corvallis, OR, USA 97331-2702

(541)737-4591

Journal of Engineering Education

<http://www.asee.org/papers-and-publications/publications/jee>

(submitted for review)

3.1 Abstract

Background:

Increasingly, instructors of large, introductory STEM courses are having students actively engage during class by answering multiple choice concept questions individually and in groups. This study investigates the influence of prompting students to individually write explanations justifying their answer selection to these questions.

Purpose (Hypothesis):

We hypothesize that prompting students to explain and elaborate on their answer choices leads to greater focus and use of normative scientific reasoning processes and will therefore improve the way they respond to questions.

Design/Method:

We use a sequential, mixed methods experimental design. First, a crossover study is employed to determine the influence of asking students to individually provide written explanations (treatment condition) of their answer choices to 42 concept questions. We then apply emergent coding to written responses of a pair of isomorphic concept questions to further elucidate the relationships between answer choices, reasoning, and contextual differences in questions.

Results:

Results show that soliciting written explanations can have a significant influence on answer choice and when it does, that influence tends to be positive. However, for a subset of questions, students choose the correct answer less often. Qualitative analyses suggest that students are also activating more sophisticated

reasoning processes to these questions even though they are not selecting the correct answer.

Conclusions:

Instructors are encouraged to solicit written explanations to multiple-choice concept questions, look for cases where their students get the correct answer using incorrect reasoning, and seek ways to help their students recognize how core concepts apply to a broad range of contexts.

3.2 Introduction

This study investigates undergraduate student responses to conceptual, qualitative multiple-choice questions in thermodynamics. We term such questions concept questions. Researchers have shown that performance on concept questions improves when students discuss their answers verbally with peers after responding individually (Singh, 2005; Smith et al., 2009; Brooks & Koretsky, 2011; Koretsky & Brooks, 2011). Researchers have also hypothesized that soliciting written responses from students improves their understanding (Taylor & Nolen, 2007; Linn & Chiu, 2011) although there is little empirical support for this belief. This study used a crossover experimental design to determine the influence of asking students to individually provide written explanations of their answer choices to concept questions. We then contrast student answer choices and written responses to two similar (isomorphic) concept questions to further elucidate the relationships between answer choices, written explanations, and contextual differences in questions.

The results from this study can help practitioners understand the value of incorporating written explanations into active learning pedagogies that use concept questions. Written explanations can be used as a tool to both foster learning and concurrently assess that learning. Additionally, from a theoretical perspective, we seek to contribute to the understanding of the role of written explanations in students' reasoning processes as they respond to concept questions. Specifically, the research questions of this study include:

1. Does requesting students to provide written explanations for their answer choices to concept questions significantly influence the distribution of their answer choices in an undergraduate introductory thermodynamics class? Is it beneficial for thinking and performance?
2. Is the reasoning in the written responses consistent with their answer choices, that is, do students choose correct answers for the right reasons?

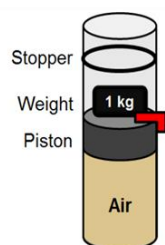
3.3 Background

Most students who can answer quantitative, algorithmic questions correctly in chemistry, physics, and engineering perform poorly when they are asked to answer concept questions on the very same topic (Papaphotis & Tsapalis, 2008; Haláková & Prokša, 2007; McDermott, 2001; Koretsky et al., 2011). VanLehn and van de Sande (2009) characterize the ability to answer concept questions correctly as “conceptual expertise” and assert, “novices often have too little practice of the right kind...” to develop conceptual expertise. Increasingly, instructors of large, introductory courses are reforming their instructional designs to attend to development towards conceptual

expertise. To provide students the opportunity for such development, instructors have students engage in answering concept questions during class using active learning pedagogies (Crouch & Mazur, 2001; Knight & Wood, 2005; McConnell et al., 2006). These pedagogies are often supported by technology such as audience response systems (Kay & LeSage, 2009).

Concept questions are designed to be conceptually challenging and typically require little or no computation so that students cannot algorithmically rely on equations to obtain the answer. They ideally focus on the most important concepts in a subject. Such questions extend assessment beyond "What does a student remember?" to "What does a student understand?" (Nurrenbern & Robinson, 1998). Specifically, 'good' concept questions should target a specific learning goal, uncover misconceptions, and elicit a range of responses (Caldwell, 2007; Crouch & Mazur, 2001; Tanner & Allen, 2005). Figure 3.1 shows a sample concept question as it was presented to students in class in the treatment condition in this study. To answer this question correctly, students need to apply their knowledge of the concept of conservation of energy, recognizing that the work done by the gas on the surroundings lowers its internal energy and, therefore, its temperature. They do not need to calculate any values, but rather reason through their answer choice based on appropriate engineering science principles. We term this logical procession "scientific reasoning."

Air at high pressure and ambient temperature is contained in a perfectly insulated piston-cylinder device. If the locks holding the piston in place are removed, the piston moves upwards to a stopper. The temperature of the air _____.



- increases
 remains the same
 decreases

Please explain your answer in the box below.

Please rate how confident you are with your answer.

- substantially unsure moderately unsure neutral moderately confident substantially confident

Submit

Figure 3.1. Expanding Piston question as it was presented in the treatment condition of Cohort B.

Concept questions with these characteristics comprise the items in concept inventories, such as the seminal force concept inventory (Hestenes, Wells & Swackhamer, 1992), and several concept inventories in general chemistry (Mulford & Robinson, 2002; Krause, Birk, Bauer, Jenkins & Pavelich, 2004), thermodynamics (Midkiff, Litzinger, & Evans, 2001; Vigeant, Prince, & Nottis, 2011), and other engineering topics (Imbrie & Reed-Rhoads, 2011). Concept questions are also a critical component in active learning pedagogies that use audience response systems

(e.g. clickers) for real-time, formative assessment (Mazur, 1997; Kay & LeSage, 2009). In the latter case, concept questions have been denoted as “ConcepTests” (Mazur, 1997) or “clicker questions” (Duncan, 2006).

Peer instruction is arguably the most well-known and widely used technology-mediated active learning pedagogy in post-secondary STEM classes (Mazur, 1997; Crouch, Watkins, Fagen & Mazur, 2007; Borrego, Cutler, Prince, Henderson & Froyd, 2013). It consists of a structured process described as follows. First, a concept question is presented to the class. Students answer the question individually. Vahey, Tatar, & Roschelle (2007) would characterize this individual response as a *private interaction*. They define private interactions as “those interactions in which students can engage with their materials and sense-making processes individually in a focused way.” The next step in peer instruction encourages students to discuss the answer choices in small groups. These group discussions could be considered *public interactions*, defined as “those interactions in which students engage in active or implicit discourse while they are simultaneously engaged with, or talking about, the product or materials of their work” (Vahey, Tatar & Roschelle, 2007). Next, students individually submit a final answer (another private interaction). Finally, this sequence is typically followed by a class wide discussion.

Studies of student performance during peer instruction show students’ frequency of correct answers usually increases after group discussion (Van Dijk, Van Der Berg & Van Keulen, 2001; Lasry, Charles, Whittaker & Lautman, 2009, Smith et al., 2011). These researchers generally link the improved performance to conceptual

learning and attribute it to the public interactions during the group discussion process between individual student responses. For example, groups are reported to demonstrate what Singh describes as “co-construction of knowledge” where students respond correctly after discussion even when none of the students in the group have the correct answer initially (Singh, 2005; Smith et al., 2009). While public interactions during peer instruction have been studied, little is known about the influences of the private interactions. Lucas (2009) argues that prompting students to first write individual explanations lessens the likelihood that high status students will dominate group discussion. In this article, we examine the influence of soliciting individual written explanations to multiple-choice concept questions (a private interaction) on performance and reasoning.

A few studies have used students’ short written explanations to examine the reasoning behind their multiple choice answer selections (Tamir, 1989, 1990; Chandrasegaran, Treagust & Mocerino, 2007; Xie & Lee, 2012; Yarroch, 1991). However, the analysis of the explanations is typically directed towards assessment, either assessing the written explanations themselves, or using the explanations in the process of item development for assessment. For example, Tamir (1990) suggests that written explanations can be used to develop two-tier multiple choice questions where students answer consecutive questions with the second question asking them to select the appropriate explanation for their answer choice to the first question. In the development of concept inventories, written explanations are often used to develop distractors and some use two-tier items (Nelson, Geist, Miller, Streveler, & Olds,

2007). In the study reported in this article, rather than assessment, we concentrate on the role of written explanations in student thinking and learning.

Our study explores if individual written explanations promote students' thinking "in a focused way" by having students in the treatment condition write an explanation justifying their answer choice. In particular, we are interested in the "sense making processes" of students. In answering the concept questions explored in this study (e.g. Figure 3.1), we hypothesize that there is a private interaction between what is inside of a student's mind, the answer choices available, and the question statement. That interaction is different and richer when the student must also construct an explanation in words on the computer screen. On a coarse level, we can imagine that when students answer a multiple-choice question, they may select their answer choice based on a guess, intuition, experience, or explicit scientific reasoning. In being asked to provide a written explanation they will "focus" on more explicit use of scientific reasoning, as that type of explanation would typically align with their conception of what the instructor values. On a fine level, we can imagine that the act of writing enables students to more fully develop reasoned arguments. Through elaborating their thoughts in the process of writing, explaining, and constructing a logical argument in defense of an answer, it becomes more likely that student will identify and correct flaws in their logic.

Our hypothesis is supported by learning scientists who have argued for decades that writing enhances thinking (Odell, 1980; Applebee, 1984; Gere, 1985; Sensenbagh, 1989; McDermott & Hand, 2010). Researchers have investigated the

influence of writing on thinking in the context of lengthy essays, journaling exercises, and shorter passages to answer and explain problems (Klein, 1999; Peasley et al., 1992; Prain & Hand, 1996). The scope in our study is similar to VanOrden (1987, 1990) who gave chemistry students short writing assignments that required the students to first solve a chemistry problem and then further explain and discuss their results in writing. She suggests students should “be able to do more than recognize or recall information” and hypothesized that the writing helps students to “(1) integrate concepts, (2) apply the integrated concepts to real life, and (3) communicate those concepts.”

Writing-to-learn (WTL) is an overarching teaching strategy that emphasizes the relation between thinking and writing by prompting students to create “texts that explore the relationships among ideas” (Klein, 1999). WTL generally takes a “constructivist” perspective (Klein, 1999; Rowell, 1997) that emphasizes that knowledge is *actively* constructed. From this perspective, students learn new knowledge by actively building upon and integrating prior knowledge. Some even argue that writing is inherently more about understanding than communication (Howard, 1988), a view consistent with Klein’s (2006) “second generation” cognitive science model. Writing has been hypothesized to help promote thinking by prompting students to take multiple perspectives (Tierney, Soter, O’Flahavan & McGinley, 1989) and encouraging conceptual reprocessing (Scardiamalia & Bereiter, 1986) both of which help build understanding by reinforcing connections between related concepts

(McDermott & Hand, 2010). In other words, the writer cannot communicate in the absence of organizing their thinking.

Our study empirically compares students' in-class responses to concept questions with and without written explanations. We also analyze the arguments in the explanations to infer whether written explanations help students further think about the concept questions. The premise is that as students think about the concept questions more fully, they better learn the concepts and practices of science.

3.4 Experimental Methods

The study has a sequential, mixed methodological basis and is divided into two parts. The first part uses a cross-over design where individual student response distributions are quantitatively analyzed to determine the effect of providing a written explanation. This part addresses research question 1. In the second part, a set of isomorphic questions is qualitatively analyzed through an emergent coding scheme to address research question 2.

3.4.1 Participants & Setting

This study is based on data obtained from two cohorts enrolled in a required, sophomore-level, undergraduate thermodynamics course at a large public university. The cohorts contained chemical, biological, and environmental engineering students. A total of 302 students (164 in Cohort A, 138 in Cohort B) participated in the study. The Institutional Review Board approved the research and participants signed informed consent forms.

The students from each cohort attended a common lecture and self-selected into one of two weekly, one-hour recitation sections. The lectures and recitations for both cohorts were taught by the same instructor. For each cohort the second recitation section was scheduled immediately after the first and held in the same room. The room was consistent across both cohorts as well.

During recitation, concept questions were posed to students and student responses were collected. The questions were timed manually, and once approximately 60% of the section had answered, the instructor typically gave a verbal warning of “30 seconds” and a verbal countdown for the last several seconds. The time spent on each concept question with the treatment condition, the written explanation prompt, is longer than is typical (Mazur 1997), largely because students are asked to provide short written explanations of why they selected a particular answer choice. All students in the course received full credit for submitting an answer, and received extra-credit for each correct answer submitted. This low-stakes approach was intended to both encourage students’ conceptual thinking processes while still providing incentive for students to respond correctly (Willoughby & Gustafson, 2009; James, 2006).

3.4.2 Study Design

In the first part of the study, we quantitatively analyzed the influence of written explanations on student answers to multiple choice questions. To minimize the effect of potential differences between sections, a crossover study was conducted. A design matrix can be seen in Table 3.1. Each section alternated between participation in the

comparison condition (no written explanation) and in the treatment condition (with written explanation) in two- to three-week intervals during the 10-week term. For the first three weeks of recitation, Section 1 was not provided with a prompt and space to explain their answer choices in writing and Section 2 was. In the fourth week, the treatment and control conditions were switched. This process was repeated in weeks seven through ten.

Table 3.1. Experimental design matrix

Week	Section 1	Section 2
1-3	Comparison	Treatment
4-6	Treatment	Comparison
7-8	Comparison	Treatment
9-10	Treatment	Comparison

The student responses to concept questions that were collected provide the core of the data for this study. Concept questions, including the multiple-choice answers and written explanations, were answered using the *AICChE Concept Warehouse* (Koretsky et al., 2014). Students were asked to first answer individually on their laptops, smartphones, or tablets, without consulting their neighbors or the instructor. Figure 3.1 (shown earlier) and Figure 3.2 depict the questions, *Expanding Piston* and *Balloon Rising*, as they were presented in the treatment condition for Cohorts B and A, respectively. For the comparison condition, the same question was asked but without the text input box or prompt that says “Please explain your answer in the box below.” To receive credit, students in the treatment condition were required to select an answer

choice but writing an explanation was optional. In the treatment condition, 89% of the student responses contained some type of written explanation.

A perfectly insulated balloon filled with an ideal gas rises into the sky. As the balloon rises, the external pressure decreases, causing the balloon to expand. What happens to the temperature of the gas inside the balloon?

- increases
- decreases
- remains the same
- need more information

Please explain your answer in the box below.

Submit

Figure 3.2. *Balloon Rising* as it was presented in the treatment condition of Cohort A.

In the two year study, a set of 61 unique questions was delivered to both treatment and comparison conditions. This number does not count questions administered near the end of the class and only answered by students in one of the two recitation sections. The questions were selected or written by the course instructor with the criteria of (i) being high quality questions following guidelines from Taylor and Nolen (2007) and (ii) having appropriate alignment with the week's learning objectives. The responses analyzed in this study are from a subset of the questions asked and were selected as follows. First, only the initial student responses to the concept questions were analyzed; we excluded responses to the second poll used as part of peer instruction. Second, each question was reviewed for clarity in two ways.

The last author evaluated the questions for ambiguous language. Separately, the first author read through written responses to identify questions that were clearly misinterpreted by more than one student. If a question was flagged by either method, it was eliminated. In all cases, the misinterpreted questions matched the questions independently identified as having ambiguous language. Ten questions were removed from the set using this process. Finally, the instructor asked a set of nine questions that directed students to select valid equations for situations in the question statement. Since such questions do not match our classification of a “concept question,” they were eliminated. Of the remaining pool of 42 individual concept questions, 21 were written by the instructor and 21 were peer reviewed questions contributed to the Concept Warehouse by others. A total of 31 questions were posed to each cohort, 20 of which overlapped and were used with both cohorts. Response distributions to the 20 overlapping questions generally aligned with one of the cohorts averaging 58% correct overall with a standard deviation of 24% and the other cohort averaging 54% with a standard deviation of 25%.

In the second part of the study, written responses to the two questions shown in Figures 3.1 and 3.2 were qualitatively analyzed. This pair of concept questions was chosen since they are isomorphic; they are also reasonably representative of the other questions. In isomorphic questions, students need to apply the same core concept, but the questions have different surface features or what Smith et al. (2009) call “different cover stories.” In one case, the representation in the cover story matched how the topic was introduced in the text and presented in class. We term this type *canonical*. For the

other case, the cover story corresponded to a realistic physical situation likely to be part of a students' daily experience; we term it *real-life*. This isomorphic question pair was intentionally chosen since the questions demonstrated different results in the first part of the study; the canonical cover story had a statistically significant positive treatment effect (i.e., the group that provided written explanations was more likely to answer the multiple choice question correctly) and the real-life cover story had a significantly negative effect. The different results were observed even though the students answered questions about the same core concept and were given similar prompts.

3.4.3 Data Analysis

To analyze student responses in the first part of the study, statistical significance was determined using non-parametric statistical tests. The statistical significance of answer distribution between treatment and comparison was determined using a Chi-squared test. Chi-squared is appropriate because it can compare two distributions and is not dependent upon an assumption of normality. Dichotomous comparisons were made using a sign test. The sign test assumes that paired samples come from populations with identical spread and central tendency. If the assumption holds true, the differences between pairs should have the same probability of positive differences as negative. Therefore, the number of positive and negative differences should follow a binomial distribution.

In the second part, coding the free-response, written explanations involved open coding, a process used to infer categories of meaning. We use a technique similar

to that of Newcomer and Steif (2008) in their analysis of written explanations to a concept question in statics. The process involves proposing a code, coding individually, comparing among the coders, modifying the code, and repeating until convergence. The first, second, and third authors, participated in this process. All three coders had domain knowledge and had either taken this course or a similar course at another institution. A hierarchical coding scheme was created for the isomorphic pair of questions that incorporates and ranks the observed concepts and misconceptions. Similar to Tamir (1990), the codes ascend from 1 (poorly reasoned) to 4 (well-reasoned) with a higher code indicating more appropriate reasoning in the explanation. The codes and category descriptions for the isomorphic pair are shown in the first two columns of Table 3.2. Sample explanations are presented in the third column of Table 3.2. They are labeled (EP) for *Expanding Piston* (Figure 3.1) and (BR) for *Balloon Rising* (Figure 3.2).

Table 3.2. Sample student explanations for Expanding Piston (EP) and Balloon Rising (BR) by code level

Code	Level Description	Sample Student Explanation
1	Other incorrect reasoning	(EP) “The temperature increase made the air molecule move faster upward, then push the piston goes up” (BR) “There will be the same amount of energy in the balloon, but it will be spread out over a larger volume. Therefore, the average temperature will decrease.”
2	Ideal gas law only	(EP) “if the pressure decreases and the volume increases then there will be no change in the temperature” (BR) “Ideal gas law. As pressure decreases volume is increased to make up for it so temperature stays the same”
3	No heat means no temperature change	(EP) “Since the device is perfectly insulated, there is no heat transfer between the system and surroundings. This means that the temperature is not changing since there is no heat entering into the system.” (BR) “If the balloon is perfectly insulated then there is no heat transfer in or out of the boundaries. There for (<i>sic</i>), the air might expand but it stays the same temperature (<i>sic</i>)”
4	Conservation of energy, including a term for work	(EP) “There is no Kinetic Energy or Heat transfer so the only source of energy to move the piston is the air's internal energy. A decrease in internal energy correlates to a decrease in temperature.” (BR) “Because $KE+KP+U=W-Q$.”

Once the code categories were developed, two researchers initially coded a subset of the written explanations. Explanations with disagreement between the coders were discussed with the research team until the team reached consensus on the appropriate code for the given explanation. The two researchers then coded all the responses and obtained an inter-rater reliability using the Cohen's κ statistic of 0.87 and 0.89 for *Balloon Rising* and *Expanding Piston*, respectively. Approximately 3% of

the explanations per coded question had a discrepancy of more than one code category.

3.5 Results

3.5.1 Quantitative Analysis of Multiple-Choice Responses

Over eleven thousand multiple-choice answers were collected and analyzed. Figure 3 presents two histograms that summarize variation in the responses. Figure 3.3(a) shows the correct answer percentage for each of the 42 questions averaged over all of the student participants. There is clearly a wide distribution in the question difficulty. Figure 3.3(b) shows the correct answer percentage for the 302 students averaged over the questions that they attempted. Again there is a wide range in performance. A large number of student scores appear across five deciles with some students answering the majority of questions correctly (77 students had scores greater than 80%) and other students struggling (60 students less than 50%).

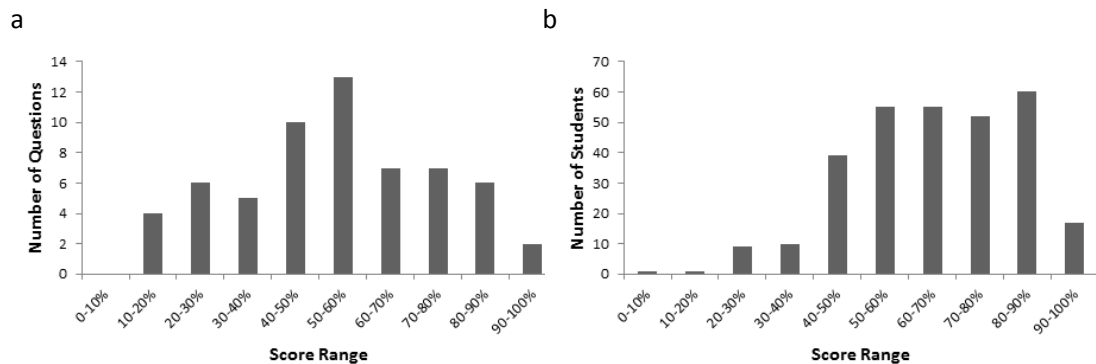


Figure 3.3 Histograms of (a) the number of questions binned by aggregate correct answer percentage across students and (b) number of students binned by aggregate correct answer percentage for each student across questions.

Figure 3.4 displays the percentage of questions where the effect of the treatment condition was (in white) and was not (shaded) found to be significantly different than the comparison ($p < 0.05$). In summary, writing explanations significantly changed the answer distribution a little less than half the time. Furthermore, Figure 3.4 delineates two different cases where such an effect is shown. The case where students in the treatment condition outperformed the students in the comparison is labeled 'Treatment > Comparison'. Questions where the students in the comparison performed better are labeled 'Treatment < Comparison.' The percentage of questions where differences in selection of answer choices between treatment and comparison were not found to be statistically significant is labeled 'Treatment = Comparison'.

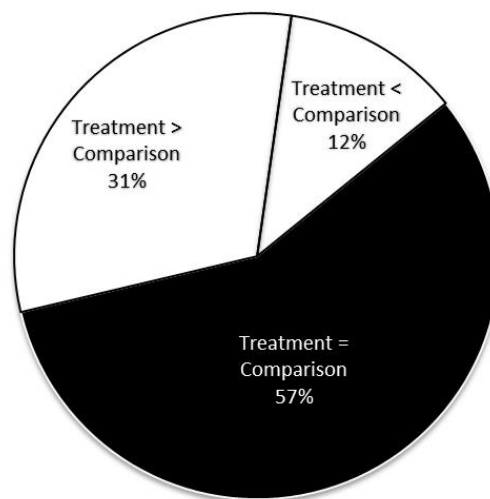


Figure 3.4 Percentage of questions where the treatment condition was (white) and was not (shaded) statistically different from the comparison.

Students in the treatment condition were significantly more likely to answer correctly than the students in the comparison condition for 31% ($n = 13$) of the

questions. However, the treatment condition was more likely to answer incorrectly for 12% ($n = 5$) of the questions. In the latter case, both recitation sections and cohorts participated in the treatment condition for at least one of the five questions, indicating this result did not arise due to a particularly weak section within one of the cohorts. These results suggest that soliciting a written explanation will generally increase the likelihood that students will answer correctly. A sign test confirms a general trend; the proportion of questions with positive treatment effects is higher than would be expected by chance ($p = 0.03$). Moreover, there were three questions with real-life cover stories in this study, and all exhibited a negative treatment effect (out of five questions total with negative effects). In other words, the canonical questions almost always showed a significantly positive influence in number of correct answers with treatment.

3.5.2 Qualitative Analysis of Isomorphic Concept Questions

While we hypothesized that the treatment condition would cause students to perform the same or better than the comparison, there was a portion of questions where students in the treatment condition performed worse (see Figure 3.4). To further investigate these cases, we selected an isomorphic concept question pair for qualitative analysis (see Figures 3.1 and 3.2). Both questions ask students to predict the change in temperature for an adiabatic, expansion process in a closed system by applying the principle of the conservation of energy. However, the question with a canonical cover story resulted in ‘Treatment > Comparison’ while the question with a real-life cover story resulted in ‘Treatment < Comparison.’

The *Expanding Piston* isomorphic concept question depicted in Figure 3.1 illustrates a case where the treatment condition outperformed the comparison. The correct answer choice is that the temperature decreases. A well-reasoned explanation might be as follows: the gas in the system does work on the surroundings (the mass on top of the piston and any gas outside) in order to expand. Since the system is perfectly insulated, the energy to do the work must come entirely from the internal energy of the air. Because the air can be assumed to be ideal, as its internal energy decreases, so must its temperature. (In this introductory thermodynamics course, consideration of intermolecular interactions between gas molecules is not expected.) The *Balloon Rising* isomorphic concept question depicted in Figure 3.2 illustrates a case where students in the treatment condition performed worse than the comparison. The correct answer choice again is that the temperature decreases and a well-reasoned explanation would be similar to that posed for *Expanding Piston*. These questions are isomorphic since they both test the same concept but have different cover stories. The cover story of *Expanding Piston* is canonical and of *Balloon Rising* is real-life.

Figure 3.5 shows the percentage correct for both treatment and comparison conditions for the isomorphic questions. Overall approximately half of the student participants chose the correct answer for *Expanding Piston* while only about one quarter of the students chose correctly for *Balloon Rising*. Furthermore for *Expanding Piston*, significantly more student participants chose the correct answer in the treatment condition than the comparison ($p < 0.001$). For *Balloon Rising*, significantly fewer students in the treatment condition chose the correct answer ($p = 0.044$).

Additionally, we can compare frequency of correct responses to one question in the isomorphic pair to the other. For the comparison condition, there is not a significant difference between the two questions ($p = 0.234$). However, there is a significant difference between the two questions in the treatment condition ($p < 0.001$). In this case, the process of explaining the answer choice in writing magnifies the influence of the cover story.

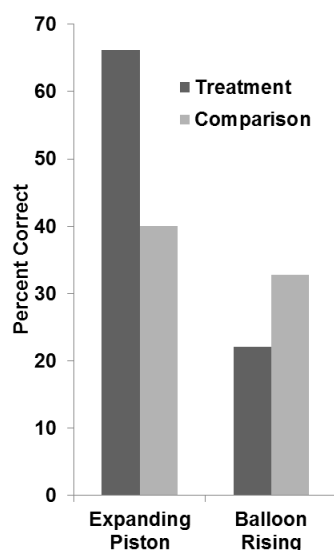


Figure 3.5 Percentage of correct answers for the treatment condition and comparison for the *Expanding Piston* and *Balloon Rising* concept questions.

Figure 3.6(a) presents the frequency of the coded explanations given by the treatment condition of Cohort B when they answered *Expanding Piston*. Code values ascend in sophistication from 1 to 4, and it is expected that the likelihood of choosing a correct answer should increase as well. Refer back to Table 3.2 for code categories and sample student explanations for each code value. As expected, most students who gave explanations and chose the correct answer did use reasoning that considered

conservation of energy and included work, e.g., “...the only source of energy to move the piston is the air's internal energy.” However, a set of students serendipitously chose the correct answer through faulty reasoning that focused on the ideal gas law, e.g., “because the lower the pressure the air is in the lower the temperature.” Some of the students that chose an incorrect answer focused on the adiabatic nature of the question; they considered energy but neglected work. For example, the student quoted in Table 3.2 said, “Since the device is perfectly insulated, there is no heat transfer between the system and surroundings. This means that the temperature is not changing since there is no heat entering into the system.”

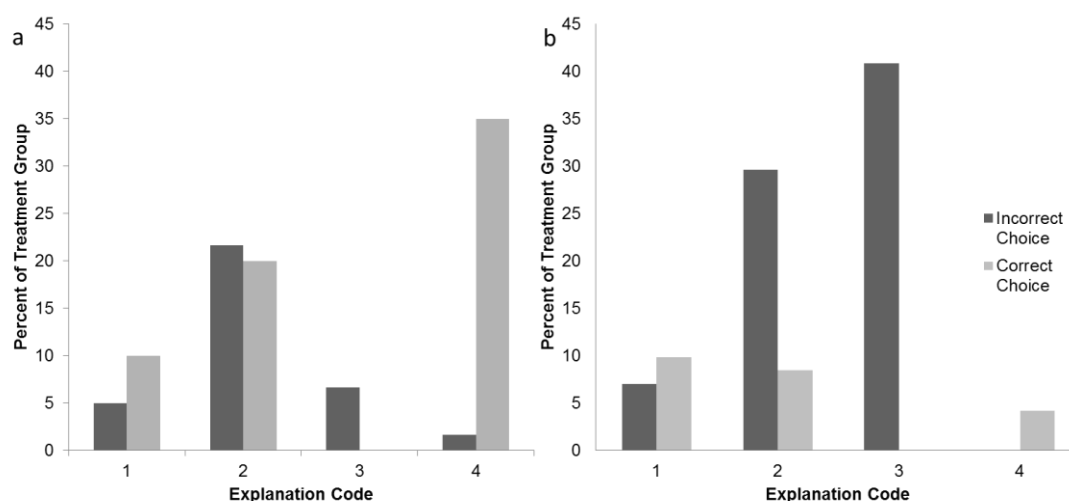


Figure 3.6. Code distribution for the explanations written by the treatment condition for (a) *Expanding Piston* and (b) *Balloon Rising*.

Figure 3.6(b) presents the percentage of the coded explanations given by the treatment condition of Cohort A when they answered *Balloon Rising*. Code values for this isomorphic question are identical to *Expanding Piston* and detailed in Table 3.2. In this case, only one student who answered correctly in the treatment condition

justified his/her choice by invoking conservation of energy. Most students who chose the correct answer used faulty reasoning related to the ideal gas law or to the “other incorrect reasoning” categories. For example, one student chose correctly and wrote, “Temperature must go down to maintain $PV=nRT$ relationship.” In other words, the student focused entirely on the ideal gas law to argue that temperature decreases since the final pressure (state 2) is less than the initial pressure (state 1), failing to recognize that the volume also changes, and neglecting to consider conservation of energy altogether. Cases in which students reason by focusing on two changing variables (e.g., pressure and temperature), but ignore a third variable that is also changing (e.g., volume) demonstrate an inability to identify multivariate processes, and is a common type of misconception in science (Loverude, Kautz, & Heron, 2002). On the other hand, most students who chose an incorrect answer directed explanations to the adiabatic nature of the question, attempting to account for energy, but failing to consider work (see Table 3.2, code 3).

There is a notable exception regarding the students that relied solely on the ideal gas law in their reasoning for *Balloon Rising*. Students in the treatment condition chose ‘need more information,’ an answer option not available for *Expanding Piston*, more frequently than the comparison condition ($p = 0.031$). Most explained that they saw too many degrees of freedom for the ideal gas law to decisively give an answer but failed to consider conservation of energy. One student correctly identified the multivariate process as follows: “The pressure decreased, which caused the volume to increase, but we don't know if that caused the PV value to be the same or not.”

3.6 Discussion

Overall, the results from this study show that soliciting a written explanation to concept questions in introductory thermodynamics can have a significant influence on answer choice and when it does, that influence tends to be positive, i.e., students more often answer correctly (research question 1). The distribution of multiple-choice answers to approximately half of the concept questions in this study is significantly different when students were prompted to write an explanation than when they were not (see Figure 3.4). The ‘focused’ nature of private interactions provides a frame to consider how students are encouraged to use more explicit scientific reasoning and to explain the positive influence on reasoning. These results are consistent with studies in cognitive neuroscience that show a prompt for a confidence judgment decreases the likelihood of guessing (Voss & Paller, 2010) and with the writing-to-learn literature that suggests that writing encourages the restructuring and reorganization of knowledge (Klein, 1999; Rivard, 1994). These results would also be anticipated by Chi, De Leeuw, Chiu & LaVancher (1994) who indicate that the active process of explaining encourages students to integrate new knowledge with existing knowledge and leads to richer conceptual understanding. By extension, a richer conceptual understanding should lead to more frequent correct answer choice selections.

However, we found that for certain questions, soliciting a written explanation actually increased the likelihood that students would answer incorrectly. To further explore this issue, we performed a more detailed investigation of contrasting responses to an isomorphic question pair. Our results highlight important subtleties in question

context of which instructors and question developers should be aware (research question 2). The two isomorphic concept questions had opposite treatment effects despite being delivered under practically identical conditions. Experts would agree that both the *Balloon Rising* question and the *Expanding Piston* question are intended to test understanding of the first law of thermodynamics in the same way. However, the effect of students' written explanations was different. The *Balloon Rising* question represents the unexpected case where students in the treatment condition chose the correct answer less frequently than students in the comparison condition. The coding of the written explanations for this question also shows that reasoning associated with better developed explanations (codes 3 and 4) was almost always associated with an incorrect answer choice while explanations indicative of less developed reasoning (codes 1 and 2) were sometimes associated with correct answer choices (see Figure 3.6a).

We argue that in both isomorphic questions the focused nature of the private interactions that take place while writing an explanation helps students think about the question more scientifically. For the *Expanding Piston* question, this effect manifests in a greater proportion of correct answers, the majority of which are well reasoned. For the *Balloon Rising* question, we posit that more students in the treatment condition utilized superior scientific reasoning as well even though they may not have arrived at the correct answer. In answering this question, students who selected the correct answer almost always did so for the wrong reason (at least in the treatment condition). By elaborating and justifying their choice through writing, students were prompted to

identify flaws in their reasoning, and change their initial answer choice. However, if their reasoning was not fully developed (e.g., code 3 instead of 4), it could lead them away from the correct answer to choose an incorrect answer. For example, significantly more students in the treatment condition selected the ‘need more information’ answer choice ($p = 0.031$). This answer selection was accompanied by explanations that indicate the students recognized the problem could not be solved using the ideal gas law by normative scientific reasoning. Specifically, they identified that they needed to consider three simultaneously changing variables (pressure, volume, and temperature). An inability to account for multi-variable relationships using the ideal gas law has been reported in physics classes (Loverude, Kautz, & Heron, 2002; Rozier & Viennot, 1991). Thus, in identifying the multivariate nature of the problem, the students who chose ‘need more information’ demonstrated better scientific reasoning even if they failed to consider the conservation of energy. Presumably, if they had not been prompted to write an explanation, some would have applied the faulty bivariate ideal gas reasoning (i.e., pressure decreases so temperature decreases) and might have selected the correct answer, as we believe was the case in the comparison condition.

Many more students were able to identify and correctly apply the appropriate scientific principle in the *Expanding Piston* question than in the isomorphic *Balloon Rising* question. Such differences illustrate how a question’s surface features can affect students’ activation of prior knowledge in science and are useful for instructors to understand as they use these concept questions for active learning in the classroom.

Drawing from the literature in conceptual change (Hammer, Elby, Scherr & Redish, 2005) and transfer (Catrambone & Holyoak 1989; Schwartz, Bransford & Sears, 2005; Schwartz, Chase & Bransford, 2012), we attribute two features as sources for differing student responses. First, the piston-cylinder assembly provides a *canonical* representation of a closed system undergoing a process where the volume changes. In this thermodynamics class, both the textbook and lecture use this representation to introduce and illustrate conservation of energy. When the concept question is framed in the canonical context, students are directly cued to apply what they might consider as a school-related concept, the conservation of energy. The additional visual cue of the simplified piston-cylinder in the question statement reinforces this contextual alignment with school. On the other hand, while most students have probably had experience with balloons in everyday life, the *Balloon Rising* concept question asks them to apply the conservation of energy principle to this more real-life context (VanOrden, 1990), which is clearly more challenging. Second, *Balloon Rising* explicitly contained the phrase “ideal gas” while *Expanding Piston* does not. The explicit term “ideal gas” may divert students from considering energy by providing an alternative and familiar school cue (ideal gas), especially as they are seeking to connect the real-life context of this question to a school related task.

In this article, we have approached the development, use, and analysis of concept questions like the real-life *Balloon Rising* question from a perspective of developing students’ thinking and reasoning processes rather than the more common view of using these questions for assessment of what students know. This intentional

choice of perspective has implications for how concept questions are selected. However, we acknowledge that this view is a simplification in that thinking, reasoning, and assessment are inextricably linked in the learning process. For assessment, question use is predicated by content validity and item validity (Yarroch, 1991). Since experts would agree that the both *Expanding Piston* and *Balloon Rising* appropriately test an underlying scientific principle (conservation of energy), they have ‘content validity’ (i.e., they test appropriate content). Item validity is fulfilled if students choose an answer that corresponds to reasoning appropriate for that answer selection (i.e., the correct multiple-choice answer is chosen for the correct reason). In the isomorphic question pair studied here, the *Expanding Piston* would be considered item valid while *Balloon Rising* would not since in the latter question students usually choose the correct answer using faulty reasoning. Therefore, from an assessment perspective, *Balloon Rising* would be rejected. In fact, of the 42 questions studied, the three concept questions that clearly present a real-life context would all be rejected by this criterion.

However, from the perspective of thinking and reasoning, rejection of these questions is not desirable. With rejection, student learning would be sequestered to the canonical representations common in the classroom and students would face greater challenges operationalizing the concepts in engineering practice. Rather, we assert that instructional design must intentionally provide students with varied contexts to apply the concepts that they are learning and help them see that the same principles that apply to “school” contexts also apply in the same way to the “real-life” contexts with

which they have had direct experience. In other words, rather than question rejection, it is more desirable to view a progression where the instructor helps students connect the canonical question (e.g., *Expanding Piston*) to the real-life one (e.g., *Balloon Rising*) and helps students develop a more expansive way to consider and apply scientific principles (Engle, 2006). Such connections are critical for developing the deep and adaptive understanding needed for professional practice (Parker et al., 2013; McKenna, 2007). Facilitating these connections may be aided by pedagogy. For example, instead of selecting either *Expanding Piston* or *Balloon Rising*, the instructor could provide both questions and ask students to reflect on their similarities and differences. In order to achieve this goal, more real-life concept questions need to be developed. The fact that only 3 of the 42 questions studied here were clearly presented in a real-life context reflects the difficulty in creating this type of question and the lack of general availability of such questions.

For questions like *Balloon Rising*, written explanations provide instructional benefit in addition to helping students focus on scientific reasoning. They allow the instructor to examine explicit student reasoning and potentially identify questions where students apply inappropriate reasoning to get the correct answer. With such understanding, the instructor can directly address the faulty reasoning and help students make connections to normative scientific principles. The recent development of methods such as lexical analysis (Haudek, et al., 2012; Urban-Lurain, et al., 2013) and tools such as the *AICHE Concept Warehouse* (Koretsky et al., 2014) will help

facilitate the ability for instructors to identify student reasoning processes without requiring overwhelming effort.

3.7 Conclusions

This study examined changes in student answer selection that result from soliciting reflective written explanations to multiple-choice concept questions. Students tended to pick the correct answer more frequently when provided with a prompt for a written explanation. However, there is a set of questions where the students that are prompted for an explanation are less likely to choose the correct answer. Qualitative analyses suggest that students may indeed also be activating more sophisticated reasoning processes to these questions, even though they tend to select the wrong answer. The results from this study suggest that instructors should be aware of the possibility of questions lacking item validity. Instructors using multiple-choice concept questions are encouraged to solicit written explanations and look for cases where their students get the correct answer using incorrect reasoning, perhaps by simply sampling the explanations from the students that selected the correct answer. Finally, instructors should consider question context in developing students' ability to apply conceptual understanding. Specifically, they should seek ways to help their students recognize that core concepts apply to a broad range of contexts, extending beyond the canonical representations presented in class to those closer to their real-life experiences.

3.8 Acknowledgements

The authors would like to acknowledge funding from the National Science Foundation (DUE 1023099) and from Oregon State University (TRF 931 and TRF 1056). The authors gratefully acknowledge helpful discussions with Shane Brown, Susan Nolen, and Debra Gilbuena. The opinions expressed are strictly those of the authors and do not necessarily represent those of the National Science Foundation.

3.9 References

- Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, 54(4), 577-596.
- Borrego, M., Cutler, S., Prince, M., Henderson, C., & Froyd, J. E. (2013). Fidelity of Implementation of Research-Based Instructional Strategies (RBIS) in Engineering Science Courses. *Journal of Engineering Education*, 102(3), 394-425.
- Brooks, B. J., & Koretsky, M. D. (2011). The influence of group discussion on students' responses and confidence during peer instruction. *Journal of Chemical Education*, 88(11), 1477-1484.
- Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. *CBE-Life Sciences Education*, 6(1), 9-20.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1147.
- Chandrasegaran, A., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293-307.
- Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3), 439-477.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science*, 5(2), 121-152.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9), 970-977.
- Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics*, 1(1), 40-95.
- Duncan, D. (2006). Clickers: A new teaching aid with exceptional promise. *Astronomy Education Review*, 5(1), 70.

- Gere, A. R., & Abbott, R. D. (1985). Talking about writing: The language of writing groups. *Research in the Teaching of English*, 19(4), 362-385.
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89-120): Information Age Publishing.
- Haláková, Z., & Prokša, M. (2007). Two kinds of conceptual problems in chemistry teaching. *Journal of chemical education*, 84(1), 172-174.
- Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What Are They Thinking? Automated Analysis of Student Writing about Acid-Base Chemistry in Introductory Biology. *CBE-Life Sciences Education*, 11(3), 283-293.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3), 141-158.
- Howard, V., & Barton, J. H. (1988). *Thinking on paper*. New York, NY: William Morrow Paperbacks.
- Imbrie, P. K., & Reed-Rhoads, T. (2011). *Expanding Access to Assessing Conceptual understanding: ciHUB, Virtual Community Supporting Conceptual Learning*. Paper presented at the ASEE 9th Global Colloquium on Engineering Education, Singapore.
- James, M. C. (2006). The effect of grading incentive on student discourse in peer instruction. *American journal of physics*, 74(8), 689-691.
- Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education*, 53(3), 819-827.
- Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11(3), 203-270.
- Klein, P. D. (2006). The challenges of scientific literacy: From the viewpoint of second-generation cognitive science. *International Journal of Science Education*, 28(2-3), 143-178.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4(4), 298-310.
- Koretsky, M., Falconer, J., Brooks, B., Gilbuena, D., Silverstein, D., Smith, C., Miller, R., Miletic, M. (2014). The AIChE Concept Warehouse: A Tool to Promote Conceptual Learning. *Advances in Engineering Education*.
- Koretsky, M., Miller, R., Falconer, J., Prince, M., Vigeant, M., Krause, S., & Silverstein, D. (2011). *SPECIAL SESSION: Educational Methods and Tools to Encourage Conceptual Learning*. Paper presented at the American Society for Engineering Education Annual Conference & Exposition, Vancouver, B.C., Canada.
- Koretsky, M. D., & Brooks, B. J. (2011). Comparison of Student Responses to Easy and Difficult Thermodynamics Conceptual Questions during Peer Instruction. *International Journal of Engineering Education*, 27(4), 897-908.
- Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. J. (2004). *Development, testing, and application of a chemistry concept inventory*. Paper presented at

- the Frontiers in Education, Savannah, Georgia.
- Lasry, N., Charles, E., Whittaker, C., & Lautman, M. (2009). *When Talking is Better than Staying Quiet*. Paper presented at the AIP Conference Proceedings, Ann Arbor, MI.
- Linn, M. C., & Chiu, J. (2011). Combining learning and assessment to improve science education. *Research and Practice in Assessment*, 6(Winter 2011), 5-14.
- Loverude, M. E., Kautz, C. H., & Heron, P. R. (2002). Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas. *American Journal of Physics*, 70(2), 137-148.
- Lucas, A. (2009). Using peer instruction and i-clickers to enhance student participation in calculus. *Primus*, 19(3), 219-231.
- Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ 07458: Prentice Hall.
- McConnell, D. A., Steer, D. N., Owens, K. D., Knott, J. R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H. (2006). Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, 54(1), 61.
- McDermott, L. C. (2001). Oersted Medal Lecture 2001: "Physics Education Research—The Key to Student Learning". *American Journal of Physics*, 69(11), 1127-1137.
- McDermott, M. A., & Hand, B. (2010). A secondary reanalysis of student perceptions of non-traditional writing tasks over a ten year period. *Journal of Research in Science Teaching*, 47(5), 518-539.
- McKenna, A. F. (2007). An investigation of adaptive expertise and transfer of design process knowledge. *Journal of Mechanical Design*, 129(7), 730-734.
- Midkiff, K. C., Litzinger, T. A., & Evans, D. (2001). *Development of engineering thermodynamics concept inventory instruments*. Paper presented at the Frontiers in Education Conference, 2001. 31st Annual.
- Mulford, D. R., & Robinson, W. R. (2002). An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *Journal of Chemistry Education*, 79(6), 739 - 744.
- Nelson, M. A., Geist, M. R., Miller, R. L., Streveler, R. A., & Olds, B. M. (2007). *How to create a concept inventory: The thermal and transport concept inventory*. Paper presented at the Annual Conference of the American Educational Research Association, Chicago, IL.
- Newcomer, J. L., & Steif, P. S. (2008). Student thinking about static equilibrium: Insights from written explanations to a concept question. *Journal of Engineering Education*, 97(4), 481-490.
- Nurrenbern, S. C., & Robinson, W. R. (1998). Conceptual questions and challenge problems. *Journal of Chemical Education*, 75(11), 1502-1503.
- Odell, L. (1980). The process of writing and the process of learning. *College Composition and Communication*, 31(1), 42-50.
- Papaphotis, G., & Tsaparlis, G. (2008). Conceptual versus algorithmic learning in high school chemistry: the case of basic quantum chemical concepts. Part 1.

- Statistical analysis of a quantitative study. *Chemistry Education Research and Practice*, 9(4), 323-331.
- Parker, W. C., Lo, J., Yeo, A. J., Valencia, S. W., Nguyen, D., Abbott, R. D., . . . Vye, N. J. (2013). Beyond Breadth-Speed-Test Toward Deeper Knowing and Engagement in an Advanced Placement Course. *American Educational Research Journal*, 50(6), 1424-1459.
- Peasley, K. L., Rosaen, C. L., & Roth, K. J. (1992). *Writing-to-learn in a conceptual change science unit*. East Lansing, MI: Center for the Learning and Teaching of Elementary Subjects, Institute for Research on Teaching, Michigan State University.
- Perez, K. E., Strauss, E. A., Downey, N., Galbraith, A., Jeanne, R., & Cooper, S. (2010). Does displaying the class results affect student discussion during peer instruction? *CBE-Life Sciences Education*, 9(2), 133-140.
- Prain, V., & Hand, B. (1996). Writing for learning in secondary science: Rethinking practices. *Teaching and Teacher Education*, 12(6), 609-626.
- Rivard, L. O. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching*, 31(9), 969-983.
- Rowell, P. M. (1997). Learning in school science: The promises and practices of writing. *Studies in Science Education*, 30, 19-56.
- Rozier, S., & Viennot, L. (1991). Students' reasonings in thermodynamics. *International Journal of Science Education*, 13(2), 159-170.
- Scardamalia, M., & Bereiter, C. (1986). *Research on written composition. Handbook of research on teaching*, 3, 778-803. (M. Wittrock Ed. Vol. 3). New York, NY: Macmillan.
- Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). *Efficiency and innovation in transfer. Transfer of learning from a modern multidisciplinary perspective*, 1-51. Charlotte, NC: Information Age Publishing, Inc.
- Schwartz, D. L., Chase, C. C., & Bransford, J. D. (2012). Resisting overzealous transfer: Coordinating previously successful routines with needs for new learning. *Educational Psychologist*, 47(3), 204-214.
- Sensenbaugh, R. (1989). ERIC/RCS: Writing Across the Curriculum: Evolving Reform. *Journal of Reading*, 32(5), 462-465.
- Singh, C. (2005). Impact of peer interaction on conceptual test performance. *American journal of physics*, 73(5), 446-451.
- Smith, M., Wood, W., Krauter, K., & Knight, J. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE-Life Sciences Education*, 10(1), 55-63.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122-124.
- Tamir, P. (1989). Training teachers to teach effectively in the laboratory. *Science education*, 73(1), 59-69.
- Tamir, P. (1990). Justifying the selection of answers in multiple choice items.

- International Journal of Science Education*, 12(5), 563-573.
- Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. *Cell Biology Education*, 4(2), 112-117.
- Taylor, C., & Nolen, S. (2007). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson-Merrill-Prentice Hall.
- Tierney, R. J., Soter, A., O'Flahavan, J. F., & McGinley, W. (1989). The effects of reading and writing upon thinking critically. *Reading Research Quarterly*, 24(2), 134-173.
- Urban-Lurain, M., Prevost, L., Haudek, K. C., Henry, E. N., Berry, M., & Merrill, J. E. (2013). *Using Computerized Lexical Analysis of Student Writing to Support Just-in-Time Teaching in Large Enrollment STEM Courses*.
- Vahey, P., Tatar, D., & Roschelle, J. (2007). *Using handheld technology to move between private and public interactions in the classroom. Ubiquitous computing in education: Invisible technology, visible impact*, 187 - 210. (M. Hooft & K. Swan Eds.). Mahwah, NJ: Erlbaum.
- Van Dijk, L., Van Der Berg, G., & Van Keulen, H. (2001). Interactive lectures in engineering education. *European Journal of Engineering Education*, 26(1), 15-28.
- VanLehn, K., & Van de Sande, B. (2009). *Expertise in elementary physics, and how to acquire it. The development of professional performance: Toward measurement of expert performance and design of optimal learning environments*, 356-378. (K. Ericsson Ed.). New York, NY: Cambridge University Press.
- VanOrden, N. (1987). Critical-thinking writing assignments in general chemistry. *Journal of Chemical Education*, 64(6), 506-507.
- VanOrden, N. (1990). Is writing an effective way to learn chemical concepts? Classroom-based research. *Journal of Chemical Education*, 67(7), 583-585.
- Vigeant, M., Prince, M., & Nottis, K. (2011). *Engineering undergraduates' conceptual understanding of thermodynamics: Assessment and change after normal instruction*. Paper presented at the Proceedings from Hawaii International Conference on Education, Honolulu, HI.
- Voss, J. L., & Paller, K. A. (2010). What makes recognition without awareness appear to be elusive? Strategic factors that influence the accuracy of guesses. *Learning & Memory*, 17(9), 460-468.
- Willoughby, S. D., & Gustafson, E. (2009). Technology talks: Clickers and grading incentive in the large lecture hall. *American Journal of Physics*, 77(2), 180-183.
- Xie, C., & Lee, H.-S. (2012). A visual approach to nanotechnology education. *International Journal of Engineering Education*, 28(5), 1006-1018.
- Yarroch, W. L. (1991). The implications of content versus item validity on science tests. *Journal of Research in Science Teaching*, 28(7), 619-629.

4. Using Word Clouds to Analyze Students' Short Written Responses

Bill J. Brooks,* Debra M. Gilbuena,* Stephen J. Krause,** Milo D. Koretsky*

*School of Chemical, Biological, and Environmental Engineering

Oregon State University

Corvallis, OR, USA 97331-2702

(541)737-4591

** Ira A. Fulton School of Engineering

Arizona State University, Tempe Campus

Tempe, AZ 85287

Washington, DC 20057

(480) 965-2050

To be Submitted to Chemical Engineering Education

<http://www.che.ufl.edu/cee/>

4.1 Abstract

Active learning in class helps students develop deeper understanding of chemical engineering principles. While the use of multiple-choice ConcepTests is clearly effective, we advocate for including student writing in learning activities as well. In this article, we demonstrate that word clouds can provide a quick analytical technique to assess student writing. We provide two examples: students' in-class explanations of their answer choice to multiple choice ConcepTests, and student responses to "muddiest point" reflection exercises.

4.2 Introduction

Imagine your department chair has just assigned you to teach Material and Energy Balances, a required course that has grown considerably in enrollment for the last several years. You taught it a few years ago and used ConcepTests (Mazur, 1997) for in-class active learning with reasonable success. You plan to use them again this time around. You recently attended a professional development seminar which described the learning benefits of asking your students to write explanations and reflections. It sounds like a great idea, so you decide to have your students write explanations to justify their answer choices to their ConcepTests. You try it the first week of class. After class you are checking out the responses, plotting the answer distributions, and then it hits you. You see the 250 written explanations. It is going to take hours to read and analyze all of these explanations! If you don't read them, will your students take them seriously? Will they continue to reflect and get the most out of them? If you do take the time to read all of them, what are you sacrificing? What part of your class preparation are you giving up? What do you do?

Many instructors have approached a dilemma similar to the one discussed in the vignette above. Sometimes it happens when you are first contemplating the implementation of a research-based instructional strategy and sometimes it comes, as

in the vignette, only after the first implementation. This article presents a potential solution to the vignette dilemma of analyzing short written responses - the use of word clouds. Word clouds provide a visual representation of word usage and frequency. They offer a quick visualization of aggregate text responses to reduce the burden of information overload (Godwin-Jones, 2006). When combined with an audience response system, they afford instructors a way to easily analyze written explanations from tens or hundreds of students in a very short time.

In this article, we describe how word clouds can be used for formative assessment in active learning. In particular, we discuss how they have been integrated into and used with the AIChE Concept Warehouse (CW). The CW is a web-based tool to help the chemical engineering education community more easily use active learning pedagogies (Koretsky, et al., 2014). We focus on the ways word clouds afford improved instruction through the CW through their use as a formative assessment tool that can provide instructors and students with valuable, timely feedback. We illustrate the use of word clouds with evidence from two active learning examples: student in-class responses to multiple choice concept questions, during the first part of peer instruction (Mazur, 1997); and student responses to “muddiest point” reflection exercises (Mosteller, 1988; Hall, Waitz, Brodeur, Soderholm & Nasr, 2002), intended to assess the most confusing topic or concept presented in lecture. In addition, we explore other potential opportunities.

4.3 Background

Active learning pedagogies have been shown to improve student conceptual understanding (Hake, 1998). Active learning means more than engaging students in classroom exercises; activities should be designed around learning outcomes, promote student reflection, and get students to think about what they are learning (Prince, 2004). Formative assessment is one integral aspect of these pedagogies that helps meet these design criteria. Assessments that include students' short written explanations (Brooks et al., in review) or reflections (Felder, 1995; Mosteller, 1988; Kelly, et al.; Carberry, Krause, Ankeny and Waters, 2013 ; Hall et al., 2002) can enhance learning (Klein, 1999; VanOrden 1987,1990; Rivard, 1994; Brooks, et al., in review). However, it is difficult to expediently examine written responses in large classes.

Word clouds, also known as "tag clouds" or "term clouds" can be a useful analytical tool to summarize text data and provide meaningful interpretations (Clough & Sen, 2008). Word clouds have been found to be beneficial because they are "highly interpretable," giving a direct visual representation of the content being measured (Chuang, Ramage, Manning & Heer, 2012). They have been used both as a research tool and as a teaching tool.

As a research tool, McNaught & Lam (2010) showed how word clouds can uncover themes in interviews consistent with those identified by other qualitative analysis methods. Similarly, word clouds have been used as a qualitative analysis tool in other cases (Williams, Parkes & Davies, 2013; Cidell, 2010). While word clouds are generally interpreted in terms of the most common words, attention to missing words

or infrequent words can be just as important (Luhn, 1958; Chuang, et al. 2012). The context from which a word cloud is created also plays an important role in the interpretation of the resultant word clouds (Chuang et al., 2012), e.g., the phrase “energy balance” holds one meaning in a chemical engineering course and takes on an entirely different meaning in oriental medicine (Qiu, 2007).

Educators have also begun to report the benefits of word clouds in teaching. Ramsden & Bate (2008) put forth a general working paper presenting word clouds as a useful teaching tool comparing different word cloud software and discussing aspects educators should consider. They note that data needs to be in a useable state (i.e., as electronic text) for word cloud analysis. In addition, they note the following potential limitations of word cloud software: spelling errors may not be taken into account; words that appear to be common may be eliminated even if they represent an important acronym, e.g., it versus IT; word clouds represent frequency, not necessarily importance; and word clouds often fail to group similar words. Ramsden and Bate (2008) suggest the use of word clouds as a complementary method to other research and teaching methods.

In practice, educators have described students construction of word clouds to promote reflection and discussion in many fields, including: accounting (Miley & Read, 2011), social studies (Berson & Berson, 2009), teaching vocabulary (Dalton & Grisham, 2011), and theology (Hamm, 2011). Word clouds have also been used in several ways for teaching reading and writing (Hayes, 2008). One general example promotes the use of texting combined with word clouds in a high school English

setting (teachingchannel.org, 2013). In this article, we present how word clouds can be used as a learning tool in chemical engineering. However, instead of having students manually construct them, word clouds are automatically constructed from students' writing for the instructor and students to use. The construction of word clouds is facilitated through the AIChE Concept Warehouse (CW).

4.4 AIChE Concept Warehouse

The AIChE CW was used as the primary data collection tool for the two examples of word cloud use reported in this article. It is a database driven website facilitating the use of concept questions with content from throughout the core chemical engineering curriculum. Currently the AIChE Concept Warehouse (CW) has more than 2,000 concept questions (ConcepTests) and 10 valid and reliable concept inventories available for searching, viewing, and using in courses. Instructor and student interfaces are available for use at <http://cw.edudiv.org>, and university faculty can obtain an account through this site. More general information about this tool can be found elsewhere (Koretsky et al., AEE, 2014). In this article, we focus on the word cloud feature that facilitated formative assessment.

For context, we provide a detailed description of the algorithm used to generate word clouds in the CW. A wide variety of word cloud algorithms are reported in the literature, e.g., some count the frequency of individual words while others count frequency of word pairs (Viegas & Wattenberg, 2008). Currently the CW summarizes the frequency of single words only. To generate the word clouds, the CW first aggregates all of the written explanations into a single string of text per answer choice.

It removes HTML tags like 'br' and 'quot' as well as filler words (e.g., able, about, above, according, accordingly, across, actually,...). The words as they were submitted temporally are mapped into the cloud horizontally and vertically according to English convention. Word frequency is mapped to the word size and color. Blue, bigger words are more frequent. Red, smaller words are less frequent. While some criticize word clouds for ignoring semantic relationships such as similar words (Ramsden & Bate, 2008; Chuang et al., 2012), the algorithm the CW uses has been improved to ignore case differences and combines similar words like singular and plural forms of words.

4.5 Word Clouds in ConcepTest Assessment

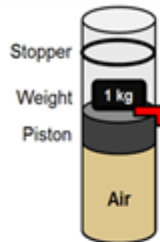
The first active learning example we use illustrates the use of word clouds in the context of polling during peer instruction. Peer instruction is arguably the most well-known and widely used technology-mediated active learning pedagogy in post-secondary Science, Technology, Engineering and Mathematics education (Mazur, 1997; Crouch et al., 2007). It consists of a structured polling process where a concept question (also called a ConcepTest or 'clicker' question) is presented to the class. Students first answer the question individually. They are then encouraged to discuss the answer choices in small groups. Finally, they individually submit a final answer. This sequence is then typically followed by a class wide discussion. The data presented below comes from the first individual answering where students are asked to explain their choice for two sample concept questions.

4.5.1 Methods

The following data was collected from two cohorts enrolled in a required, sophomore-level, undergraduate energy balances course at a large public university. The cohorts contained chemical, biological, and environmental engineering students. Between 60 and 70 students provided written explanations for each of the two questions presented in this article. These students came from subset of a larger study population, reported elsewhere (Brooks et al., in review). The lectures and recitations for both cohorts were taught by the same instructor, in the same room, using the CW to deliver the ConcepTests. The Institutional Review Board approved the research and participants signed informed consent forms.

Figure 4.1 and Figure 4.2 depict two isomorphic concept questions as they were presented to the students in their respective cohort; one cohort answered one question, the other cohort answered the other question. In isomorphic questions, students need to apply the same core concept, but the questions have different surface features, like the balloon or piston, or what Smith et al. (2009) calls “different cover stories.” To answer these questions correctly, ideally, students apply their knowledge of energy balances, recognizing that the work done by the gas on the surroundings lowers its internal energy and, therefore, its temperature. For this question pair, however, the correct answer can also be obtained from faulty reasoning using the ideal gas law. In that case a student may reason, since $PV=nRT$, as P decreases, T also must decrease. This reasoning process fails to account for changing volume, and is, therefore, classified as faulty reasoning.

Air at high pressure and ambient temperature is contained in a perfectly insulated piston-cylinder device. If the locks holding the piston in place are removed, the piston moves upwards to a stopper. The temperature of the air _____.



- increases
- remains the same
- decreases

Please explain your answer in the box below.

Please rate how confident you are with your answer.

- substantially moderately neutral moderately substantially
unsure unsure confident confident
-

Submit

Figure 4.1. Sample question (piston question) as it was presented to students who wrote explanations.

A perfectly insulated balloon filled with an ideal gas rises into the sky. As the balloon rises, the external pressure decreases, causing the balloon to expand. What happens to the temperature of the gas inside the balloon?

- increases
- decreases
- remains the same
- need more information

Please explain your answer in the box below.

Submit

Figure 4.2. Sample question (balloon question) as it was presented to students who wrote explanations.

4.5.2 Results

Tables 4.1 and 4.2 present the word clouds for the explanations by answer option for the questions depicted in Figures 4.1 and 4.2, respectively. The first column in each table shows the answer options for the respective question. The second column contains the word cloud that was produced from the written explanations for the corresponding answer option. The third column contains a representative explanation given by a student that selected the corresponding answer. All written explanations for both questions were also iteratively coded, detailed results of which are presented elsewhere (Brooks et al., in review).

Students predominantly chose the correct answer for the piston question (Figure 4.1) using scientifically valid reasoning, yet students who answered the balloon question (Figure 4.2) correctly predominantly used faulty reasoning related to

the ideal gas law. For the balloon question most students chose ‘remains the same.’ They did so because they apparently thought that since the balloon was “perfectly insulated”, no heat meant there could be no temperature change. The students that answered the piston question with ‘remains the same’ used different reasoning. They apparently thought that the decrease in pressure was compensated for by the increase in volume. Students who select this answer to the two conceptually similar questions do so using different reasons, and the word clouds capture this difference.

Table 4.1. Multiple-choice answer options, word clouds, and representative explanations for the concept question depicted in Figure 4.1 (piston question).

Answer Option	Word Cloud of Written Explanations	Representative Explanation (emphasis added)
remains the same	<p>air temperature pressure volume decrease increase perfectly insulated heat remain system</p>	<p>“if the pressure decreases and the volume increases then there will be no change in the temperature”</p>
increases	<p>temperature increased piston air gas law volume increase upward</p>	<p>“volume will increase so the temperature will increase also”</p>
decreases	<p>pressure air temperature decrease piston energy work internal volume heat weight system increase</p>	<p>“the system does work on the surroundings therefore it expends energy and the temperature decreases”</p>

Table 4.2. Multiple-choice answer options, word clouds, and representative explanations for the concept question depicted in Figure 4.2 (balloon question).

Answer Option	Word Cloud of Written Explanations	Representative Explanation (emphasis added)
remains the same	<p>balloon perfectly insulated remain heat increase volume temperature</p>	<p>the balloon is perfectly insulated so the temperature of the balloon does not change.</p>
increases	<p>nrt volume increase temperature pressure increase pv</p>	<p>$pV=nRT$, V is increase, so T is increase</p>
decreases	<p>balloon pressure temperature pv decrease gas energy molecules volume INCREASE nrt lower</p>	<p>Temperature must go down to maintain $PV=nRT$ relationship.</p>
need more information	<p>volume pressure decreasing balloon pv nrt temperature ideal gas law change constant increase decrease</p>	<p>$PV=nRT$, or $T=P/(nR)$. Because P is decreasing, but V is increasing, we need to know how exactly they are related in order to know if temperature is increasing or decreasing.</p>

4.5.3 So what can we learn from these word clouds?

First, let us focus on the correct answer, ‘decreases.’ The students who predominantly chose the correct answer for the piston question using correct reasoning

have a corresponding word cloud in which the words 'energy' and 'work' can be seen. However, students who answered the balloon question correctly, using faulty reasoning related to the ideal gas law, have a corresponding word cloud in which 'energy' is present but, 'work' does not appear; terms like "pv" and "nrt" can be seen instead. In the case of the balloon problem, we see an example of when a missing word is as important, or more so than the words that appear (Luhn, 1958; Chuang, et al. 2012). In this case, the word cloud without the word 'work' suggests that even though many students chose correctly, they may still need attention regarding the role of work in closed system energy balances.

We can also consider the word clouds associated with explanations for the distractors to provide insight into the ideas expressed by students who chose a wrong answer. The students that answered the piston question with 'remains the same' thought that the decrease in pressure was compensated for by the increase in volume. Notice that the words 'pressure', 'volume', 'decrease', and 'increase' are in almost equal proportion. For the balloon question, most students chose 'remains the same.' since the balloon was "perfectly insulated", no heat meant there could be no temperature change. Therefore, neither 'energy' nor 'work' appear in that word cloud. Again, the reasons for similar answers are different. In the case of the distractors, each distractor should have a particular misconceptions with which is it most associated, but it could have several. The word clouds may provide an instructor with enough information to identify which misconception is most prevalent for the majority of students who chose each distractor.

4.6 Word Clouds to Examine Muddiest Point Reflections

The second example illustrates the use of Muddiest Point Reflections for formative assessment. In a Muddiest Point Reflection, an instructor asks students to write a brief, anonymous written comment describing the concept or topic that they found to be the most difficult to understand during class (Mosteller, 1988; Angelo & Cross, 1993; Kelly, et al., 2010). With this information, the instructor can strategize to adjust his/her teaching and pedagogy to address issues specific to a many students. The CW software allows word clouds of Muddiest Point Reflections to be available either immediately after students have responded or a short time after their responses.

4.6.1 Methods

Data were collected in several materials science classes at a large public university with class sizes of 40-45 students. Figure 4.3 presents a screenshot of the Muddiest Point Reflection as it is presented to students on their laptops, cell phones, or tablets using the CW. The Muddiest Point Reflection was assigned at the end of class and students could answer on their electronic devices; however, the assignment was allowed to be submitted up to six hours after class because some students do not bring laptops or cell phones to class. Students were offered up to five percent extra credit on their final grade for answering at least 20 of the 24 Muddiest Point Reflections over the term. These exercises have an estimated 65% response rate. The data collected for this research was approved by the Institutional Review Board.

When the exercise was first presented to students, the instructor discussed with students the purpose of the exercise, both from a student learning and instructor

feedback standpoint. At the beginning of the class following each Muddiest Point Reflection submission, the instructor thanked the students for their submissions. In addition, the instructor showed the word cloud from the previous submission, presented student quotes, and led a discussion regarding the student learning issues that identified in the previous submission. The discussion used the method of Socratic questioning in working toward resolution of the student learning issues. The instructor also reiterated that responses to the Muddiest Point Reflection would help improve not only the course for the current cohort of students, but as well as future cohorts.

What was the muddiest point in lecture today?

How muddy is it?

Almost Clear	Slightly Muddy	Moderately Muddy	Very Muddy	Extremely Muddy
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>


Figure 4.3. Muddiest point reflection as it was delivered to students.

4.6.2 Results

The example word cloud presented here contains students' Muddiest Point Reflections after the topic of failure in metals was covered in class. This topic has important real-world consequences, since engineering systems such as airplanes, chemical plants and bridges are susceptible to failures with consequent loss of lives.

Table 4.3 gives the results of that Muddiest Point response exercise in the form of the word cloud and two representative sample quotes.

Table 4.3. Word cloud and sample quotes for the sample Muddiest Point Reflection.

Word cloud of Muddiest Point Reflections	Sample Quotes
 <p>The word cloud contains the following words: 'calculating', 'stress', 'crack', 'failure', 'mechanism', 'types', 'general', 'microstructural', 'appearance', 'fracture', 'microscopic', and 'views'. The words 'failure' and 'mechanism' are the largest, followed by 'types' and 'calculating'. 'stress', 'crack', 'general', 'microstructural', 'appearance', 'fracture', 'microscopic', and 'views' are smaller.</p>	<p>“classifying the general failure mechanism and microstructural mechanisms.”</p> <p>“Many of the mechanisms used unfamiliar vocabulary. Appearance.”</p>

4.6.3 So what can we and students learn from the Muddiest Point Reflection word cloud?

In the prior class discussion of this topic, the four main types of failures were described, along with the failure mechanisms, fracture appearances, and testing methods that have predictive capabilities. "Failure", "mechanism", and "types" were the largest words seen, indicating that failure types and associated mechanisms were the most prominent muddiest points as opposed to fracture appearances or testing methods. The major difficulty that a significant fraction of the students were grappling with was the connections between the different aspects of a given "type of failure mechanism", which was clearly reflected in the size of the words in the word cloud.

A reading of the student comments confirmed the diagnosis that was first quickly highlighted by the word cloud. In order to address this particular learning

issue, the instructor was inspired in the next class to create a well-detailed table about failure mechanism type and associated characteristics. The table, as such, clearly delineated the characteristics of the failure mechanism types and features and included: a real world example; conditions causing failure, mechanism of failure, fracture surface appearance, and test methods for predicting lifetime associated with different mechanisms. Most of the students vigorously took notes and copied the table during the discussion. This is just one example that illustrates how the use of word clouds in Muddiest Point Reflections helps the teacher who can improve and adjust instruction, as well as the students, who are empowered to impact their own learning through thoughtful reflections. The rapid feedback with the Muddiest Point and associated word cloud can have a significant impact on student learning.

Research has shown that addressing learning issues as quickly as possible with rapid feedback is very effective for improving motivation and learning (Shute, 2008). Frequent feedback plays an important role in the progression of a learner from the level of "novice" toward "expert" understanding & performance in a given domain. In a review on the acquisition of expert skills, Ericsson, et al. (1993) cites as one important condition for optimal learning and improving performance is that learners will *receive immediate and informative feedback and knowledge of results of their performance* on a given task. This is reflected by the response of the students to a survey about the use of word clouds.

In a section of the introductory materials science class an end of semester survey was administered to 33 students on the impact of the word cloud use in the

classroom. Sixty-seven per cent of the students agreed or strongly agreed that, “The word clouds helped me visualize what the most confusing concepts in the class were.” Seventy-six per cent of the students agreed or strongly agreed with the statement that, “the word clouds informed me about issues other students were having with the class content.” Overall, as an instructor the word clouds and Muddiest Point Reflection provide a quick and measured diagnosis of student learning issues for adjusting current and future instruction. For students, the word clouds serve as a visual indicator of issues that they and others in class may be grappling with and are more motivated to engage in discussion and dialogue in addressing those issues to improve their knowledge and learning on more difficult concepts and content. Thus, instructors and students are mutual beneficiaries of the use of word clouds in materials classes.

4.7 Word Clouds for other Short Written Exercises

In the previous two sections we discussed how word clouds have been used for specific types of exercises, ConcepTests and Muddiest Point Reflections. While further research is required to evaluate the utility of word clouds to examine other types of short written exercises, in this section we briefly explore a few other areas where word clouds may be beneficial. In general, word clouds can be used for any type of short written response. For example, Vigeant, Prince & Nottis (2011) describe inquiry-based activities for thermodynamics and heat transfer. In these activities students are prompted to predict results of an experiment before the experiment and explain their prediction in writing. They then have students run or observe an experiment. After experimentation, students compare results with their predictions in

writing, and discuss with their peers. Finally they write answers to post-activity questions. Each of the writing steps present an opportunity for word cloud use to visualize aggregate student responses. As their inquiry-based activities continue to be implemented at different schools in different contexts word clouds might offer another quick way to examine if the students in these new contexts give similar responses to those in the original context. Other scaffolded activities that have a similar ‘predict - observe - explain’ structure (Champagne, Klopfer & Anderson, 1980), such as the interactive virtual laboratories recently incorporated in the CW, may also benefit from word clouds.

4.8 Planned Word Cloud Improvements for the AIChE Concept Warehouse

For exercises like Muddiest Point Reflections and other short written exercises, the current word cloud analytical algorithm may be sufficient. However, this type of analysis may benefit from including the option of using word pairs (Viegas & Wattenberg, 2008) as a basis, an option we are currently exploring. In addition, we are considering modifications to address one of the concerns reported by Ramsden & Bate (2008); to prevent the elimination of seemingly common words with special meanings (it versus IT) we intend to incorporate a custom list in which instructors can identify words to exclude or words to include.

4.9 Conclusions & Implications

Active learning can help students develop deeper understanding of chemical engineering principles. While multiple-choice ConcepTests are useful, we advocate for including student writing in learning activities as well. Written explanations and

reflections can help students organize their thinking and explicitly reflect on what has been covered. Writing also provides information that faculty can use to focus instruction. Writing reveals students' faulty reasoning and misconceptions, and can help the instructor identify concepts and topics that are difficult.

In this article, we demonstrate that word clouds can provide a quick analytical technique to assess student written explanations and reflections. The AICHE Concept Warehouse automatically generates word clouds and can facilitate collection and analysis of student writing, even in large classes. Unlike external applications like wordle where text needs to be manually entered, word clouds are automatic and quick. We are also working to improve and better integrate word clouds and other analysis options into this tool to help make deep, concept-based learning more effective for students and easy for faculty to implement.

4.10 Acknowledgements

The authors would like to acknowledge funding from the National Science Foundation (DUE 1023099, 1225456, 1226325) and from Oregon State University (TRF 931, 1056). The opinions expressed are strictly those of the authors and do not necessarily represent those of the National Science Foundation.

4.11 References

- Angelo, T. A., & Cross, K. P. (1993). Classroom assessment techniques.
- Berson, I. R., & Berson, M. J. (2009). Making sense of social studies with visualization tools. *Social Education*, 73(3), 124-126.
- Brooks, B. J., White, R. M., Bowen, A. S., Higgins, A. Z., & Koretsky, M. D. (in review). Promoting Scientific Reasoning through Written Explanations to Multiple-Choice Concept Questions. *Journal of Engineering Education*.

- Carberry, A., Krause, S., Ankeny, C., & Waters, C. (2013). "Unmuddying" course content using muddiest point reflections. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Oklahoma, City, OK.
- Champagne, A. B., Klopfer, L. E., & Anderson, J. H. (1980). Factors influencing learning of classical mechanics. *American Journal of Physics*, 48(12), 1074-1079.
- Chuang, J., Ramage, D., Manning, C., & Heer, J. (2012). *Interpretation and trust: designing model-driven visualizations for text analysis*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Cidell, J. (2010). Content clouds as exploratory qualitative data analysis. *Area*, 42(4), 514-523.
- Clough, P., & Sen, B. (2008). Evaluating tagclouds for health-related information research. *Health Info Management Research*.
- Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics*, 1(1), 40-95.
- Dalton, B., & Grisham, D. L. (2011). eVoc strategies: 10 ways to use technology to build vocabulary. *The Reading Teacher*, 64(5), 306-317.
- Ericsson, K. A. , Krampe , R. T. , & Tesch-Römer , C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review* , 100 , 363–406.
- Felder, R. M. (1995). A longitudinal study of engineering student performance and retention. IV. Instructional Methods. *Journal of Engineering Education*, 84(4), 361-367.
- Godwin-Jones, R. (2006). EMERGING TECHNOLOGIES Tag Clouds in the Blogosphere: Electronic Literacy and Social Networking. *Language Learning & Technology*, 10(2), 8-15.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66, 64.
- Hall, S. R., Waitz, I., Brodeur, D. R., Soderholm, D. H., & Nasr, R. (2002). *Adoption of active learning in a lecture-based engineering class*. Paper presented at the Frontiers in Education, 2002. FIE 2002. 32nd Annual.
- Hamm, S. E. (2011). Using word clouds for reflection and discussion in an online class. *Teaching Theology & Religion*, 14(2), 156-156.
- Hayes, S. (2008). Wordle. *Voices from the Middle*, 16(2), 66-68.

- Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11(3), 203-270.
- Koretsky, M., Falconer, J., Brooks, B., Gilbuena, D., Silverstein, D., Smith, C., Miller, R., Miletic, M. (2014). The AIChE Concept Warehouse: A Tool to Promote Conceptual Learning. *Advances in Engineering Education*.
- Kelly, J., Graham, A., Eller, A., Baker, D., Tasooji, A., and Krause, S. (2010). Supporting student learning, attitude, and retention through critical class reflections. *2010 ASEE Annual Conference Proceedings*.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ 07458: Prentice Hall.
- McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The qualitative report*, 15(3), 630-643.
- Miley, F., & Read, A. (2011). Using Word Clouds to Develop Proactive Learners. *Journal of the Scholarship of Teaching and Learning*, 11(2), 91-110.
- Mosteller, F. (1988). Broadening the scope of statistics and statistical education. *The American Statistician*, 42(2), 93-99.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, 93(3), 223-231.
- Qiu, J. (2007). Traditional medicine: a culture in the balance. *Nature*, 448(7150), 126-128.
- Ramsden, A., & Bate, A. (2008). Using word clouds in teaching and learning.
- Rivard, L. O. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching*, 31(9), 969-983.
- Schute, V. J., (2008) Focus on Formative Feedback. *Review of Educational Research*, 78, 153-189.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122-124.
- teachingchannel.org. Text What You Learned: Using Technology to Assess. Retrieved 11/25/13, 2013, from <https://www.teachingchannel.org/videos/texting-to-assess-learning>
- VanOrden, N. (1987). Critical-thinking writing assignments in general chemistry. *Journal of Chemical Education*, 64(6), 506-507.
- VanOrden, N. (1990). Is writing an effective way to learn chemical concepts? Classroom-based research. *Journal of Chemical Education*, 67(7), 583-585.

- Vigeant, M., Prince, M., & Nottis, K. (2011). Fundamental Research in Engineering Education. Development of Concept Questions and Inquiry-Based Activities in Thermodynamics and Heat Transfer: An Example for Equilibrium vs. Steady-State. *Chemical Engineering Education*, 45(3), 211-218.
- Viégas, F. B., & Wattenberg, M. (2008). Timelines tag clouds and the case for vernacular visualization. *interactions*, 15(4), 49-52.
- Williams, W., Parkes, E. L., & Davies, P. (2013). Wordle: A method for analysing MBA student induction experience. *The International Journal of Management Education*.

5. Student Response Times to In-class Thermodynamics Concept Questions

Bill J. Brooks,* Dedra N. Demaree** & Milo D. Koretsky*

*School of Chemical, Biological, and Environmental Engineering

Oregon State University

Corvallis, OR, USA 97331-2702

(541)737-4591

** Center for New Designs in Learning & Scholarship

Georgetown University

3520 Prospect St. NW #314

Washington, DC 20057

(202)687-5295

American Journal of Physics

<http://scitation.aip.org/content/aapt/journal/ajp/>

(submitted for review)

5.1 Abstract

The time it takes students to respond to cognitive tasks is a consideration for psychometricians, cognitive scientists, and education researchers. In this study, we measured engineering students' response times to multiple choice thermodynamics concept questions while students were engaged in active learning in class. We examine response time differences between correct and incorrect answers, with and without written explanations to justify the answers, and as a function of question difficulty. Our findings show that students who answer correctly take less time to answer easy questions than those who answer incorrectly, but more time for more difficult questions. Response times are correlated with question difficulty when the correct answer is chosen, but not for incorrect answers. When accounting for difficulty, it takes students about 50 seconds more to provide a written explanation justifying their answer choice.

5.2 Introduction

Response times to cognitive tasks have allowed cognitive scientists and psychometricians to study various aspects of learning and its measurement. Response times have been used to examine thought process complexity (Bokenholt, 2012), the relationship between confidence and answer popularity (Koriat, 2008) and student aptitude in the context of item response theory (van der Maas et al., 2011; Abdelfattah, 2007). They have also been used to study student effort (Wise & DeMars, 2006; Wise & Kong, 2005) and response strategies (Mislevy, 1993). In the context of physics education, Lasry, Watkins, Mazur & Ibrahim (2013) found that students who chose

correct answers to conceptual multiple-choice problems responded faster than those who answer incorrectly. They also found that response time is inversely correlated with confidence, a finding consistent with cognitive science literature (Koriat, 2008). Lasry, et. al. (2013) express a need “for better understanding of the cognitive processes in answering [multiple-choice] test questions.” While that focus was on test questions, we posit the same is true of multiple-choice concept questions used in active learning pedagogies such as peer instruction (Mazur, 1997).

Peer instruction (PI) is a student centered active learning pedagogy that has been shown to be effective at fostering conceptual understanding (Lasry, Mazur & Watkins, 2008; Crouch & Mazur, 2001). It is a relatively simple modification to lecture that includes formative assessment in the form of polling using a concept question, also known as a “conceptest” (Mazur, 1997) or “clicker question” (Duncan, 2006). Depending on the outcome of the initial polling, students may be encouraged to discuss the reasoning for their answer with their peers before being polled again.

Educators considering the implementation of this type of active learning pedagogy refer to limited class time, including the time it takes students to answer questions, as a rationale not to adopt such pedagogies. Dancy and Henderson (2010) studied physics faculty approaches to research-based instructional strategies, citing a lack of research into the optimal time that should be spent engaging students in activities such as answering concept questions and discussing with their peers. They show that time is a factor that prevents faculty from using these more research-based instructional strategies. Of their respondents, 24.1% mentioned time generally as an

impediment and 8.0% explicitly said that research-based instructional strategies require too much class time that could otherwise be used to cover the course content. Similarly, time and coverage of content is a concern when using constructed response items like written explanations (Wainer & Thissen, 1993).

This study investigates the time it takes for thermodynamics students to answer multiple-choice concept questions in an active learning class environment. Sometimes the questions incorporated a prompt for students to write explanations justifying their answer choice. We investigate the influence that writing has on response time and how it relates to question difficulty in multiple thermodynamics courses. We ask the following research questions:

1. Do students in thermodynamics who chose the correct multiple choice answer respond faster or slower (if either) than students who chose an incorrect answer?
2. Does response time correlate with question difficulty, and if so, how?
3. Does response time increase significantly (and if so, by how much) when students are asked to write an explanation?
4. Does writing an explanation affect the response time of students who chose the correct multiple-choice answer in the same way as students who chose an incorrect answer?

5.3 Methods

The questions examined in this study were asked in two thermodynamics courses at a large public university over the span of six years. Student participants

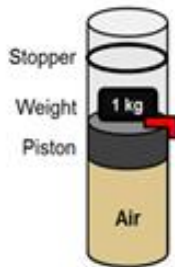
($N_{\text{student}} = 358$) were enrolled in chemical, environmental, or biological engineering. One course was at the introductory-level and required for 2nd year students in all three programs while the other is an advanced junior-level course required by chemical engineering majors and occasionally taken by students from the other disciplines. The courses were taught by the same two instructors over the time of the study. Almost all students in the advanced course had successfully completed the introductory course. The research was approved by the Institutional Review Board and participants signed informed consent forms.

The questions were delivered in class as part of active learning activities that often resulted in PI. Students answered individually on their laptops, tablets, or cellphones using database driven web pages that stored the responses (Koretsky & Brooks, 2008; Koretsky et al., 2014). The students had experience using both the specific technology and the associated active learning pedagogy in at least one prerequisite course prior to the courses studied here.

Figure 5.1 shows a sample conceptual, qualitative multiple choice question as it was presented to students on their electronic device. Generally, these questions are conceptually challenging and typically require little or no computation so that students cannot algorithmically rely on equations to obtain the answer. This question shows a case when students were prompted to write an explanation. Selection of questions and which questions incorporated a written explanation prompt was up to the instructors of the courses. To answer this sample question correctly, ideally students apply their knowledge of the conservation of energy, recognizing that the work done by the gas

on the surroundings lowers its internal energy (whether or not intermolecular interactions are considered) and, therefore, its temperature.

Air at high pressure and ambient temperature is contained in a perfectly insulated piston-cylinder device. If the locks holding the piston in place are removed, the piston moves upwards to a stopper. The temperature of the air _____.



- increases
- remains the same
- decreases

Please explain your answer in the box below.

Please rate how confident you are with your answer.

- | | | | | |
|-------------------------|-----------------------|-----------------------|-------------------------|----------------------------|
| substantially
unsure | moderately
unsure | neutral | moderately
confident | substantially
confident |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Figure 5.1. Sample concept question.

For this study, in-class responses to multiple-choice questions ($N_{\text{response}} = 12,028$) were recorded in a database. The data analyzed included: timestamp when the polling started, timestamp when each student answered, the student's multiple choice answer selection, whether an explanation was prompted, and the text of the student's explanation. There were 65 concept questions in total with 46 questions overlapping

between instances where a written explanation was prompted ($N_{\text{question}} = 59$) and when no explanation was requested ($N_{\text{question}} = 52$). Since we are interested in learning about cognitive processing as students respond to these questions, we only include initial delivery of questions in the analysis. In other words, if the instructor chose to ask the same question a second time (e.g., re-polling in PI), that instance would not be included in the study.

Response times were computed by subtracting the individual answer timestamp from the polling start timestamp. In the case of written explanations students submit the answer choice and explanation simultaneously (see Figure 5.1). We use correctness of answer selection (correct or incorrect), the existence of a written explanation, and question difficulty as independent variables. Response times were categorized as “written explanations” when an explanation had a non-zero length. Difficulty indices were calculated using the ratio of the total number of correct responses across conditions to number of total responses across conditions. Therefore, a difficulty index is the percent correct; i.e., a smaller number indicates a more difficult question (Steif & Dantzler, 2005).

5.4 Results

Table 5.1 provides a summary of average student response times. Response time was shortest when students chose correctly and did not write an explanation, resulting in an average of about three minutes. Response time was longest when students chose incorrectly and wrote an explanation, averaging almost five minutes. A multi-factor ANOVA shows that performance and written explanations both

contribute significantly to response time. Students who answer correctly are faster than those who answered incorrectly by an average of about twenty seconds whether or not students write an explanation ($F(1,12025) = 158, P < 0.001$). Furthermore, it takes an average of 1.3 additional minutes to write an explanation ($F(1,12025) = 1665, P < 0.001$).

Table 5.1. Average time taken (in seconds) for correct and incorrect answers, with and without written explanations. N indicates the number of responses in that condition.

	⟨t⟩ without explanation	⟨t⟩ with explanation
Correct	183 ($N_{\text{response}} = 3,036$)	261 ($N_{\text{response}} = 3,475$)
Incorrect	206 ($N_{\text{response}} = 2,865$)	288 ($N_{\text{response}} = 2,652$)
Correct and Incorrect	194	273

To look for general trends, we further grouped the data by question. We then performed a sign test to compare average response time between correct and incorrect responses on a per question basis. The sign test assumes that paired samples come from the same population with both identical spread and central tendency. If that assumption holds, the differences between randomly sampled pairs should have equal probability of being either negative or positive. Therefore, they should follow a binomial distribution with a probability of 0.5.

Of the questions where students did not write explanations ($N_{\text{question}} = 52$), average response times of students who chose correctly were shorter for 36 questions and longer for 16 questions ($P < 0.01$). This result is consistent with Table 5.1.

However, when students wrote explanations ($N_{\text{question}} = 59$), average response times

for correct answers were shorter for only 25 questions and longer for 34 questions ($P = 0.05$). This latter result is not consistent with Table 5.1 and indicates in the majority of cases where students provide written explanations, students who chose the correct answer are slower than those who chose incorrectly. This result prompted further investigation, identifying question difficulty as a potential contributing factor.

Figure 5.2 depicts the average response times versus the difficulty index for questions without a written explanation (5.2a - correct responses, 5.2b - incorrect responses). Similarly, Figure 5.3 shows the corresponding data when an explanation was written (5.3a - correct responses, 5.3b - incorrect responses). Correct response, irrespective of writing condition, appears correlated to question difficulty and incorrect responses do not. The correlation is stronger in the case where students write explanations. It appears that the students who correctly answered *easy* questions answer more quickly than the students who chose incorrectly. However, as the questions become more difficult, the students who chose correctly responded more slowly than the students who chose incorrectly. We define a “crossover difficulty index” as when the response time for correct answers equals the time for incorrect answers. The crossover difficulty index for the no written explanation group is 0.35 and the crossover difficulty for the written explanation group is 0.63.

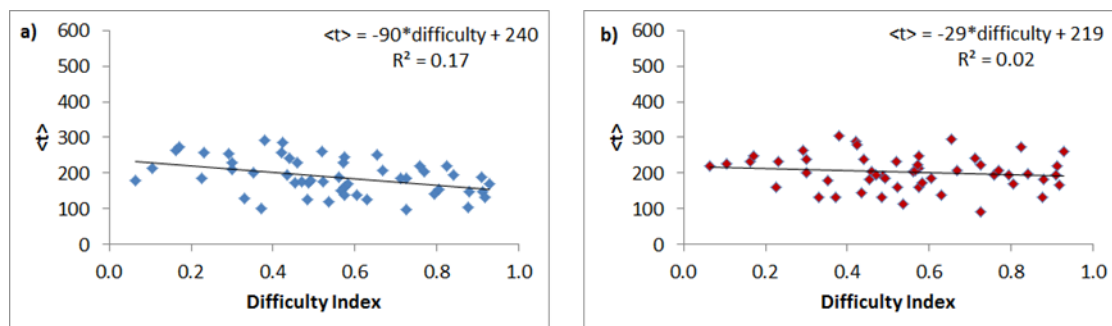


Figure 5.2. Average response time vs. question difficulty when students did not write an explanation for (a) correct answers ($P < 0.001$) and (b) incorrect answers.

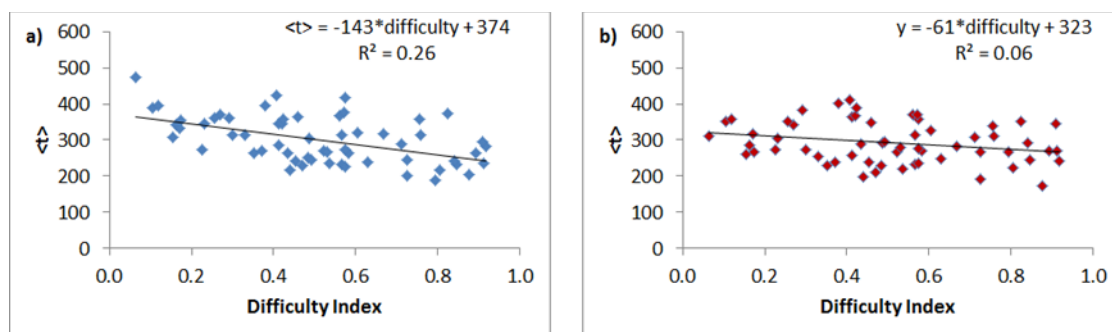


Figure 5.3. Average response time vs. question difficulty when students wrote an explanation for (a) correct answers ($P < 0.001$) and (b) incorrect answers.

A multivariate linear regression model with interactions was constructed to determine the effects of the three independent variables studied on response time. The model includes binary factors *correctness* in answer choice (+1 = correct; -1 = incorrect) and *explanation* (+1 = written explanation; -1 = no written explanation), together with the parametric variable *difficulty* index (0.0 to 1.0). Retaining only the significant terms, we obtain the following model equation that explains 53% of the variation in the measured response times to the set of 65 thermodynamics questions:

$$\begin{aligned}
 \text{ResponseTime(seconds)} = & 290 - 19 * \text{correctness}^{\#} - 49 * \text{explanation}^{\#\#} - 82 * \text{difficulty}^{\#\#} \dots \\
 & - 38 * \text{correctness} * \text{difficulty}^{\#} \\
 & (\#p < 0.05, \#\#p < 0.001)
 \end{aligned}$$

According to the model, difficulty, explanation, and correctness are all significant factors in response time, as well as an interaction of correctness and difficulty (as seen in Figures 5.2 and 5.3). Difficulty has the largest effect with difficult questions requiring up to 82 seconds longer than easy ones. The effect of writing an explanation adds an average of 49 seconds to the response time.

5.5 Discussion

We characterized the influence of correctness, prompting a written explanation, and question difficulty on the time it took for students to respond to multiple choice questions. Sixty-five questions in two different thermodynamics classes were studied. While the numerical values are specific to this study, the results are useful to the practitioner who is concerned about how class time is used. The largest influence on response time was question difficulty. A hard question could take students up to an average of 82 seconds longer than an easy question. Written explanations add a little less than a minute to the time it takes for students to respond. However, this “cost” of class time should be weighed in comparison to the benefit. Several education scholars have endorsed the use of written justification to improve student understanding (Taylor & Nolen, 2007; Linn & Chiu, 2011), and we advocate using written explanations, if possible. Like Lasry et al. (2013), we observed faster responses for correct answers, but the magnitude was only around 20 seconds (10%).

Additionally, there is an interaction between correctness and difficulty. Students who answer correctly are quicker on easy questions but slower on more difficult questions.

The results from this study can also be considered from the perspective of student thinking as they engage in these types of concept questions in lecture. Bokenholt (2012) posed a theory that categorizes thinking processes into two groups: simpler processes are “unconscious, rapid, effortless and automatic” and more complicated processes are “conscious, slow, effortful and deliberative.” The latter are both language- and rule-based. The more complicated processes are also potentially responsible for metacognitive processes of monitoring and may “override” the simpler processes. He reported response time data showing that simpler processes are faster than more complicated processes. This model can be applied to the response time data reported here as we posit that choosing correctly or writing explanations can both be related to more complicated thinking processes especially with difficult questions.

Response times for correct choices were found to be correlated to difficulty while incorrect choices were not. Such a result suggests that the thinking processes in students who answer correctly are associated with the difficulty of the question. Said another way, the way students who answer correctly think about the questions depends on the nature of the question. Conversely, while the number of students who choose incorrectly increases with question difficulty, the average response times do not. It appears when students do not understand the concept well enough to apply it to select the correct answer, they use similar thinking processes, whether the question is easy or difficult. Additionally, response times are correlated more strongly to difficulty when

students write explanations justifying their answer choice. We propose written explanations can elicit complex, deliberate thinking processes. In this case, students are prompted to organize their thinking to develop logical arguments that explain the reasoning in their answer choice (Brooks, et al., in review).

5.6 Conclusions & Implications

Correctness, written explanations, and difficulty all influence student response times to in class multiple-choice questions in thermodynamics. Written explanations increase response times by around 50 seconds. Students who answered correctly took longer to respond to more difficult questions while those who answered incorrectly took about the same amount of time regardless of question difficulty. Thus, students who chose the correct answer on easy questions responded faster than those who chose an incorrect answer, but slower on more difficult questions. We believe the difference may be due to better developed or well-reasoned thought processes on the part of the students who choose correctly.

5.7 Acknowledgements

The authors would like to acknowledge funding from the National Science Foundation (DUE 1023099) and from Oregon State University (TRF 931 and TRF 1056). The opinions expressed are strictly those of the authors and do not necessarily represent those of the National Science Foundation. We would also like to thank Debra Gilbuena for constructive feedback.

5.8 References

Abdelfattah, F. A. (2007). *Response latency effects on classical and item response theory parameters using different scoring procedures*. Ohio University.

- Brooks, B. J., White, R. M., Bowen, A. S., Higgins, A. Z., & Koretsky, M. D. (in review). Promoting Scientific Reasoning through Written Explanations to Multiple-Choice Concept Questions. *Journal of Engineering Education*.
- Böckenholt, U. (2012). The Cognitive-Miser Response Model: Testing for Intuitive and Deliberate Reasoning. *Psychometrika*, 77(2), 388-399.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69, 970.
- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78, 1056.
- Koretsky, M., & Brooks, B. (2008). *A Web-based Interactive Science and Engineering Learning Tool that Promotes Concept-Based Instruction*. Paper presented at the Proceedings of the Annual Conference of the American Society for Engineering Education.
- Koretsky, M., Falconer, J., Brooks, B., Gilbuena, D., Silverstein, D., Smith, C., Miller, R., Miletic, M. (2014). The AIChE Concept Warehouse: A Tool to Promote Conceptual Learning. *Advances in Engineering Education*.
- Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 945.
- Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, 76, 1066.
- Lasry, N., Watkins, J., Mazur, E., & Ibrahim, A. (2013). Response times to conceptual questions. *American Journal of Physics*, 81, 703.
- Linn, M. C., & Chiu, J. (2011). Combining learning and assessment to improve science education. *Research and Practice in Assessment*, 6(Winter 2011), 5-14.
- Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ 07458: Prentice Hall.
- Steif, P. S., & Dantzer, J. A. (2005). A statics concept inventory: Development and psychometric analysis. *Journal of Engineering Education*, 94(4), 363-371.
- Taylor, C., & Nolen, S. (2007). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson-Merrill-Prentice Hall.
- van der Maas, H. L., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological review*, 118(2), 339.
- Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, 6(2), 103-118.
- Wise, S. L., & DeMars, C. E. (2006). An Application of Item Response Time: The Effort-Moderated IRT Model. *Journal of Educational Measurement*, 43(1), 19-38.
- Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2),

163-183.

6. Conclusions

This dissertation presents four studies of technology-enabled pedagogy in the engineering classroom. The innovative technology supports pedagogy meant to foster conceptual understanding. The focus of this dissertation is examining the effect of prompting written explanations of answer choice selections to multiple choice concept questions. The outcomes of this research are valuable because engineers are often required to solve novel problems where conceptual understanding is critical. And conceptual understanding has been shown to improve through the use of active learning pedagogies.

Previous research has often been limited to studying aggregated effects over a term. Also, the literature with respect to variation in active learning pedagogies is sparse. Turpen and Finkelstein (2009) characterized variation in physics faculty's implementation of Peer Instruction and identified 13 dimensions of practice. This dissertation adds to the knowledge base applicable to some of these dimensions. The results from the use of written explanations can provide insight into 'clicker' question types, student-student collaboration, time 'constraints', student explanations, student voice, formative assessment, and prior knowledge. Fine grained analyses like these are needed so that practitioners can make informed decisions about pedagogy.

We present evidence that suggests engineering students are more likely to use more sophisticated, scientific reasoning when they write explanations to multiple choice concept questions, even if it does not help them choose the correct answer (Chapter 3). This finding provides empirical evidence supporting the writing to learn literature's assertion that writing goes beyond assessment and communication. The

writing to learn literature often takes a constructivist perspective and this finding can also be interpreted from a constructivist perspective. From the cognitive side, the improved reasoning we observed could be explained through a student's reorganization of prior knowledge and integration of new knowledge while interacting with the computer. From the sociocultural side, the use of better reasoning may be motivated by the construction of a socially acceptable logical argument. Undergraduate students in engineering ought to know that the use of scientific reasoning in defense of an answer is more acceptable to their instructors and ultimately future supervisors than, for example, answering "intuitively" or merely guessing.

We also show correctness, written explanations, and difficulty influence students' response times to multiple choice concept questions (Chapter 5). Writing requires an additional minute. Response time is directly proportional to difficulty when students choose correctly, irrespective of the writing. However, response time does not appear to depend upon difficulty when students choose wrong answers. These results imply different reasoning processes associated with choosing correct or wrong answers. These different reasoning processes reinforce and give added perspective to the findings of Chapter 3 on the influence of written explanations. These findings also extend some of the work of Lasry et al. (2013) and add to the body of literature for cognitive scientists and psychometricians. These findings on the influence of written explanations should also be kept in mind when considering the results of the study of peer instruction.

It is not surprising that there is a popularity effect with the use of peer instruction. We found that when an intermediate answer distribution is displayed students tend to change their answer choice in favor of the popular choice. This result has also been reported by Perez et al. (2010). However, we also found that students tended to change to the popular answer even when the intermediate answer distribution was not displayed, an effect we did not anticipate. The choice to display an intermediate answer distribution did influence self-reported confidence. Students are more confident when they are shown the intermediate result distribution and choose the popular answer choice after peer discussion. Students who choose the unpopular answer are less confident, even if it is correct. These results extend the work of Perez, add to the literature on peer instruction and active learning, and show broader applicability of what Koriat (2008) terms the consensuality principle.

The tendency to change to the popular answer without display of the intermediate distribution is surprising. One interpretation of the tendency toward choosing a popular answer is because of the perceived popularity of the answer itself. This interpretation would make sense if students were choosing the popular answer because of the display of an intermediate answer distribution. Another view is that the effect comes from a common interpretation of the question (mediated by common language, culture, experience, etc.) that makes the answer and associated reasoning common in the first place. This latter interpretation is supported by the apparent lack of influence of the display of the intermediate answer distribution on answer choice

when written explanations are solicited. However, local majorities could also have an influence.

While more research is needed to further explain the observation that students tend to choose the popular answer, there was often an improvement in quality of written explanation. This improvement is even seen for the students who choose correctly both before and after peer discussion as well as students who choose incorrectly before peer discussion. We can say written explanations generally improve after peer discussion. Presumably thinking and understanding improve as well. We saw some instances of co-construction. Co-construction has been reported based on multiple choice answers for cases where all group members choose incorrectly at first and then choose correctly after discussion (Singh, 2005; Smith et al., 2009). With our approach, we are able to observe instances of co-construction through analysis of students' written explanations which provide additional evidence and understanding. For example, in one group, one student had an acceptably correct explanation. Another student had a typical incorrect explanation. After discussion, the second student not only changed to the correct answer, but provided an exceptionally lucid justification, far superior to either explanation given by the first student (Koretsky and Brooks, 2011).

This research leads to a number of recommendations for practitioners. Practitioners should be encouraged to include written explanations and multiple choice concept questions in their classes. The use of word clouds is proposed as a tool to help practitioners quickly interpret written explanations. Analyzing individual written

explanations in real time can be difficult as class sizes become large. One way to help is by integrating the aggregation and summarization into a word cloud with the collection of the written explanations. Additionally, it may help if emphasis is placed on how practitioners should be attentive to the existence or lack of unexpected words as well as expected words.

Reasoning improves through discussions of answer choices with peers. Therefore, practitioners should be encouraged to use peer instruction. They should be further encouraged because reasoning improves for the high achieving students. Therefore the gains are not relegated to lower achieving students. However, practitioners should also be aware of the tendency of students to choose the popular answer whether it is correct or wrong. Sometimes the correct answer is not popular. There are also occasions where students choose the correct answer using faulty reasoning. Yaroch (1991) saw this tendency to use faulty reasoning in the selection of a correct answer choice decades ago. Others have reported students' misapplication of the ideal gas law (Rozier and Viennot, 1991; Loverude et al. 2002) as we observed with sophomores and juniors in chemical engineering. Students' serendipitously choosing the correct answer using the wrong reasoning is a significant problem that cannot be uncovered through the multiple choice answer selections alone. Practitioners are encouraged to spend the extra minute needed to also solicit written explanations when they poll students using multiple choice concept questions. These explanations add both increased fidelity to assessment in the polling process and help improve student reasoning.

Finally, the work in this dissertation provides some interesting preliminary data on the use of concept questions that have a 'real world' context. While there is limited data on this type of question, we found that students may have a stronger tendency to choose the correct answer using faulty reasoning. However, including questions with real world contexts help students with transfer. The important thing is that practitioners are attentive to the possibility that their students may use alternative thinking processes to answer these types of questions. This topic may be fruitful for further research.

Bibliography

- Abdelfattah, F. A. (2007). *Response latency effects on classical and item response theory parameters using different scoring procedures*. Ohio University.
- Allen, K. (2006). The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics. *Available at SSRN 2130143*.
- Angelo, T. A., & Cross, K. P. (1993). Classroom assessment techniques.
- Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, 54(4), 577-596.
- Asch, S. E. (1955). Opinions and social pressure. *Readings about the social animal*, 17-26.
- Barber, M., & Njus, D. (2007). Clicker evolution: seeking intelligent design. *CBE-Life Sciences Education*, 6(1), 1-8.
- Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36(2), 147-168.
- Barker, V., & Millar, R. (2000). Students' reasoning about basic chemical thermodynamics and chemical bonding: what changes occur during a context-based post-16 chemistry course? *International Journal of Science Education*, 22(11), 1171-1200.
- Berson, I. R., & Berson, M. J. (2009). Making sense of social studies with visualization tools. *Social Education*, 73(3), 124-126.
- Borrego, M., Cutler, S., Prince, M., Henderson, C., & Froyd, J. E. (2013). Fidelity of Implementation of Research-Based Instructional Strategies (RBIS) in Engineering Science Courses. *Journal of Engineering Education*, 102(3), 394-425.
- Bowen, G. M., Roth, W. M., & McGinn, M. K. (1999). Interpretations of graphs by university biology students and practicing scientists: Toward a social practice view of scientific representation practices. *Journal of Research in Science Teaching*, 36(9), 1020-1043.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*: National Academy Press Washington, DC.
- Brooks, B. J., Gilbuena, D. M., Falconer, J. L., Silverstein, D. L., Miller, R. L., & Koretsky, M. D. (2012). *Preliminary Development of the AIChE Concept Warehouse*. Paper presented at the ASEE Annual Conference and Exposition, San Antonio, TX.
- Brooks, B. J., & Koretsky, M. D. (2011). The influence of group discussion on students' responses and confidence during peer instruction. *Journal of Chemical Education*, 88(11), 1477-1484.
- Brooks, B. J., White, R. M., Bowen, A. S., Higgins, A. Z., & Koretsky, M. D. (in review). Promoting Scientific Reasoning through Written Explanations to Multiple-Choice Concept Questions. *Journal of Engineering Education*.
- Böckenholt, U. (2012). The Cognitive-Miser Response Model: Testing for Intuitive and Deliberate Reasoning. *Psychometrika*, 77(2), 388-399.

- Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. *CBE-Life Sciences Education*, 6(1), 9-20.
- Carberry, A., Krause, S., Ankeny, C., & Waters, C. (2013). "Unmuddying" course content using muddiest point reflections. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Oklahoma, City, OK.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1147.
- Champagne, A. B., Klopfer, L. E., & Anderson, J. H. (1980). Factors influencing learning of classical mechanics. *American Journal of Physics*, 48(12), 1074-1079.
- Chandrasegaran, A., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293-307.
- Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3), 439-477.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science*, 5(2), 121-152.
- Chuang, J., Ramage, D., Manning, C., & Heer, J. (2012). *Interpretation and trust: designing model-driven visualizations for text analysis*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Cidell, J. (2010). Content clouds as exploratory qualitative data analysis. *Area*, 42(4), 514-523.
- Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, 24(4), 467-563.
- Clough, P., & Sen, B. (2008). Evaluating tagclouds for health-related information research. *Health Info Management Research*.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9), 970-977.
- Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics*, 1(1), 40-95.
- Dalton, B., & Grisham, D. L. (2011). eVoc strategies: 10 ways to use technology to build vocabulary. *The Reading Teacher*, 64(5), 306-317.
- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78, 1056.
- Draper, S., Cargill, J., & Cutts, Q. (2002). Electronically enhanced classroom interaction. *Australian journal of educational technology*, 18(1), 13-23.
- Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of*

- computing in higher education*, 7(2), 3-47.
- Duncan, D. (2006). Clickers: A new teaching aid with exceptional promise. *Astronomy Education Review*, 5(1), 70.
- Elby, A. (1999). Another reason that physics students learn by rote. *American Journal of Physics*, 67, S52.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3), 363.
- Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The Physics Teacher*, 40, 206.
- Falconer, J. L. (2007). Conceptests for a Chemical Engineering Thermodynamics Course. *Chemical Engineering Education*, 41(2), 107-114.
- Felder, R. M. (1995). A longitudinal study of engineering student performance and retention. IV. Instructional Methods. *Journal of Engineering Education*, 84(4), 361-367.
- Gere, A. R., & Abbott, R. D. (1985). Talking about writing: The language of writing groups. *Research in the Teaching of English*, 19(4), 362-385.
- Godwin-Jones, R. (2006). EMERGING TECHNOLOGIES Tag Clouds in the Blogosphere: Electronic Literacy and Social Networking. *Language Learning & Technology*, 10(2), 8-15.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66, 64.
- Hall, S. R., Waitz, I., Brodeur, D. R., Soderholm, D. H., & Nasr, R. (2002). *Adoption of active learning in a lecture-based engineering class*. Paper presented at the Frontiers in Education, 2002. FIE 2002. 32nd Annual.
- Haláková, Z., & Prokša, M. (2007). Two kinds of conceptual problems in chemistry teaching. *Journal of chemical education*, 84(1), 172-174.
- Hamm, S. E. (2011). Using word clouds for reflection and discussion in an online class. *Teaching Theology & Religion*, 14(2), 156-156.
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89-120): Information Age Publishing.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. doi: 10.3102/003465430298487
- Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What Are They Thinking? Automated Analysis of Student Writing about Acid-Base Chemistry in Introductory Biology. *CBE-Life Sciences Education*, 11(3), 283-293.
- Hayes, S. (2008). Wordle. *Voices from the Middle*, 16(2), 66-68.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3), 141-158.
- Howard, V., & Barton, J. H. (1988). *Thinking on paper*. New York, NY: William Morrow Paperbacks.

- Imbrie, P. K., & Reed-Rhoads, T. (2011). *Expanding Access to Assessing Conceptual understanding: ciHUB, Virtual Community Supporting Conceptual Learning*. Paper presented at the ASEE 9th Global Colloquium on Engineering Education, Singapore.
- James, M. C. (2006). The effect of grading incentive on student discourse in peer instruction. *American journal of physics*, 74(8), 689-691.
- Kalman, C. S., Milner-Bolotin, M., & Antimirova, T. (2010). Comparison of the effectiveness of collaborative groups and peer instruction in a large introductory physics course for science majors. *Canadian Journal of Physics*, 88(5), 325-332.
- Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education*, 53(3), 819-827.
- Kelly, J., Graham, A., Eller, A., Baker, D., Tasooji, A., & Krause, S. (2010). *Supporting student learning, attitude, and retention through critical class reflections*. Paper presented at the 2010 ASEE Annual Conference Proceedings.
- Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning*, 21(4), 260-268.
- Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11(3), 203-270.
- Klein, P. D. (2006). The challenges of scientific literacy: From the viewpoint of second-generation cognitive science. *International Journal of Science Education*, 28(2-3), 143-178.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4(4), 298-310.
- Koretsky, M., & Brooks, B. (2008). *A Web-based Interactive Science and Engineering Learning Tool that Promotes Concept-Based Instruction*. Paper presented at the Proceedings of the Annual Conference of the American Society for Engineering Education.
- Koretsky, M., Falconer, J., Brooks, B., Gilbuena, D., Silverstein, D., Smith, C., . . . Miletic, M. (2014). The AIChE Concept Warehouse: A Tool to Promote Conceptual Learning. *Advances in Engineering Education*.
- Koretsky, M., Miller, R., Falconer, J., Prince, M., Vigeant, M., Krause, S., & Silverstein, D. (2011). *SPECIAL SESSION: Educational Methods and Tools to Encourage Conceptual Learning*. Paper presented at the American Society for Engineering Education Annual Conference & Exposition, Vancouver, B.C., Canada.
- Koretsky, M. D., & Brooks, B. J. (2011). Comparison of Student Responses to Easy and Difficult Thermodynamics Conceptual Questions during Peer Instruction. *International Journal of Engineering Education*, 27(4), 897-908.
- Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

- 34(4), 945.
- Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. J. (2004). *Development, testing, and application of a chemistry concept inventory*. Paper presented at the Frontiers in Education, Savannah, Georgia.
- Lasry, N., & Aulls, M. W. (2007). The effect of multiple internal representations on context-rich instruction. *American Journal of Physics*, 75, 1030.
- Lasry, N., Charles, E., Whittaker, C., & Lautman, M. (2009). *When Talking is Better than Staying Quiet*. Paper presented at the AIP Conference Proceedings, Ann Arbor, MI.
- Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, 76, 1066.
- Lasry, N., Watkins, J., Mazur, E., & Ibrahim, A. (2013). Response times to conceptual questions. *American Journal of Physics*, 81, 703.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*: Cambridge university press.
- Linn, M. C., & Chiu, J. (2011). Combining learning and assessment to improve science education. *Research and Practice in Assessment*, 6(Winter 2011), 5-14.
- Loverude, M. E., Kautz, C. H., & Heron, P. R. (2002). Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas. *American Journal of Physics*, 70(2), 137-148.
- Lucas, A. (2009). Using peer instruction and i-clickers to enhance student participation in calculus. *Primus*, 19(3), 219-231.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ 07458: Prentice Hall.
- Mazur, E. (2009). Farewell, lecture. *Science*, 323(5910), 50-51.
- McConnell, D. A., Steer, D. N., Owens, K. D., Knott, J. R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H. (2006). Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, 54(1), 61.
- McDermott, L. C. (2001). Oersted Medal Lecture 2001: "Physics Education Research—The Key to Student Learning". *American Journal of Physics*, 69(11), 1127-1137.
- McDermott, M. A., & Hand, B. (2010). A secondary reanalysis of student perceptions of non-traditional writing tasks over a ten year period. *Journal of Research in Science Teaching*, 47(5), 518-539.
- McKenna, A. F. (2007). An investigation of adaptive expertise and transfer of design process knowledge. *Journal of Mechanical Design*, 129(7), 730-734.
- McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The qualitative report*, 15(3), 630-643.
- Meltzer, D. E. (2004). Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course. *American Journal of Physics*, 72, 1432.

- Midkiff, K. C., Litzinger, T. A., & Evans, D. (2001). *Development of engineering thermodynamics concept inventory instruments*. Paper presented at the Frontiers in Education Conference, 2001. 31st Annual.
- Miley, F., & Read, A. (2011). Using Word Clouds to Develop Proactive Learners. *Journal of the Scholarship of Teaching and Learning*, 11(2), 91-110.
- Miller, R. L., Streveler, R. A., Nelson, M. A., Geist, M. R., & Olds, B. M. (2005). *Concept inventories meet cognitive psychology: Using beta testing as a mechanism for identifying engineering student misconceptions*. Paper presented at the Proceedings of the American Society for Engineering Education Annual Conference.
- Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiken, R. Mislevy & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19-39).
- Mosteller, F. (1988). Broadening the scope of statistics and statistical education. *The American Statistician*, 42(2), 93-99.
- Mulford, D. R., & Robinson, W. R. (2002). An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *Journal of Chemistry Education*, 79(6), 739 - 744.
- Nayak, L., & Erinjeri, J. P. (2008). Audience response systems in medical student education benefit learners and presenters. *Academic Radiology*, 15(3), 383-389.
- Nelson, M. A., Geist, M. R., Miller, R. L., Streveler, R. A., & Olds, B. M. (2007). *How to create a concept inventory: The thermal and transport concept inventory*. Paper presented at the Annual Conference of the American Educational Research Association, Chicago, IL.
- Newcomer, J. L., & Steif, P. S. (2008). Student thinking about static equilibrium: Insights from written explanations to a concept question. *Journal of Engineering Education*, 97(4), 481-490.
- Nicol, D. J., & Boyle, J. T. (2003). Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education*, 28(4), 457-473.
- Nurrenbern, S. C., & Pickering, M. (1987). Concept learning versus problem solving: Is there a difference? *Journal of Chemical Education*, 64(6), 508.
- Nurrenbern, S. C., & Robinson, W. R. (1998). Conceptual questions and challenge problems. *Journal of Chemical Education*, 75(11), 1502-1503.
- Odell, L. (1980). The process of writing and the process of learning. *College Composition and Communication*, 31(1), 42-50.
- Papaphotis, G., & Tsaparlis, G. (2008). Conceptual versus algorithmic learning in high school chemistry: the case of basic quantum chemical concepts. Part 1. Statistical analysis of a quantitative study. *Chemistry Education Research and Practice*, 9(4), 323-331.
- Parker, W. C., Lo, J., Yeo, A. J., Valencia, S. W., Nguyen, D., Abbott, R. D., . . . Vye, N. J. (2013). Beyond Breadth-Speed-Test Toward Deeper Knowing and Engagement in an Advanced Placement Course. *American Educational Research Journal*, 50(6), 1424-1459.

- Peasley, K. L., Rosaen, C. L., & Roth, K. J. (1992). *Writing-to-learn in a conceptual change science unit*. East Lansing, MI: Center for the Learning and Teaching of Elementary Subjects, Institute for Research on Teaching, Michigan State University.
- Penuel, W. R., Abrahamson, L., & Roschelle, J. (2006). Theorizing the transformed classroom: A sociocultural interpretation of the effects of audience response systems in higher education. *Audience response systems in higher education: Applications and cases*, 187-208.
- Perez, K. E., Strauss, E. A., Downey, N., Galbraith, A., Jeanne, R., & Cooper, S. (2010). Does displaying the class results affect student discussion during peer instruction? *CBE-Life Sciences Education*, 9(2), 133-140.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science education*, 66(2), 211-227.
- Poulis, J., Massen, C., Robens, E., & Gilbert, M. (1998). Physics lecturing with audience paced feedback. *American Journal of Physics*, 66, 439.
- Prain, V., & Hand, B. (1996). Writing for learning in secondary science: Rethinking practices. *Teaching and Teacher Education*, 12(6), 609-626.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, 93(3), 223-231.
- Qiu, J. (2007). Traditional medicine: a culture in the balance. *Nature*, 448(7150), 126-128.
- Ramsden, A., & Bate, A. (2008). Using word clouds in teaching and learning.
- Rivard, L. O. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching*, 31(9), 969-983.
- Rowell, P. M. (1997). Learning in school science: The promises and practices of writing. *Studies in Science Education*, 30, 19-56.
- Rozier, S., & Viennot, L. (1991). Students' reasonings in thermodynamics. *International Journal of Science Education*, 13(2), 159-170.
- Sanger, M. J., & Badger, S. M. (2001). Using computer-based visualization strategies to improve students' understanding of molecular polarity and miscibility. *Journal of Chemical Education*, 78(10), 1412.
- Sawrey, B. A. (1990). Concept learning versus problem solving: Revisited. *Journal of Chemical Education*, 67(3), 253.
- Scardamalia, M., & Bereiter, C. (1986). *Research on written composition* (M. Wittrock Ed. Vol. 3). New York, NY: Macmillan.
- Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, 87(3), 433.
- Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). *Efficiency and innovation in transfer*. Charlotte, NC: Information Age Publishing, Inc.
- Schwartz, D. L., Chase, C. C., & Bransford, J. D. (2012). Resisting overzealous transfer: Coordinating previously successful routines with needs for new learning. *Educational Psychologist*, 47(3), 204-214.

- Sensenbaugh, R. (1989). ERIC/RCS: Writing Across the Curriculum: Evolving Reform. *Journal of Reading*, 32(5), 462-465.
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Simon, B., Kohanfars, M., Lee, J., Tamayo, K., & Cutts, Q. (2010). *Experience report: Peer instruction in introductory computing*. Paper presented at the Proceedings of the 41st ACM technical symposium on Computer science education.
- Singh, C. (2005). Impact of peer interaction on conceptual test performance. *American journal of physics*, 73(5), 446-451.
- Smith, M., Wood, W., Krauter, K., & Knight, J. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE-Life Sciences Education*, 10(1), 55-63.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122-124.
- Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: Development and psychometric analysis. *Journal of Engineering Education*, 94(4), 363-371.
- Steif, P. S., Lobue, J. M., Kara, L. B., & Fay, A. L. (2010). Improving problem solving performance by inducing talk about salient problem features. *Journal of Engineering Education*, 99(2), 135-142.
- Tamir, P. (1989). Training teachers to teach effectively in the laboratory. *Science education*, 73(1), 59-69.
- Tamir, P. (1990). Justifying the selection of answers in multiple choice items. *International Journal of Science Education*, 12(5), 563-573.
- Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. *Cell Biology Education*, 4(2), 112-117.
- Taylor, C., & Nolen, S. (2007). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson-Merrill-Prentice Hall.
- teachingchannel.org. Text What You Learned: Using Technology to Assess. Retrieved 11/25/13, 2013, from <https://www.teachingchannel.org/videos/texting-to-assess-learning>
- Tierney, R. J., Soter, A., O'Flahavan, J. F., & McGinley, W. (1989). The effects of reading and writing upon thinking critically. *Reading Research Quarterly*, 24(2), 134-173.
- Turpen, C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction. *Physical Review Special Topics-Physics Education Research*, 5(2), 020101.
- Urban-Lurain, M., Prevost, L., Haudek, K. C., Henry, E. N., Berry, M., & Merrill, J. E. (2013). *Using Computerized Lexical Analysis of Student Writing to Support Just-in-Time Teaching in Large Enrollment STEM Courses*.
- Vahey, P., Tatar, D., & Roschelle, J. (2007). *Using handheld technology to move*

- between private and public interactions in the classroom.* (M. Hooft & K. Swan Eds.). Mahwah, NJ: Erlbaum.
- van der Maas, H. L., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological review*, *118*(2), 339.
- Van Dijk, L., Van Der Berg, G., & Van Keulen, H. (2001). Interactive lectures in engineering education. *European Journal of Engineering Education*, *26*(1), 15-28.
- VanLehn, K., & Van de Sande, B. (2009). *Expertise in elementary physics, and how to acquire it* (K. Ericsson Ed.). New York, NY: Cambridge University Press.
- VanOrden, N. (1987). Critical-thinking writing assignments in general chemistry. *Journal of Chemical Education*, *64*(6), 506-507.
- VanOrden, N. (1990). Is writing an effective way to learn chemical concepts? Classroom-based research. *Journal of Chemical Education*, *67*(7), 583-585.
- Vigeant, M., Prince, M., & Nottis, K. (2011). *Engineering undergraduates' conceptual understanding of thermodynamics: Assessment and change after normal instruction*. Paper presented at the Proceedings from Hawaii International Conference on Education, Honolulu, HI.
- Vigeant, M., Prince, M., & Nottis, K. (2011). Fundamental Research in Engineering Education. Development of Concept Questions and Inquiry-Based Activities in Thermodynamics and Heat Transfer: An Example for Equilibrium vs. Steady-State. *Chemical Engineering Education*, *45*(3), 211-218.
- Viégas, F. B., & Wattenberg, M. (2008). Timelines tag clouds and the case for vernacular visualization. *interactions*, *15*(4), 49-52.
- Voss, J. L., & Paller, K. A. (2010). What makes recognition without awareness appear to be elusive? Strategic factors that influence the accuracy of guesses. *Learning & Memory*, *17*(9), 460-468.
- Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, *6*(2), 103-118.
- Watters, D. J., & Watters, J. J. (2007). Approaches to learning by students in the biological sciences: Implications for teaching. *International Journal of Science Education*, *29*(1), 19-43.
- Williams, W., Parkes, E. L., & Davies, P. (2013). Wordle: A method for analysing MBA student induction experience. *The International Journal of Management Education*.
- Willoughby, S. D., & Gustafson, E. (2009). Technology talks: Clickers and grading incentive in the large lecture hall. *American Journal of Physics*, *77*(2), 180-183.
- Wise, S. L., & DeMars, C. E. (2006). An Application of Item Response Time: The Effort-Moderated IRT Model. *Journal of Educational Measurement*, *43*(1), 19-38.
- Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee

- motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163-183.
- Xie, C., & Lee, H.-S. (2012). A visual approach to nanotechnology education. *International Journal of Engineering Education*, 28(5), 1006-1018.
- Yarroch, W. L. (1991). The implications of content versus item validity on science tests. *Journal of Research in Science Teaching*, 28(7), 619-629.

APPENDICES

Appendix A – The Influence of Group Discussion on Students' Responses and Confidence during Peer Instruction Supporting Information

Bill J. Brooks and Milo D. Koretsky

Supporting information for the following paper:

Journal of Chemical Education

<http://pubs.acs.org/doi/abs/10.1021/ed101066x>

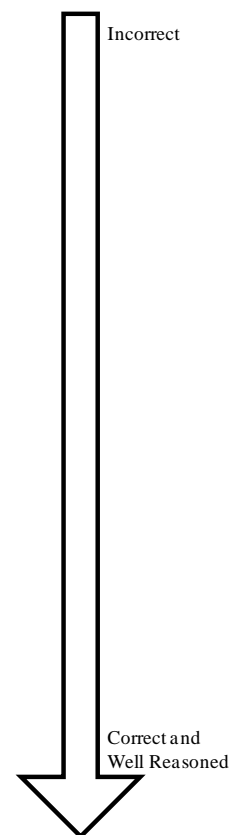
November 2011, Volume 88, Issue 11, pages 1477-1484

A.1 Coding Scheme

The coding scheme used for the 5 question pairs in this study is shown in Table A.1 The coding methodology is described in the manuscript.

Table A.1 Coding scheme for conceptual questions.

Code	Throttling Valve	Spray Can	Mixing	Equilibrium	Adiabatic Air
1	Ideal gas law – implicit constant volume assumption	Attractive forces would hold the molecules together, reducing the pressure	Since delta h is positive, energy is released, raising the temperature	Condenses because of constant volume	Generally incorrect – adiabatic means no temperature change
2	Ideal gas law – explicit constant volume assumption	Pressure increases because of the attractive molecular interactions	Some element is missing, i.e. positive delta h means endothermic and temperature goes down	Vaporizes because of Raoult's Law	Ideal gas law – P is directly proportional to T
3	Misconceived energy balance	Enthalpy is less, to achieve equilibrium, entropy must be less and pressure greater	The positive delta h means energy is needed for the mixing and it comes from sensible heat	Saturation pressure and liquid mole fraction are constant so water must vaporize to increase vapor phase mole fraction	Implicit energy balance – system does work, losing heat
4	Proper energy balance	Same as 3 and explicitly states that the increase in pressure comes from decrease in entropy	Same as 3 and mixing requires energy because the unlike intermolecular interactions have less stability	Same as 3 and includes a statement about how no matter how much the vapor phase mole fraction can be increased, it can't reach the original equilibrium conditions	Explicit energy balance – some of the energy in the initial temperature will be converted to work to raise the piston



A.2 Conceptual questions used in this study

The seven concept questions, as presented to the students using the WISE interface are presented. Figure A.1 shows *Throttling Valve*, as provided to students before group discussion while Figure A.2 shows the same exercise after group discussion where students selected their group members. Figure A.3 shows *Equilibrium*; Figure A.4 shows *Spray Can*; Figure A.5 shows *Mixing*; Figure A.6 shows *Adiabatic Air*; Figure A.7 shows *Solid-Liquid Melt*; and Figure A.8 shows *Chem Reaction*.

WISE Learning Tool
Oregon State UNIVERSITY

Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and the pressure drops through the valve to a lower value, P_2 .

Assuming the valve is well insulated and inlet and outlet pipes connected to the valve are the same diameter, what is the relationship of the outlet temperature T_2 to the inlet temperature T_1 ?

$T_2 < T_1$
 Can't answer until I know what gas is flowing
 $T_2 > T_1$
 $T_2 = T_1$

Multiple choice answers

Please explain your reasoning.

Short answer follow-up explanation

Please rate how confident you are with your answer.

substantially unsure moderately unsure neutral moderately confident substantially confident

Confidence follow-up

Submit

Source: Ron Miller

Figure A.1. *Throttling Valve* exercise pre discussion



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

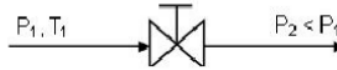
- [Virtual Hand Raise Exercises](#)
- [Study Group](#)
- [My Statistics](#)
- [Homepage](#)

Logged in as:
Milo ([logout](#))

Contact point:
[Milo Koretsky](#)

An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and the pressure drops through the valve to a lower value, P_2 .

Assuming the valve is well insulated and inlet and outlet pipes connected to the valve are the same diameter, what is the relationship of the outlet temperature T_2 to the inlet temperature T_1 ?



- $T_2 = T_1$
- Can't answer until I know what gas is flowing
- $T_2 < T_1$
- $T_2 > T_1$

Multiple choice answers

Please explain your reasoning.

Short answer follow-up explanation

Please rate how confident you are with your answer.

- | | | | | |
|-------------------------|-----------------------|-----------------------|-------------------------|----------------------------|
| substantially
unsure | moderately
unsure | neutral | moderately
confident | substantially
confident |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Confidence follow-up

Please select group members.

- Minimize/Expand selection
- Student A
- Student B
- Student C
- Student D
- Student E
- Student F

Group member selection

Source: Ron Miller

Figure A.2. Throttling Valve exercise post discussion



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

[Virtual Hand Raise](#)

[Exercises](#)

[Study Group](#)

[My Statistics](#)

[Homepage](#)

Logged in as:

[\(logout\)](#)

Contact point:

[Milo Koretsky](#)

Pure water is present and is in vapor liquid equilibrium at 1.2 atm in a piston/cylinder assembly. You inject 5 cm³ of air into the system, but keep P and T constant. What happens?

- All the water vaporises.
- All the water condenses
- Some of the water vaporises.
- Some of the water condenses.

Multiple choice answers

Explain your answer

Short answer follow-up explanation

Please rate how confident you are with your answer.

- | | | | | |
|-------------------------|-----------------------|-----------------------|-------------------------|----------------------------|
| substantially
unsure | moderately
unsure | neutral | moderately
confident | substantially
confident |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Confidence follow-up

Submit

Source: John Falconer

Figure A.3. Equilibrium exercise



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

[Virtual Hand Raise Exercises](#)
[Study Group](#)
[My Statistics](#)
[Homepage](#)

Logged in as:
[\(logout\)](#)

Contact point:
[Milo Koretsky](#)

In class, we solved the problem shown below. In the solution presented in class, the ideal gas model was assumed and the pressure was calculated to be 9.1 atm. If we were to solve this problem accounting for CF_2Cl_2 as a real gas, how would the final pressure compare to the 9.1 atm calculated in class?

Class examples

Consider the use of CF_2Cl_2 as a dispersing agent for aerosol spray cans. Estimate the pressure a can has to withhold at 40°C . Its enthalpy of vaporization at its normal boiling

point (244 K) is $\Delta h_{\text{vap}} = 20.25 \left[\frac{\text{kJ}}{\text{mol}} \right]$.

Why is CF_2Cl_2 no longer used in spray cans?

- The pressure would be less than 9.1 atm.
- The pressure would be greater than 9.1 atm.
- The pressure would still be equal to 9.1 atm.
- There is no way to determine how the pressure would change

Multiple choice answers

Please explain your reasoning below.

Short answer follow-up explanation

Please rate how confident you are with your answer.

substantially unsure moderately unsure neutral moderately confident substantially confident

Confidence follow-up

Submit

Figure A.4. *Spray Can exercise*



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

[Virtual Hand Raise](#)

[Exercises](#)

[Study Group](#)

[My Statistics](#)

[Homepage](#)

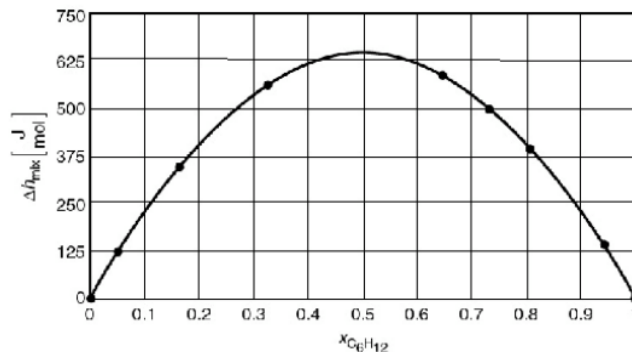
Logged in as:

[\(logout\)](#)

Contact point:

[Milo Koretsky](#)

The enthalpy of mixing for a mixture of cyclohexane and toluene is shown below. Consider the adiabatic mixing of 1 mole of cyclohexane and 1 mole of toluene at 25 °C. The final temperature will be:



- Less than 25 °C
 Greater than 25 °C
 Equal to 25 °C
 Need more information

Multiple choice answers

Explain your answer.

Short answer follow-up explanation

Please rate how confident you are with your answer.

- substantially
 unsure
-
- moderately
 unsure
-
- neutral
-
- moderately
 confident
-
- substantially
 confident
-

Confidence follow-up

Figure A.5. *Mixing exercise*



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

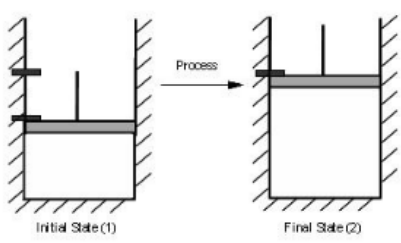
- [Virtual Hand Raise Exercises](#)
- [Study Group](#)
- [My Statistics](#)
- [Homepage](#)

Logged in as:
[\(logout\)](#)

Contact point:
[Milo Koretsky](#)

Air at high pressure and ambient temperature is contained in a perfectly insulated piston-cylinder assembly as shown. Stops prevent the piston from moving up. The stops are then removed and the piston quickly rises into the atmospheric pressure air above it until a second set of stops is encountered that prevents it from leaving the cylinder.

The temperature of the air in the cylinder:



- Decreases
- Increases
- Remains the same
- Insufficient information

Multiple choice answers

Explain Your answer

Written explanation

Please rate how confident you are with your answer.

- | | | | | |
|-------------------------|-----------------------|-----------------------|-------------------------|----------------------------|
| substantially
unsure | moderately
unsure | neutral | moderately
confident | substantially
confident |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Confidence follow-up

Source: Midkiff, Clark

Figure A.6. Adiabatic Air exercise



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

[Virtual Hand Raise](#)

[Exercises](#)

[Study Group](#)

[My Statistics](#)

[Home Page](#)

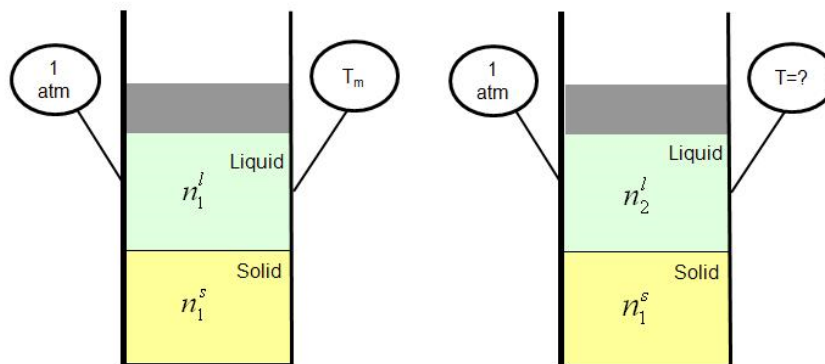
Logged in as:

[\(logout\)](#)

Contact point:

[Milo Koretsky](#)

The diagram on the left shows the normal melting point of pure solid 1 to be T_m . Consider now that the same pure solid 1 is melting into a liquid that initially contains pure 2, as shown on the right. How does the temperature at which 1 will melt into the liquid, T , compare to the case on the left? You may assume that the enthalpy of mixing of liquid 1 and 2 is zero.



- You cannot tell without more information
 $T > T_m$
 $T = T_m$
 $T < T_m$

Multiple choice answers

Explain your answer [using the concepts presented in class this term](#)

Short answer follow-up explanation

Please rate how confident you are with your answer.

substantially moderately neutral moderately substantially
 unsure unsure confident confident

Confidence follow-up



Submit

Figure A.7. Solid-Liquid Melt exercise



Web-based Interactive Science and Engineering Learning Tool

Conceptualization Exercise

[Virtual Hand Raise](#)

[Exercises](#)

[Study Group](#)

[My Statistics](#)

[Home Page](#)

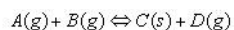
Logged in as:
[\(logout\)](#)

Contact point:
[Milo Koretsky](#)

Consider the reaction below where gaseous components A and B react to form solid C. This reaction proceeds to equilibrium at 500 K and 4 bar where the following quantities of species are present:

- 0.25 mol A
- 0.25 mol B
- 0.5 mol C
- 0.5 mol D

What is the equilibrium constant at 500 K? You may assume ideal gas behavior.



- 0.25
- 2
- 0.5
- 1
- 4
- The value of K is different than those provided above.

Multiple choice answers

Explain your answer

Short answer follow-up explanation

Please rate how confident you are with your answer.

substantially unsure moderately unsure neutral moderately confident substantially confident

Confidence follow-up

-
-
-
-
-

Submit

Figure A.8. Chem Reaction exercise

A.3. ANOVA of Student Confidence

Table A.2 shows the results of multi-factor ANOVA with main effects and interactions for the data set. Consensuality, correctness, and pre/post group discussion show an influence on reported confidence. Two significant interactions - the display of an intermediate bar graph with pre/post group discussion and consensuality with correctness - influence student confidence as well.

Table A.2. ANOVA: Confidence as a function of Correctness, Consensuality, Pre/Post Peer Interaction, and whether or not a bar graph of intermediate responses was displayed.

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:Correct	4.82	1	4.82	5.38	0.020
B:Consensual	13.59	1	13.59	15.16	0.000
C:Pre/Post PI	22.17	1	22.17	24.73	0.000
D:Intermediate Result	0.94	1	0.94	1.05	0.306
INTERACTIONS					
AB	16.26	1	16.26	18.14	0.000
AC	0.59	1	0.59	0.65	0.419
AD	0.27	1	0.27	0.30	0.582
BC	1.29	1	1.29	1.44	0.230
BD	1.59	1	1.59	1.77	0.183
CD	3.93	1	3.93	4.39	0.036
RESIDUAL	667.89	745	0.90		
<hr/>					
TOTAL (CORRECTED)	747.32	755			

Appendix B – Preliminary Development of the AIChE Concept Warehouse

Bill J. Brooks, Debra M. Gilbuena, John L. Falconer, David L. Silverstein, Ron L. Miller, and Milo D. Koretsky

Proceedings of the 119th American Society for Engineering Education Annual Conference & Exposition, June 2012, San Antonio, Texas

B.1 Introduction

The AIChE Concept Warehouse is being developed with the goal of creating a community of learning within the discipline of chemical engineering (ChE) focused on concept-based instruction. Many engineering educators and industry partners emphasize the need for students to apply their knowledge to new and challenging problems.¹ In order to do so, our students must *learn with understanding*.² A lack of conceptual understanding has been shown to severely restrict students' ability to solve new problems, since they do not have the functional understanding to use their knowledge in new situations.³ However, science and engineering classrooms often reward students more for rote learning than for conceptual understanding.^{4, 5} There is clearly a need for more emphasis on conceptual understanding and concept-based instruction.

Concept-based instruction (e.g., ConcepTests, concept inventories) often depends on high quality concept questions. These questions can be time consuming and difficult to construct, posing one of the biggest barriers keeping faculty from implementing this type of pedagogy.^{6, 7} *The AIChE Concept Warehouse* decreases this barrier by housing questions pertinent to courses throughout the core chemical engineering (ChE) curriculum (Material and Energy Balances, Thermodynamics, Transport Phenomena, Kinetics and Reactor Design, Process Control, and Materials Science). This cyber-enabled infrastructure is maintained through the Education Division of the American Institute of Chemical Engineers (AIChE), the discipline's major professional society. With careful and intentioned design to promote widespread

use, this tool has the potential to catalyze the use of concept-based pedagogy throughout the chemical engineering education community. In this paper we present a description of the current status and available features, which should prove useful for educators who are interested in incorporating concept-based instruction into their courses. Similarly, a description of the design and development process is included to provide potential designers with a useful reference to inform future design processes. In addition, we report on initial deployment, community building activities, and future plans for the *AICHE Concept Warehouse*.

B.1.1 Concept-based Pedagogy

Concept-based pedagogies have been studied in the physics education research community for decades and provide a model that engineering education researchers have been adopting. Two concept-based pedagogical tools have dramatically reshaped how conceptual teaching and learning are viewed in college physics classrooms: ConcepTests⁸ and concept inventories.³ Both of these tools require high quality concept questions in order to be effective. High quality concept questions are typically multiple choice, conceptually challenging, and require little to no calculation so students cannot rely on equations to reach an answer. In the following, we provide a brief description of ConcepTests and concept inventories. In addition, we discuss a framework, Diffusion Theory, used to inform development and promote widespread adoption.

B.1.2 ConcepTests & Active Learning

In his book *Peer Instruction*, Mazur describes the use of *ConcepTests* to engage students in conceptual learning during lecture.⁸ *ConcepTests* consist of one or a small number of concept questions. *ConcepTests* can be used with peer instruction in class. The process includes structured questioning in which all students respond independently. Students then discuss their answers in small groups and respond again individually. Peer instruction encourages students to reflect on conceptual problems, think through the arguments being developed, and put them into their own words. It also provides both student and instructor with feedback regarding student understanding of the concept being tested.

Studies of more than 5,000 science and engineering students have found that classes using active learning methods such as peer instruction had double the conceptual learning gains⁹ and 25% higher pass rate¹⁰ than traditional lecture. Deslauriers et al. compared the performance of two sections of students studying electromagnetic waves; the section taught by an inexperienced postdoctoral fellow using active learning techniques outperformed, by 2.5 standard deviations, the section taught by an experienced professor with a record of high evaluations.¹¹ The use of *ConcepTests* and peer instruction has recently been reported in chemical engineering, with similar positive results.¹²⁻¹⁴

B.1.3 Concept Inventories

While *ConcepTests* are used to facilitate active learning, concept inventories are one tool used to evaluate and demonstrate the effectiveness of active learning. The

first concept inventory, the Force Concept Inventory (FCI),^{3,15} provided a reliable and valid instrument to measure students' fundamental conceptual understanding of Newtonian mechanics. Concept inventories are valid and reliable instruments that consist of high quality concept questions. Validity is a measure of how well a concept inventory measures the intended concepts, as evaluated by experts, student observations, or other means within a single population of students or a variety of populations.¹⁶ Reliability is a measure of the degree to which repeated administrations of a concept inventory produce the same results.¹⁶ Concept inventories are used as an objective pre/post measure of an intervention and have been used to inform instruction by identifying student misconceptions. Since development of the FCI, concept inventories have been created in a variety of engineering subjects, including statics,¹⁷ dynamics,¹⁸ and fluid mechanics¹⁹ among others.

A set of concept inventories pertinent to chemical engineering have also been developed. One of these, the Thermal and Transport Concept Inventory (TTCI),^{20, 21} covers concepts in heat transfer, fluid mechanics, and thermodynamics. Research used to develop the TTCI also informed the construction of the Heat and Energy Concept Inventory,²² which was used to show the effectiveness of inquiry-based activities on repairing misconceptions in heat transfer and thermodynamics.²³ The Materials Concept Inventory has been used to assess conceptual gains in introductory materials engineering courses.²⁴ Other concept inventories in chemical engineering that have been used to assess student learning include, the Engineering Thermodynamics Concept Inventory,²⁵ and the Material and Energy Balances Concept Inventory.²⁶

B.2 Design for Diffusion

Although significant effort has been made in engineering education research to identify best practices in teaching, such as the cases discussed above, widespread adoption of best practices can be met with resistance and is often slow.²⁷ For widespread use, a bridge needs to be formed between innovation and general practice. The intent of the *AIChE Concept Warehouse* is to provide such a bridge between chemical engineering researchers developing concept inventories and ConcepTests and chemical engineering educational practitioners in the classroom.

Diffusion theory

provides a useful framework to inform the development process and identify the design elements that will better enable widespread

Table B.1. Attributes of an innovation described by Rogers²⁸

Attribute	Description
Relative advantage	“the degree to which an innovation is perceived as better than the idea it supersedes.” (p229) ²⁸
Compatibility	“the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters.” (p240) ²⁸
Complexity	“the degree to which an innovation is perceived as relatively difficult to understand and use.” (p257) ²⁸
Observability	“the degree to which the results of an innovation are visible to others.” (p258) ²⁸
Triability	“the degree to which an innovation may be experimented with on a limited basis.” (p257) ²⁸

adoption. Diffusion, defined by Rogers, is “the process in which an innovation is communicated through certain channels over time among the members of a social system.”²⁸ Diffusion theory classifies the perceived attributes of an innovation that most impact the rate at which it is adopted. These attributes, described in Table B.1, are being attended to both in initial design and throughout development.

B.2.1 The AIChE Concept Warehouse – February 2012 Status

Currently, the *AIChE Concept Warehouse* has more than 1,200 concept questions available for searching, viewing, and using in courses through the user interfaces. Student and instructor interfaces are available at <http://cw.edudiv.org> for the community, and university faculty can obtain an account through this site. A third interface, for administration purposes, includes moderator and testing functions such as adding news bulletins, organizational information, and debugging. The administrator interface will not be described in detail in this paper.

B.2.2 Adding Questions & Concept Inventories

One relative advantage of the *AIChE Concept Warehouse* is the availability of quality concept questions

for core chemical

engineering courses. A

substantial effort has been

made to populate the

database with questions

during the initial development

period. Figure B.1 shows

the total number of questions available in the first year of development.

Questions have been incorporated from the authors' personal lists, ConcepTest PowerPoint files from www.learncheme.com, and Materials Science PowerPoint files provided by Stephen Krause (ASU). Complete concept inventories are also available including each of three portions of the Thermal and Transport Concept Inventory

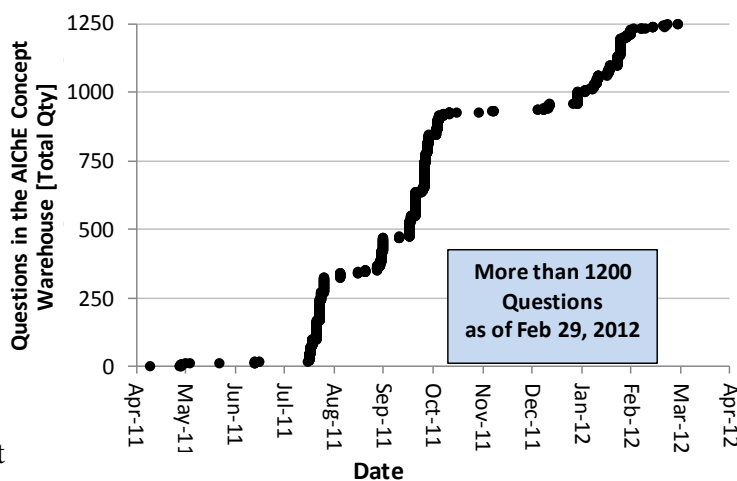


Figure B.1. Number of questions added to the *AIChE Concept Warehouse* during the initial development period.

(fluids, thermodynamics, and heat transfer),¹⁶ the Heat and Energy Concept Inventory,²² and the Materials Concept Inventory.²⁹ These questions and concept inventories are accessible via the user-friendly interfaces that have been developed.

B.2.3 Instructor Interface

The instructor interface is organized into seven main sections, accessible by their corresponding tabs: Home, ConcepTests, Concept Inventories, Classes, Profile, Review, and Resources. Table B.2 provides a description of the functions available to instructors in the *AICHE Concept Warehouse*, organized by submenu sections within each tab.

Table B.2. Summary of instructor interface functions organized by sub menu sections within each tab

Tab	Submenu Sections	Description of Functions
Home		View a brief summary of the latest changes, including: highlights of <i>AICHE Concept Warehouse</i> news, added questions, new tutorials and comments about submitted questions.
ConcepTests	Search	View, filter, and search for questions. Then, select question(s) for use in class.
	Manage Tests	Organize, group, download (MS Power Point, MS Word), or assign (via projection in-class or sent to student laptops or smartphones) ConcepTests. Confidence and short answer explanation prompts can be added to questions during assignment.
	Statistics	View information after questions have been answered, including all or a subset of options: question title and text, question images, percent correct, answer choice distribution, average confidence, number of students answered, and pedagogy recommendation.
	New Question	Add new questions. Question information can include: question title, question and answer images, answer options, comments for faculty, and applicable research data. Question types include: multiple choice – single right answer or multiple right answer, short answer, and ranking. Questions can also be tagged by class, misconception, topic, and textbook.
Concept Inventories	Browse	View available concept inventories and select for use in class. Additional concept inventory information is available for viewing, such as: research data, development history, list of individual questions and an answer key.
	Manage Inventories	Assign concept inventories (via projection in-class or sent to student laptops or smartphones), either in complete form or subsections. Confidence and short answer explanation prompts can be added to questions during assignment.
	Statistics	View information after concept inventory questions have been answered, including all or a subset of options: question name, question text, question image, percent correct, answer distribution, average confidence, and number of students answered.

Classes	Class List	Create and delete personal classes and associate personal classes with general classes.
	Manage Class	Manage the class roster and grade sheet. Add students to or remove them from the class roster. View and download the grade sheet, which includes student responses to questions (correct/incorrect, written responses, written explanations, confidence)
Profile	Preferences	Set personal preferences such as: show or hide tooltips, show or hide answer option comments, and show or hide the correct answer indicator in question previews.
	Demographic Information	Report institution, schedule type, approximate first year teaching, and approximate first year using active learning.
	My Clicker	Select clicker type, download clicker integration application, and register clicker receiver.
	User Agreement	Displays the end user license agreement (EULA). Accepting the EULA is a required prerequisite to use of the <i>AIChE Concept Warehouse</i> and is displayed upon initial log-in.
Review	Under development	
Resources	Intro to Pedagogy	View a collection of journal articles and videos to facilitate the integration of concept-based pedagogy into their classes.
	Tutorials	Watch videos to help instructors use the <i>AIChE Concept Warehouse</i> .
	FAQ	Find answers to frequently asked questions (e.g., How do I search for concept questions?)
	Chat Forum	Interact with other community members through a bulletin board.
	Helpful Links	Links to external resources on concept-based pedagogy, and active learning.
	News Archive	View the full history of the news about the <i>AIChE Concept Warehouse</i> .

In order to maximize compatibility and minimize complexity, an effort was made to design the instructor interface to match with the current practices of potential adopters, be familiar, and user-friendly. One way of accomplishing this design objective was to predict and accommodate different way users might leverage the *AIChE Concept Warehouse*. The remainder of this section presents a few potential scenarios of how an instructor might engage, how they fit into the community, and the features that enable them.

If an instructor would like to participate at a basic level they can simply find and select a set of concept questions to download (via MS Word or PowerPoint). They will have access to easily search through the concept questions currently available, and

can use these questions on homework, on tests, or with external clicker systems as they please. In this way, instructors already using peer instruction or active learning with concept questions need only minor changes to current practices and the *AICHE Concept Warehouse* may even save them valuable preparation time. They also need not expose their students to the site or be involved further themselves.

If an instructor wants to use more of the features available, instead of downloading questions they can integrate the use of clickers or have students log in and answer ConcepTests and inventories on their laptops or smart phones (either in-class or for homework). If instructors solicit responses via laptops or smartphones, they can prompt short answer explanations and confidence follow-ups in addition to the multiple choice answers. Such written reflection is perceived by students as helpful¹². These more involved features require students to interface with the site; the student interface is described in the next section.

For faculty, the results from assignments are aggregated, tabulated and archived for later use. They are also downloadable in MS Excel format. Finally, for the engineering education research community, usage data can be used to inform ConcepTest and concept inventory development. For example, data from student selections could be used for item testing, a critical step in concept inventory development. This synergy will allow the question pool for use in concept inventories to greatly expand.

In addition, instructors may actively contribute by adding their own questions to the database. They can create multiple choice, multiple correct multiple choice,

ranking, and short answer questions. These questions can then be associated with classes, the misconceptions they are testing, textbooks, and topics. Comments can also be left for instructors when they are searching for or deciding on using questions. These comments can include information about the correct answer, a distractor, or the use of the question.

B.2.4 Student Interface

Having only three tabs, the student interface is much simpler than the instructor interface. The tabs, submenu sections, and function descriptions are presented in Table B.3. The questions tab, shown in Figure B.2, shows the highest priority concept-based assignment for a selected class. It is in the questions tab that students can answer questions from in-class activities or homework assignments. Also depicted in Figure B.2, is the student view of three question components: a multiple choice question, a short answer follow-up, and a confidence follow-up. A sample (incorrect) student answer is also shown.

Table B.3. Summary of AICHE Concept Warehouse Student functions organized by tab and submenu sections


Tab	Submenu Sections	Description of Functions
Home		Alerts students to the number of concept-based assignments with unanswered questions, categorized by class.
Questions		View the highest priority concept-based assignment for a selected class. Students can answer questions from in-class activities or homework assignments.
Profile	Demographics	Report voluntary demographic information (e.g., birth year, first year at the university, gender, race, and major).
	Informed consent	Students can allow or deny the use of their response data for research purposes.
	Clicker registration	Aligns clicker responses with the student's account.

Class : **Thermodynamics** ▾

Time remaining: 00:03:10

An **ideal gas** flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 and T_1 and the pressure drops through the valve to a lower value, P_2 .

Assuming the valve is well insulated and inlet and outlet pipes connected to the valve are the same diameter, what is the relationship of the outlet temperature T_2 to the inlet temperature T_1 ?



$T_2 > T_1$
 $T_2 = T_1$
 $T_2 < T_1$
 Can't tell without knowing the type of gas.

Please explain your answer in the box below.

since well insulated and no work is being done, deltaU has to equal zero. This means that the energy lost from the drop in pressure would lead to an increase in temperature.

Please rate how confident you are with your answer.

substantially moderately neutral moderately substantially
 unsure unsure confident confident

Figure B.2. Screenshot of a student's answer from the questions tab from the student interface

To obtain an account, students provide their email address and click the “log in” button, while leaving the password field empty. A link is then emailed to the provided email address that allows students to create a password. The same process is used to reset forgotten passwords. Student email addresses and passwords are encrypted; therefore password reminders cannot be sent. After an account is setup, students can log in for the first time and are greeted with a prompt to provide

voluntary user information and an opportunity to provide informed consent to allow their anonymized responses to be incorporated into aggregate usage information for research purposes, as shown in the profile tab in Table B.3.

Technology use in the classroom, in the form of laptops,^{30, 31} tablet PCs,^{32, 33} and clickers³⁴ has been well documented. The use of internet-capable cell phones is also emerging as a learning tool, seen in a recent study on active learning³⁵ and as evidenced by audience response system companies releasing cell phone applications as an alternative to purchasing separate hardware.³⁶⁻³⁹ The student interface was designed with these considerations in mind and is enabled by the software structure detailed in the next section.

B.3 The AIChE Concept Warehouse – Design & Development

B.3.1 Software Structure

The *AIChE Concept Warehouse* software structure is based on a synergy between a web-based user interface (programmed using PHP 5.3) and a commercial database (MySQL 5.1). Advantages of this combination are presented in Figure B.3. The *AIChE Concept Warehouse* is being developed on both Windows and Linux platforms to accommodate developer preferences, institutional constraints, and provide ease of portability. Instructor and student usernames and passwords, along with any links between student and instructor accounts are all encrypted.

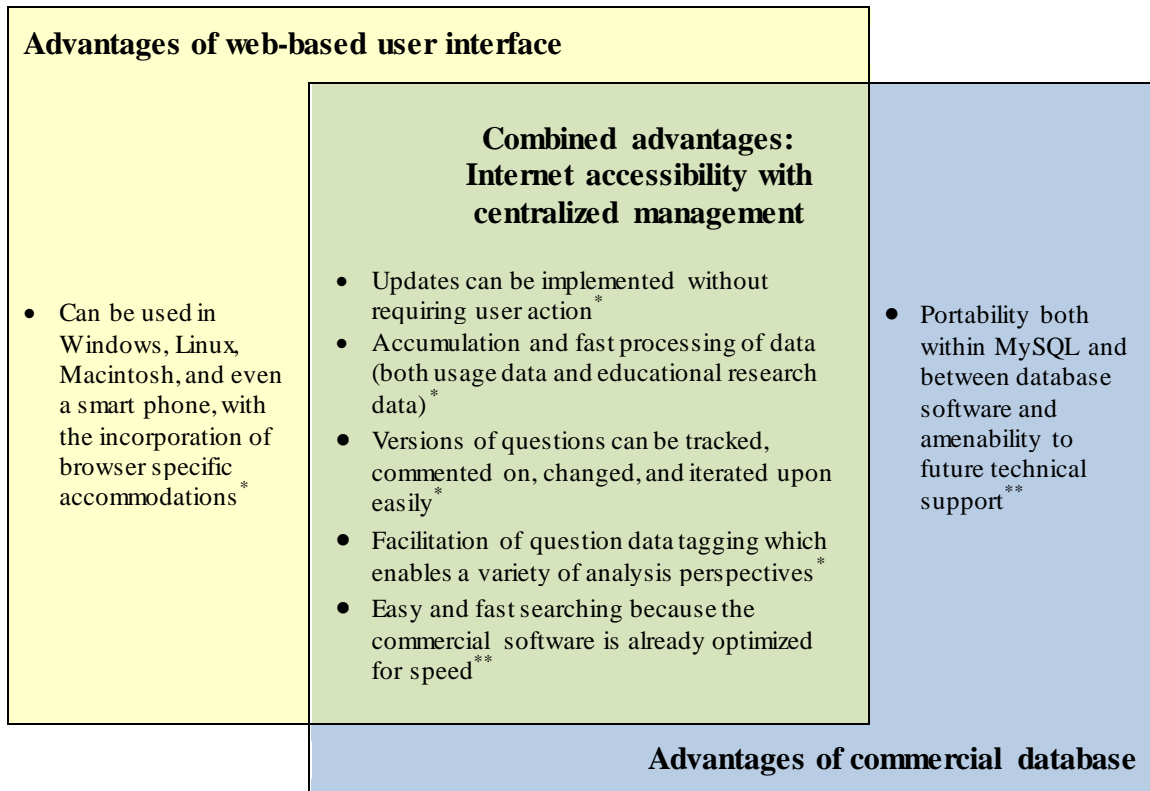


Figure B.3. Advantages of a web-based user interface and a commercial database *O'Reilly⁴⁰, **Bulger et al.⁴¹

Development of the *AIChE Concept Warehouse* software structure leverages the development experience of two of the authors on another education-focused tool. This previously developed tool, the Web-based Interactive Science and Engineering (WISE) Learning Tool, also has a web-based user interface and is database driven.⁴² Other similarities include a concept question focus, the facilitation of active, concept-based pedagogies and efficient data collection. While there are some similarities, there are other key differences. The WISE Learning Tool was designed for individual instructors to manage and deliver questions to their own classes and track research data for their classes alone. With the WISE Learning Tool, the questions lacked the

capability to be easily shared between instructors. In the *AIChE Concept Warehouse* instructors can still manage their own questions and classes; however, they can also share their questions with the community and search through and use any of questions others have added. Some research data in the *AIChE Concept Warehouse* is also automatically aggregated. Greater care has also been taken to make the user interface of the *AIChE Concept Warehouse* as simple as possible while offering necessary and useful features.

B.3.2 User Interface Design & Development

While an underlying structure was being developed, design of the user interfaces was also underway. Initially the design team generated a list of functions that needed to be incorporated into the *AIChE Concept Warehouse*. Separate function lists were generated for each interface. Next, each function list was categorized and grouped into website sections. A similar design process is used for each section, which includes the following steps:

- Developing a function list to be incorporated into the page
- Creating a storyboard of a page that includes the listed section functions
- Implementing the storyboard concept in a live webpage,
- Design team testing of live webpages and modification to enhance functionality and usability.
- External testing of live webpages and modification to enhance functionality and usability.

To illustrate the design process, we provide a detailed description of the evolution of the *question search* section starting from the function list and progressing to the latest version.

The initial function list for the question search section was comprised of three simple functions: (i) find questions, (ii) view a list of questions, and (iii) select questions for use. A storyboard was created in Microsoft PowerPoint which was intended to be familiar to a typical user and incorporate the ability to carry out the specified functions through various user input options (e.g., buttons, links, checkboxes, text input fields). In some cases, storyboards also incorporated features and functions that were not originally included in the function list.

The question search section storyboard, shown in Figure 4, was created in November 2010. To accomplish the first specified function, find questions, two aspects were incorporated (a) filtering options and (b) a keyword search, labeled accordingly with yellow shaded circles in Figure B.4. Filtering would allow the user to check a filter option box, resulting in a question list that only contained questions pertinent to the checked filter option. The keyword search would be similar to a typical web search, and only return questions pertinent to the keywords used. Along with the keyword search, an advanced search option was envisioned but not fully detailed in the storyboard.

(a) Filtering options – when selected each filter category would appear in the lower half of the filter sidebar with a list of category-specific filters

(b) Keyword search with an advanced search option

(c) List of question previews

(d) Add button - allows an instructor to add a question to the list of questions they want to use in class

The screenshot shows a web interface for searching questions. At the top, there are navigation tabs: Home, Classes, Tests, Author, Review. Below these is a search bar containing the text "Throttling valve, ideal gas, energy balance" and an "Advanced Search" button. The search results are displayed in a table with four rows, each showing a "Throttling Valve" question. Each row includes a diagram of a valve, a description of the question, and metadata such as author, date, class, and usage. An "Add to Set" button is visible at the end of each row.

Figure B.4. Step 2: Storyboard for the question search section with keyword search and filtering options highlighted, November 2010.

The storyboard included a list of question previews, labeled (c) in Figure B.4, as a way to view the information for multiple questions in a relatively compact, yet informative manner, accomplishing the second function. The four question previews in this storyboard are all identical and preview the “Throttling Valve” question, adapted from the TICI.¹⁶ Question previews included: the question title, a question image if one was associated with the question, the first ~150 characters of the question text, the question author(s), the date the question was added to the *AICHe Concept Warehouse*, question type, statistical information, class association, concept association, and usage information.

Each question preview also included a button in the lower right corner which would allow a question to be added to the instructor's personal list of questions, the list of questions that the instructor intended to use in his or her class. This button, one of which is labeled (d) in Figure B.4, would afford the third required function, allowing the instructor to select questions for use.

After storyboards were designed, they were used as a basis to create active prototype pages with partial functionality. Access to these pages was limited to the design team. An initial page, prototype V1, for the question search section is shown in Figure 5, from February 2011.

AICHE CONCEPT WAREHOUSE Education Division username [logout](#) | [help](#)

HOME CLASSES TESTS AUTHOR REVIEW

MANAGE SETS CURRENT SET BUILD SET RESULTS/STATISTICS

Filter By

Showing 0-2 of 2 results

General Classes

- Design (0)
- Energy Balances (0)
- Fluid Dynamics (0)
- Material Balances (3)
- Material Science (0)
- Reaction Engineering (0)
- Transport Phenomena (0)
- Thermodynamics

<p>Throttling Valve</p>	<p>An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P1 and T1 and the pressure drops t...</p> <p>Author(s): Brooks, Ron Miller Class: Thermodynamics Type: MC, SA Overall rating ★★★★☆</p> <p><input type="button" value="Add to Set"/></p>
<p>Adiabatic Air</p>	<p>Air at high pressure and ambient temperature is contained in a perfectly insulated piston-cylinder assembly as shown. Stops prevent the piston from mo...</p> <p>Author(s): Brooks, Clark Midkiff Class: Thermodynamics Type: MC Overall rating ★★★★☆</p> <p><input type="button" value="Add to Set"/></p>

OSU UK Copyright 2011 AIChE Concept Warehouse

Figure B.5. Step 3: Implementing the question search section of AICHE Concept Warehouse prototype V1, February 2011.

Where navigation in the “Tests” submenu was limited and largely unspecified in the storyboard (Figure B.4), prototype V1 (Figure B.5) has four functional submenu links: manage sets, current set, build set, and results/statistics. These working submenu titles continued to change over the course of development. There were cosmetic differences between the storyboard and the prototype V1, and certain features illustrated in the storyboards were excluded in the initial prototype. In the question search section exclusions were advanced searching, search result sorting, and some of the question information. Other aspects included in the initial prototype were not in the original storyboards. For example, an overall rating for each question was included that was intended to serve as a rough rating of question quality based on user feedback. Along with research data and user comments, similar to those available at Amazon.com, the overall rating could serve as a metric for instructors to reference as they evaluated which questions to use in their classes.

The initial prototype pages were tested by the members of the design team and revised for user-friendliness and functionality. Figure B.6 shows a later revision, prototype V2, of the question search section, dated August 2011.

The screenshot displays the AICHE Concept Warehouse interface. At the top, the logo 'AICHE Education Division CONCEPT WAREHOUSE' is visible on the left, and 'Silver username help' is on the right. A navigation bar contains buttons for HOME, QUESTIONS, INVENTORIES, CLASSES, REVIEW, PROFILE, and RESOURCES. Below this is a secondary navigation bar with buttons for SEARCH, NEW QUESTION, CURRENT TEST, MANAGE TESTS, and RESULTS/STATISTICS.

The main content area shows a search for 'pressure'. A search bar contains the text 'pressure' and a 'Search' button. To the right of the search bar is a link for 'Advanced Search'. Below the search bar, it indicates 'Showing 1-10 of 188 results'. On the left side, there is a sidebar with a 'Classes' section containing a list of categories: Energy Balances, Fluid Dynamics, Heat Transfer, Mass Transfer, Material Balances, Materials Science, Other, Reaction, Engineering, Workshop, and Thermodynamics (which is checked). Below the sidebar are sections for 'My Questions', 'Topics', 'Type', and 'Unused (self)'. At the bottom of the sidebar is a 'Suggestions' button.

The search results are displayed in a grid. The first result is titled 'Modified Throttling Valve'. It includes a diagram of a throttling valve with inlet conditions P_1, T_1 and outlet conditions $P_2 < P_1$. The text of the question is: 'An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1 '. The author is 'Brooks, Ron Miller', the class is 'Energy Balances Thermodynamics', the type is 'Multiple Choice', and the overall rating is five stars. An 'Add to Set' button is located at the bottom right of the result box.

The second result is titled 'Equilibrium Constant Change with Pressure'. The text of the question is: 'For the gas phase reaction: $A \rightarrow 3B$. The equilibrium constant is 10 at 2 atm.' The author is 'Falconer', the class is 'Thermodynamics', the type is 'Multiple Choice', and the overall rating is five stars. An 'Add to Set' button is located at the bottom right of the result box.

At the bottom of the results grid, a 'Suggestions' section is partially visible, showing 'Heated Ideal Gas Tank' and 'A rigid tank contains CO_2 at 2 atm and $50^\circ C$. When the

Figure B.6. Step 4: Revision after design team testing of the question search section of AICHE Concept Warehouse prototype V2, August 2011.

In prototype V2, additional functionality was integrated. The question previews were transformed from a purely informational display to incorporate links to more complete question summary pages. These links are apparent in Figure B.6 by the presence of underlined font in all of the question related text. Webpage naming and organization also evolved over time. The “Build Set” submenu became the “Search” submenu, a more appropriate name. User-friendliness was enhanced between prototype V1 and prototype V2 as well with modification of filter option organization. In prototype V1 there had been a main filtering options category, labeled “Filter By,”

that would enable or disable filter categories. In order to filter by one of the options in a category, the user would need to first expand the “Filter By” list by clicking on the image of a warehouse crate next to the “Filter By” text. Next, the user would check the checkbox next to the category in the “Filter By” list, expand the category, and finally click on the check box next to the particular option. The design team recognized that there would be limited number of categories present, so for user-friendliness the extra clicks required to enable a category were removed and all filter categories were listed on the left side of the page, in prototype V2. A suggestions button was also added so that users could suggest filter categories. Main tab titles and submenu titles were also changed and rearranged between versions.

After prototype V2, semi-structured interviews were performed with two users of the WISE Learning Tool. Both interviewees are faculty members who have used the WISE Learning Tool in sophomore-level courses with large class sizes (more than 100 students). Both interviewees were junior faculty with less than 5 years in a faculty position. These semi-structured interviews included broad questions such as “Walk me through a typical day in which you used WISE?”, “What did you like about WISE?”, and “If you could, what would you change about WISE?” After asking broad questions about WISE, the interviewees were introduced to prototype V2 of the *AIChE Concept Warehouse*, allowed to briefly test it themselves and comment on different aspects of the prototype. Interviewers took notes during the interviews. Some of the issues the interviewees had experienced with the WISE Learning Tool had been addressed in the *AIChE Concept Warehouse* prototype V2. Other issues had not, and

were later incorporated into the *AIChE Concept Warehouse* as a result of interviewee comments. The majority of the question search section was well-liked by the interviewees and, as a result, largely unmodified based on interviewee comments. However, the interviewees provided insightful recommendations for class management in general, a need for a user-friendly interface, and some comments on specific sections in the *AIChE Concept Warehouse*, including and “Manage Tests,” “Statistics,” and “New Question.”

After a limited amount of external testing, some minor changes were incorporated into prototype V3 of the question search section, shown in Figure B.7. This prototype is the working version of the *AIChE Concept Warehouse*, dated December 2011.

The screenshot displays the AIChE Concept Warehouse search interface. At the top, the logo for AIChE Education Division is visible, along with the text 'CONCEPT WAREHOUSE'. The user is logged in as 'Silver' with a 'username' and a 'help' link. A navigation bar contains buttons for HOME, CONCEPTTESTS, CONCEPT INVENTORIES, CLASSES, PROFILE, and RESOURCES. Below this is a secondary navigation bar with links for ABOUT, SEARCH, MANAGE TESTS, STATISTICS, and NEW QUESTION.

The main search area shows a search box containing the word 'pressure', with 'Search', 'Reset', and 'Advanced Search' buttons. Below the search box, it indicates 'Showing 1-20 of 258 results (in 0.029289s)' and 'Results per page: 20'. A sidebar on the left lists categories: Classes, Misconceptions, Topics, My Questions, Unused (self), and Type.

The search results are displayed in a table-like format. The first result is 'ID:1 Modified Throttling Valve', which includes a diagram of a throttling valve with inlet conditions P_1, T_1 and outlet conditions $P_2 < P_1$. The description states: 'An ideal gas flows steadily through the piping system and valve shown below. The inlet pressure and temperature are P_1, T_1 ...'. The contributor is Bill Brooks, the author is Ron Miller, and the class is Energy Balances Thermodynamics. The question type is Multiple Choice, and it has a rating of five stars.

The second result is 'ID:20 Equilibrium Constant Change with Pressure', with the text 'For the gas phase reaction:'.

Figure B.7. Step 5 (in progress): Revision after initial external testing of the question search section of *AIChE Concept Warehouse* prototype V3, December 2011.

Text buttons were converted to icon buttons. To reduce page clutter, the question preview text is no longer underlined; it does still link to a complete question view, but the hyperlink is now indicated by a changing mouse pointer. In addition, checkboxes were added to each question preview so instructors can now select multiple questions at once and click a single add button to add all of the selected questions to their personal question list for use in class.

The *AIChE Concept Warehouse* will continue to be improved both through design team recommendations, beta tester feedback, and community involvement.

B.4 The AIChE Concept Warehouse – Beta Testing

Beta testing of the *AIChE Concept Warehouse* began in November 2011, and is currently underway. Beta testers have been and are still being recruited via the *AIChE Concept Warehouse* website. In addition, beta testers were recruited at the 2011 ASEE Annual Conference and the 2011 AIChE Annual Meeting, described in more detail in the Community Development section of this paper. After being recruited, potential beta testers go through the following process.

First, instructors are asked to fill out the online application form available at the *AIChE Concept Warehouse* website (<http://cw.edudiv.org/>). Currently seven beta testers have signed up to participate. The application form includes questions in five categories: general information about the applicant, teaching philosophy, course information, information about their computer system, and other. General information about the beta testers includes name, contact information, computer expertise and

previous beta testing experience. The teaching philosophy category asks for a brief description of the applicant's general teaching philosophy, information about their experience with concept-based pedagogy and specifically asks about previous experience using clickers. Information gathered in this category enables us to determine and provide the appropriate level of support throughout the beta testing process. The course information collected from beta testers helps prioritize question entry. Information about computer systems is used to ensure compatibility with an instructor's hardware, software, and technological preferences, including operating system, presentation software, and web browser.

The next step in the beta testing process consists of the design team initiating discussions with the potential beta testers. This step allows for further evaluation of the suitability of potential beta testers for pre-release testing. So far, no beta testers have been eliminated based on a lack of suitability. This initial discussion also introduces potential beta testers to the *AICHE Concept Warehouse* prototype and affords development of a preliminary, personalized test plan.

Beta testers then implement their personalized test plan by using the *AICHE Concept Warehouse* as it best fits into their course(s). After use, follow-up interviews are conducted and surveys are given. Interviews and surveys serve two purposes. First, they provide feedback on the user interface and functionality of the tool. This feedback will be used to improve the *AICHE Concept Warehouse*. Second, the interviews and surveys inform the diffusion process. Specific questions in the interviews and surveys are designed to characterize the attributes of this innovation.

B.5 Community Development Activities

Involvement of chemical engineering educators is crucial for the success of the *AICHE Concept Warehouse*. In order to foster community engagement, two types of activities have either started or are planned for the future. These types of activities include professional society related activities (e.g., special sessions, posters, papers, and presentations) and independent department visits. In general, the activities are intended to help faculty interested in incorporating educational methods and tools into their classrooms to encourage students to think more deeply about concepts central to chemical engineering.

At the 2011 American Society for Engineering Education (ASEE) Annual Conference, a special session, titled "Educational Methods and Tools to Encourage Conceptual Learning," was presented, also with an accompanying paper.⁴³ This interactive special session had approximately 70 attendees and lasted 90 minutes. Seven presenters emphasized the need for concept-based pedagogy and described pedagogical and instructional tools that can be used to promote conceptual learning, and identify and repair misconceptions. In addition, a presenter discussed how technology can be used to enable concept-based pedagogy in the ever-changing classroom, highlighting the utility of the *AICHE Concept Warehouse*. Presentations ended with a reflection on the benefits and challenges of integrating concept-based pedagogy into classroom instruction. Activities were included throughout the presentations, which had participants get into pairs, reflect on what was discussed, and

extend their understanding to consider when might be most appropriate to use different pedagogical tools and why some misconceptions are so deeply held. Participants also used an audience response system (Turning Technologies ResponseCard NXT clickers) to answer conceptual questions that were delivered via the *AIChE Concept Warehouse*. After the presentations, the special session concluded with a lively panel discussion of participant questions.

At the 2011 AIChE Annual Meeting, a workshop titled “Tools and Techniques for Conceptual Learning” was facilitated. This 2.5 hour workshop had five presenters and approximately 15 participants. It had presentations similar to the ASEE special session, with an emphasis on the need for concept-based pedagogy and a discussion of pedagogical tools, repair of misconceptions and classroom experiences. However, it ended with participants gaining hands-on experience using the AIChE Concept Warehouse. Participants went through an activity to develop their own concept questions, which led some to appreciate the difficulty in constructing a high quality concept questions. Participants were shown how to add questions to the AIChE Concept Warehouse, create ConcepTests, assign ConcepTests to their classes, and how to view student results. The contrast in attendance between the ASEE special session and the AIChE workshop is reflective of the challenge in attracting mainstream faculty to education reform.

In addition to the activities already completed, professional society related activities are planned for the future. One such activity is a 3-hour, two-part workshop at the ASEE Chemical Engineering Faculty Summer School in July, 2012. The goal of

this workshop is to provide early career faculty members with the education methods and tools they will need to incorporate concept-based pedagogy into their classrooms. Pedagogical content will be similar to that of the ASEE special session and the AIChE workshop. The workshop will have many interactive components, including discussions, clicker and laptop content delivery and response, and scaffolded hands-on use of the *AIChE Concept Warehouse*. During scaffolded use, participants will develop conceptual questions, add them to the *AIChE Concept Warehouse*, and use them to practice a sample lesson using Peer Instruction. At the end of this workshop, participants will walk through the other resources and materials available on the website, and provide feedback regarding the workshop and the *AIChE Concept Warehouse*.

Along with the professional society related activities, we intend to visit departments who are interested in adopting this tool. It is expected that these workshops will incorporate similar material as has been included in previous workshops; however, they will be tailored to the individual needs of the departments. If interested in hosting a department workshop, please contact the corresponding author.

B.6 Future Plans

Future plans include continued application development and testing, and community development activities. Iterative development and testing is focused on finalizing *AIChE Concept Warehouse* sections, incorporating additional concept

questions and concept inventories, performing beta testing, and incorporating feedback from beta testers. Community development activities will primarily be in the form of workshops, introducing the *AICHE Concept Warehouse* to more members of the community. Planned workshops include the 2012 ASEE ChE Summer School and independent department visits.

As researchers improve current tools and develop new concept-based pedagogical tools, improvements and new additions will be integrated into the *AICHE Concept Warehouse*. For example, one improvement expected for concept inventories is the development of diagnostic models to interpret the results of concept inventories. These models can be incorporated into the *AICHE Concept Warehouse* to provide users with a more descriptive explanation of student understanding. New additions might include more scaffolding for the repair of misconceptions and additional approaches to facilitate conceptual learning.

In addition, we intend to study the longitudinal diffusion of this tool in order to inform the design, development, and diffusion of other engineering education best practices. It is expected that by carefully tracking the diffusion process and gathering information regarding decisions to adopt this tool, we will be able to inform others trying to bridge the gap between engineering education researchers and practitioners.

B.7 Summary & Implications for the Future

In this paper, we have reported on the initial development and current status of the *AICHE Concept Warehouse*. This web accessible tool is currently in beta testing. It houses more than 1,200 concept questions, five concept inventories, and provides

pedagogical scaffolding to promote the use of concept-based pedagogy throughout chemical engineering.

While the design of this tool and incorporation of questions provides a strong foundation, participation from the community is critical for success. We encourage the broader community of chemical engineering educators to participate.

If you would like to use it, please visit us at <http://cw.edudiv.org>

B.8 Acknowledgments

The authors gratefully acknowledge support from the National Science Foundation's Course, Curriculum and Laboratory Improvement Program, under the grants NSF 1023099, 1022957, 1022875, 1022785 "Collaborative Research: Integration of Conceptual Learning throughout the Core Chemical Engineering Curriculum." We would also like to thank our beta testers. Finally, we appreciate the other individuals that have contributed concept questions and constructive feedback.

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.

B.9 References

1. *Educating the engineer of 2020: Adapting engineering education to the new century*. The National Academies Press: Washington, DC, 2005.
2. Bransford, *How people learn: Brain, mind, experience, and school*. National Academy Press: Washington, DC, 2000.
3. Hestenes, D.; Wells, M.; Swackhamer, G., Force Concept Inventory. *The Physics Teacher* 1992, 141-158.
4. Elby, A., Another reason that physics students learn by rote. *American Journal of Physics* 1999, S52.
5. Felder, R. M.; Brent, R., Understanding Student Differences. *Journal of Engineering Education* 2005, 57-72.

6. Crouch, C.; Watkins, J.; Fagen, A.; Mazur, E., Peer Instruction: Engaging Students One-on-One, All At Once. *Research-Based Reform of University Physics* 2007.
7. Fagen, A.; Crouch, C.; Mazur, E., Peer Instruction: Results from a Range of Classrooms. *The Physics Teacher* 2002, 206-209.
8. Mazur, E., *Peer Instruction*. Prentice-Hal: Upper Saddle River, NJ, 1997.
9. Hake, R. R., Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 1998, 64-74.
10. Poulis, J.; Massen, C.; Robens, E.; Gilbert, M., Physics lecturing with audience paced feedback. *American Journal of Physics* 1998, 66, 439-441.
11. Deslauriers, L.; Schelew, E.; Wieman, C., Improved Learning in a Large-Enrollment Physics Class. *Science* 2011, 862-864
12. Koretsky, M.; Brooks, B., Student Attitudes in the Transition to an Active-Learning Technology. *Chemical Engineering Education* 2012, 289-297.
13. Falconer, J., Conceptests for a Thermodynamics Course. *Chemical Engineering Education* 2007, 41, 107.
14. Falconer, J., Use of ConcepTests and Instant Feedback in Thermodynamics. *Chemical Engineering Education* 2004, 64-67.
15. Halloun, I., The initial knowledge state of college physics students. *American Journal of Physics* 1985, 1043.
16. Nelson, M. A.; Geist, M. R.; Miller, R. L.; Streveler, R. A.; Olds, B. M. In *How to Create a Concept Inventory: The Thermal and Transport Concept Inventory*, 2007.
17. Steif, P.; Dantlzer, J., A Statics Concept Inventory: Development and Psychometric Analysis. *Journal of Engineering Education* 2005, 363-371.
18. Gray, G.; Costanzo, F.; Evans, D.; Cornwell, P.; Self, B.; Lane, J. In *The Dynamics Concept Inventory Assessment Test: A Progress Report and Some Results*, Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition, American Society for Engineering Education: 2005.
19. Martin, J.; Mitchell, J.; Newell, T. In *Work In Progress: Analysis of Reliability of The Fluid Mechanics Concept Inventory*, ASEE/IEEE Frontiers in Education Conference, Savannah, GA, IEEE: Savannah, GA, 2004.
20. Streveler, R.; Olds, B.; Miller, R.; Nelson, M. In *Using a Delphi Study to Identify the Most Difficult Concepts for Students to Master in Thermal and Transport Science*, Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition, American Society for Engineering Education: 2003.
21. Streveler, R.; Miller, R.; Santiago-Roman, A.; Nelson, M.; Geist, M.; Olds, B., Rigorous Methodology for Concept Inventory Development: Using the 'Assessment Triangle' to Develop and Test the Thermal and Transport Science Concept Inventory (TTCI). *International Journal of Engineering Education* 2011, 968-984.
22. Prince, M.; Vigeant, M.; Nottis, K. In *Development of a Concept Inventory in Heat Transfer*, ASEE Conference Proceedings, Austin, TX, American Society of Engineering Education: Austin, TX, 2009.
23. Vigeant, M. A. S.; Prince, M. J.; Nottis, K. In *The use of inquiry-based activities to repair student misconceptions related to heat, energy, and temperature*, Proceedings from American Association for Engineering Education, Vancouver, BC, Vancouver, BC, 2011.
24. Krause, S.; Decker, J. C.; Griffin, R. In *Using a materials concept inventory to assess conceptual gain in introductory materials engineering courses*, IEEE: 2003; pp T3D-7-11 Vol. 1.
25. Midkiff, K. C.; Litzinger, T. A.; Evans, D. L. In *Development of Engineering Thermodynamics Concept Inventory instruments*, Frontiers in Education Conference, 2001. 31st Annual, 2001; 2001; pp F2A-F23 vol.2.
26. D.C, S., A concept inventory for material and energy balances. *Education for Chemical Engineers* 2010, 5, e1-e12.

27. Borrego, M., Development of engineering education as a rigorous discipline: A study of the publication patterns of four coalitions. *JOURNAL OF ENGINEERING EDUCATION-WASHINGTON-* 2007, 96, 5.
28. Rogers, E. M., *Diffusion of innovations*. 5th ed.; Free Press: New York, 2003; p xxi, 551 p.
29. Krause, S.; Decker, J. C.; Niska, J.; Alford, T.; Griffin, R. In *A Materials Concept Inventory for introductory materials engineering courses*, NASA; 1998: 2003; pp 413-424.
30. Efav, J.; Hampton, S.; Martinez, S.; Smith, S., Miracle or Menace: Teaching and Learning with Laptop Computers in the Classroom. *Educause Quarterly* 2004, 27, 9.
31. Griffin, J., O. Hayden; Walker, T. D. L. In *Faculty and Student Response to a Laptop Computer Requirement for Engineering Freshmen*, American Society for Engineering Education Annual Conference and Exposition, Portland, Oregon, Portland, Oregon, 2005.
32. Cromack, J. In *Technology and learning-centered education: Research-based support for how the tablet PC embodies the Seven Principles of Good Practice in Undergraduate Education*, IEEE: 2008; pp T2A-1-T2A-4.
33. Koile, K.; Singer, D., Development of a tablet-pc-based system to increase instructor-student classroom interactions and student learning. *The Impact of Pen-based Technology on Education: Vignettes, Evaluations, and Future Directions* 2006.
34. Duncan, D., Clickers in the Classroom. *Addison: San Francisco, CA* 2005.
35. Koretsky, M. D.; Brooks, B. J., Student Attitudes in the Transition to an Active-Learning Technology. *Chemical Engineering Education* 2012, 46, 289-297.
36. Turning Technologies, L. ResponseWare.
<http://www.turningtechnologies.com/audienceresponseproducts/responseoptions/responseware/>.
37. i>clicker web>clicker mobile audience response system.
<http://www.iclicker.com/Products/webclicker/>.
38. Inc., Q. Student Response Systems | Interactive Classroom Voting | Qwizdom | Education | Q5 Remote. <http://www.qwizdom.com/education/qvr.php>.
39. eInstruction vClickerSM Mobile Edition | eInstruction.com.
<http://www.einstruction.com/products/student-response-systems/vclicker-mobile-edition>.
40. O'Reilly, T., What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies* 2007, 65, 17-37.
41. Bulger, B.; Greenspan, J.; Wall, D., *MySQL/PHP database applications*. Wiley Pub.: Indianapolis, IN, 2004.
42. Koretsky, M.; Brooks, B. In *A Web-based Interactive Science and Engineering Learning Tool that Promotes Concept-based Instruction*, American Society for Engineering Education Annual Conference and Exposition, Pittsburgh, PA, American Society for Engineering Education: Pittsburgh, PA, 2008.
43. Koretsky, M.; Miller, R. L.; Falconer, J. L.; Prince, M. J.; Vigeant, M. A.; Krause, S. J.; Silverstein, D. L. In *Educational Methods and Tools to Encourage Conceptual Learning in Chemical Engineering*, American Society for Engineering Education Annual Conference and Exposition, Vancouver, B.C., Canada, Vancouver, B.C., Canada, 2011.

*Appendix C – Item Response Theory Analysis of Multiple-choice Concept Questions
using the AIChE Concept Warehouse*

Bill J. Brooks, Sam Mihelic, Adam Z. Higgins, Marita Barth, Paula Weis, and Milo D. Koretsky

Unpublished Description of Data Analysis Using Item Response Theory

C.1 Introduction

The main intent of this appendix is to document our understanding of Item Response Theory (IRT) and its integration into the Concept Warehouse (CW). It is useful to the chemistry department (CH) at OSU specifically and potentially future users of the CW generally. A secondary reason for inclusion is that the method was used in the attempt to characterize the multiple choice questions used in Chapter 3. It did not work for Chapter 3 but the modification, as well as an error I made, may still be useful even though it did not work for that study.

I keep most of the content appropriate to the school of chemical, biological, and environmental engineering (CBEE) study on written explanations (Chapter 3) as it was written up for a conference paper. It provides background for the model and our modifications. I add chemistry context and results where appropriate instead of reworking it from that perspective. I also intentionally keep a dyslexic error I made in part of the normal ogive assumption. The erroneous model fit the CBEE data whereas the corrected model did not. Furthermore, with either model, there appear to be a higher concentration of low guessing parameters for the set of questions where there was a treatment effect. I suspect there may still be something to the error, the method, or a combination thereof even though I do not see it and went another direction.

Originally, while studying the influence that written explanations had on performance on multiple-choice concept questions, I determined that some of the concept questions we were using were unreliable. However, since they all had face or content validity, I was not sure which ones were 'good' or 'bad'. IRT was suggested

as a characterization method and a slight modification to account for our treatment effect looked promising.

The chemistry department at OSU is working on a project to improve their assessment methodologies. Part of that includes comparing and contrasting online and on-campus general chemistry course offerings. To better do that and better assess their students and learning gains more generally, they intend to use more rigorous methods to iteratively refine their collection of concept questions.

C.2 Background & Theoretical Framework

We sought to determine the impact of soliciting a short answer explanation in conjunction with a student's answer choice in technology-based active learning pedagogies. According to constructivist learning theory, new knowledge is actively built upon prior knowledge and experience. It is hypothesized that encouraging students to provide an explanation will result in better scores because it encourages students to actively restructure their knowledge.

Initially a lack of technology provided a significant barrier to the study and practice. Now short answers/explanations are an increasingly available option, even in large classes.^[1,2,3,4] Chemical engineering students also look upon the ability to give written explanations favorably. However, those same students complain about the weight, battery life, and their forgetfulness when laptop computers are used to replace clickers.^[5] Furthermore, in large classes, it can be time consuming to analyze explanations.

Two studies have examined the effect of reflection and discussion by comparing students' answers to multiple choice questions before and after group discussion. ^[6,7] Both studies found cases in which the correct answer was identified as a result of group discussion when all of the individuals were wrong initially. From these studies, researchers infer that group discussion is beneficial. Previously, we have analyzed written explanations and self-reported confidence during a modified version of peer instruction and found a consensus-correctness interaction in correct choice adoption and improvement in quality of explanation. We also found that the display of an answer distribution before discussion influences confidence over answer choice. ^[8]

One complication to determining the impact of soliciting short answer explanations is the need for quality questions. Most clicker questions are not considered in terms of reliability and validity. Even in cases where questions come from concept inventories, the validity and reliability are likely not transferrable. Instrument validity and reliability are for particular populations and within the context of an instrument's complete set of questions. A method to vet a question set was needed. One such method commonly referenced in the literature, Classical Test Theory (CTT), has been used in the development of concept inventories. ^[9] However, CTT assumes chance guessing, among other things.

An extension of CTT that does not make this assumption is Item Response Theory (IRT). ^[10] IRT assumes that the probability a student will get a question correct is a function of both the student's natural 'aptitude' as well as characteristics inherent in the question and attempts to model it. Once an IRT model is fit to the data from a

battery of questions, an ‘aptitude’ parameter for each student and discrimination, difficulty, and guessing parameters for each question are generated. It is hypothesized that questions with enticing distractors, incorrect multiple choice answers that are compelling if the student holds a particular misconception, lower frequency or probability of guessing.

C.3 Methods

Participants & Setting

This study began observing a single term of a required, sophomore-level, undergraduate energy balances course at OSU. The class contained chemical, biological, and environmental engineering (CBEE) students. A total of 138 students from CBEE and 465 students from chemistry (CH) participated in the study. The research was approved by the Institutional Review Board and participants signed informed consent forms.

The CBEE class was divided into two sections for a weekly, one-hour recitation. The lectures and recitations for both sections were taught by the same instructor. The second recitation section was scheduled immediately after the first (which should minimize cross talk) and held in the same room. In recitation, conceptual questions were posed to students and student responses were collected. The questions were timed manually; the instructor judged when it was time to stop polling once approximately 60% of the section had answered. Once the threshold was reached, the instructor typically gave a verbal warning of “30 seconds” before a verbal countdown for the last several seconds.

All students in the CBEE course received full credit for submitting an answer, and received extra-credit for each correct answer submitted. A grading incentive has been shown to affect student discourse in active learning, and it has been suggested that discussion is suppressed in a “high stakes” environment. ^[11,12] The low-stakes approach implemented in this study was intended to both encourage student discussion, while still providing incentive for students to respond correctly.

The CH classes were a combination of online and on-campus general chemistry offerings. Both courses were assessed via BlackBoard™ with short concept question quizzes before and after topic/concept level modules. Compared to CBEE, the CH methods were more self-paced. The CH concept questions also did not have a prompt for a written explanation.

Study Design

To minimize the effect of potential differences between sections in CBEE a crossover study was conducted. A design matrix can be seen in Table C.1. Each section alternated between participation in the control group (no written explanation) and in the treatment group (with written explanation) in two- to three-week intervals during the 10-week term. For the first three weeks of recitation, *Section 1* (S1) was not provided with a prompt or space to explain their answer choices and *Section 2* (S2) was. In the fourth week, the first quiz was administered. After the first quiz, the test parameters were switched. The time spent on each conceptual question with the treatment condition, the written explanation prompt, is longer than is typical ^[13],

largely because students are asked to provide short answer written explanations of why they selected an answer.

Table C.1. Experimental design matrix

Week	Section 1	Section 2	
1-3	without written explanations	with written explanations	← Exam 1
4-6	with written explanations	without written explanations	
7-8	without written explanations	with written explanations	← Exam 2
9-10	with written explanations	without written explanations	← Exam 3
			← Final Exam

The CH classes did not have a treatment effect. In a previous analysis, there was no difference seen between online and on-campus offerings.^[14]

Data Sources, Collection & Analysis

First, CBEE student performance data from their material balance course (i.e., course grades) were used to compare the two sections and determine if they are statistically representative of the same population. The final grades from the previous course in the sequence is viewed as a proxy for student performance. Both a Mann-Whitney W test and a Kolmogorov-Smirnov test do not show a significant difference between the sections.

Second, student responses to conceptual questions were collected and provide the heart of the data for this study. A total of 63 and 55 conceptual questions were asked of S1 and S2, respectively. These questions included the original conceptual

questions and repeated follow-ups used as part of Peer Instruction. Some conceptual questions were end-of-class questions that time permitted in only one of the two sections because the other section spent more time writing explanations or engaging in lengthy student initiated discussions. Repeated follow-up questions were removed from the question set analyzed in this study.


Conceptual questions, including the multiple choice answers, written explanations, and self-reported confidence ratings, were assigned through a relatively new tool available to the chemical engineering community, The AIChE Concept Warehouse. ^[4] The AIChE Concept Warehouse is an NSF-supported, multi-institution, collaborative effort to facilitate the development and use of concept-based pedagogies in the core undergraduate chemical engineering curriculum. Students were asked to first answer individually on their laptops, smart phones, or tablets, without consulting their neighbors or the instructor. For the questions analyzed in this study, students selected a multiple choice answer and reported how confident they were. If they were in the treatment section, they wrote a short answer explanation as well.

The AIChE Concept Warehouse was also used in the collection of the responses for the CH assessments through quiz object integration into BlackBoard™. A total of 37 questions were used in the CH analysis, 20 questions total as a pre-assessments before each unit and 17 after.

Figure C.1 depicts a sample question as it was presented in the treatment condition in CBEE. For the control condition, the identical question was asked except it did not have the text input box or prompt that says “Please explain your answer in

the box below.” To receive credit, all students were required to select an answer choice but writing an explanation and selecting how confident they were with their answer was optional. Similarly formatted questions were used in the CH study but they had neither the written explanations nor confidence follow-ups.

An endothermic reaction takes place in a continuous flow, isothermal reactor at constant pressure.
The enthalpy of the effluent is _____ the inlet.



greater than
 lesser than
 the same as

Please explain your answer in the box below.

Please rate how confident you are with your answer.

substantially unsure	moderately unsure	neutral	moderately confident	substantially confident
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure C.1. Question “Outlet vs. Inlet Enthalpy in Isothermic Endothermic Reaction” as presented to the treatment group

Analysis of student responses was performed through an item analysis, similar to the analysis used by Wang and Bao ^[10] in their IRT analysis of scores on the Force Concept Inventory. Their analysis resulted in a model fit that included 3 parameters to model each question as well as a parameter for each student or student grouping to predict the odds of choosing a correct answer. The question parameters are the

discrimination (a), difficulty (b), and guessing (c). The values for the discrimination and difficulty range from negative two to positive two and the guessing range from zero to one, corresponding to the odds that a student can guess and achieve a correct answer. The student parameter, aptitude (θ), ranges from negative two to positive two and relates to difficulty.

We modified the Wang and Bao model to include an additional aptitude parameter, one for when a given student was a member of the control group and the other when that same student was part of the treatment group. The *correct* model is given by the following equation:

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp(-1.7a_i(\theta_j - b_i))} \quad (\text{C.1})$$

where i is the index of the question and j is the index of the ‘aptitude’ for a particular student.

To highlight some of the influence that the IRT parameters have on the model predictions, two sample questions from CBEE are presented in Figure C.2. The dashed line represents the odds (vertical axis) that a student of a given aptitude (horizontal axis) will get that question correct. It starts at the lower left because its guessing parameter (c) is equal to zero. Since it has discrimination (a) near positive two, the probability of choosing a correct answer increases with higher aptitude. The question modeled by the solid line has discrimination (a) of negative two. Therefore, the odds that a low aptitude student would get it correct is higher and decreases as aptitude increases. The difficulty parameter (b) shows where the transition, the magnitude of

which is determined by a , is centered. The dashed line has a b of about -0.6 and is therefore shifted to the left and the solid line has a b of about 0.8 , shifting it to the right.

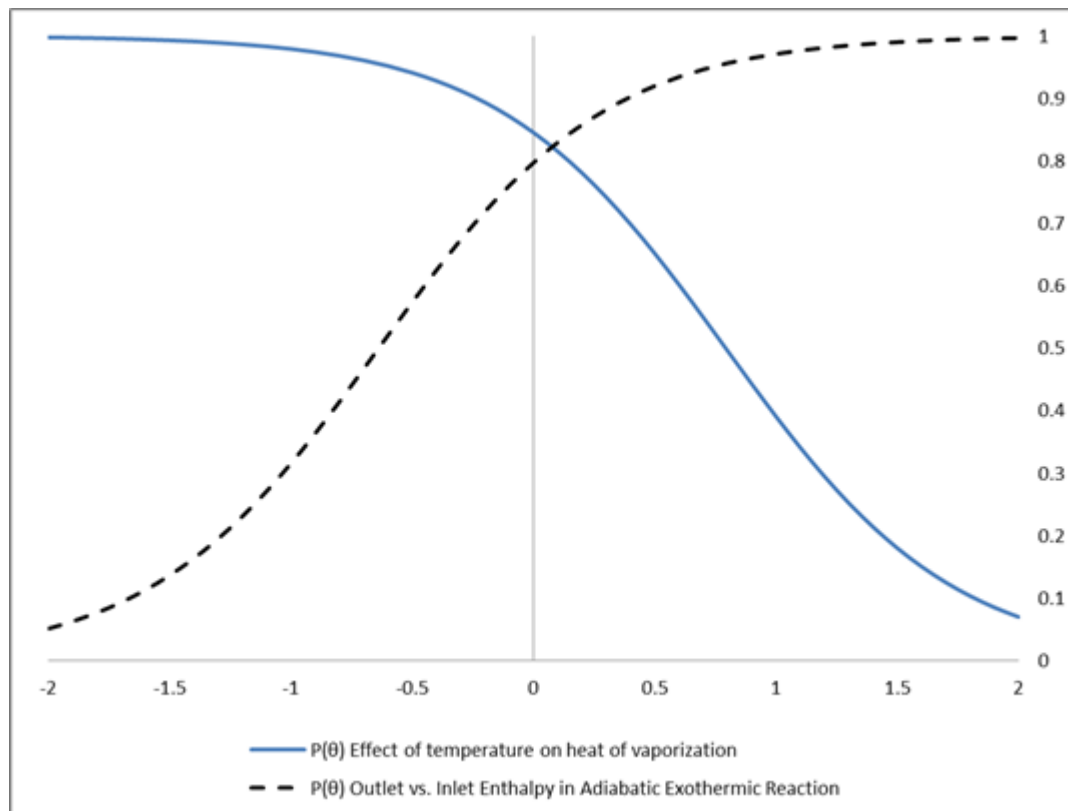


Figure C.2. Sample questions to highlight the effect of model parameters on probability prediction. A) dashed line ($a=2$; $b=-0.6$; $c=0$) B) solid line ($a=-2$; $b=0.8$; $c=0$)

In other words, in the case of the dashed line in Figure C.2, the discrimination parameter is high, indicating that the question is more likely to be answered correctly by high aptitude students (i.e., the students who, on average, perform better on conceptual questions). The difficulty parameter is indicative of the aptitude at which students start choosing the correct answer. The higher the difficulty, the more the

curve is shifted to the right (where only the higher aptitude students choose correctly). The guessing parameter in this case is the limiting probability of answering correctly as the aptitude parameter goes to negative infinity – i.e., the chances that an extremely low performing student would answer correctly. In the case of the solid line, because the discrimination parameter is negative, the difficulty represents the location of the transition between low aptitude students choosing correctly and high aptitude students selecting a wrong answer. This relation no longer has anything to do with the colloquial difficulty. The guessing parameter is the limiting probability of answering correctly as the aptitude parameter goes to positive infinity – i.e., the chances that an extremely high performing student would answer correctly.

Of the 28 questions analyzed for CBEE using something other than a normal ogive assumption (see Equation C.2), 14 had a negative discrimination parameter. While the constant before the discrimination parameter is wrong (i.e., does not assume normal ogive behavior), this result is still informative. One might think that this many questions with a negative discrimination means that a high aptitude student would perform worse than a low aptitude student half of the time. That should be impossible, unless there is no appreciable difference between ‘high’ aptitude students and ‘low’ aptitude students (e.g., if every student in the class got essentially the same percentage of the concept questions correct). While we could provide a plot that shows aptitude vs. overall score that proves aptitude and score are correlated, and that the scores range from around 40% to around 90%, there can still be questions as to whether there really is a difference in performance between high and low ‘aptitude’ students. To

help with further explanation the *erroneous* version of Equation C.1 is reproduced below as Equation C.2:

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp(-1.07a_i(\theta_j - b_i))} \quad (\text{C.2})$$

where i is the index of the question and j is the index of the ‘aptitude’ for a particular student.

To explain how half of the questions can have positive discrimination and the other half negative, while maintaining a correlation between aptitude and score, discussion of the interaction of some of the parameters is important. The closer the discrimination parameter (a) is to zero, the less the difficulty (b) and aptitude (θ) matter. The higher the guessing parameter (c), the more the guessing parameter matters and the less the other parameters matter. There are a number of questions where the model predicts approximately equal chances across the board because of the high guessing parameter and low magnitude discrimination. Even if we neglect the guessing parameter and assume a high magnitude in discrimination, difficulty then becomes important. It is aptitude in the reference frame of the difficulty that really determines whether that aptitude is ‘good’ or ‘bad’ and difficulty is a question by question concern. We can get 40-90% and correlation to aptitude if, for example, the difficulty is positive for some of the negative discrimination questions so that many of the higher ‘aptitude’ students have the same high odds of choosing correctly as the lower ‘aptitude’ students.

Note: In the current version of the model fitting algorithm for the CW, the discrimination parameters are not allowed to go negative and the guessing parameter is not allowed to be greater than 0.5 or 50% guessing chances.

C.4 Results & Discussion

Preliminary results from CH are presented below. They generally show that the correct IRT model fits well for the post-tests but did not fit the pre-tests.

Representative response curves from the pre-tests (Figure C.3) and post-tests (Figure C.4) are presented.

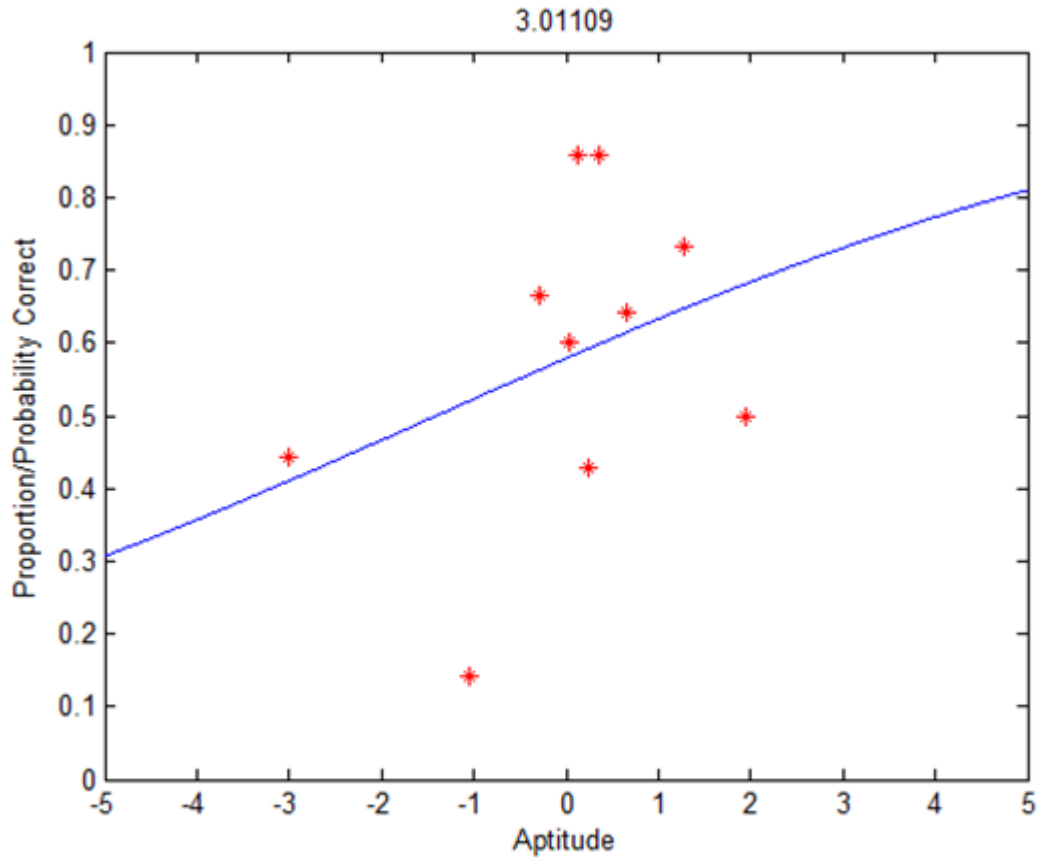


Figure C.3. Representative question (CW question ID: 2213) response curve from the pre-test. The blue line is the IRT model fit, predicting the probability students will select a correct answer as a function of their 'aptitude'. The red stars represent the average score for students in their aptitude group. The number at the top (3.01109) gives the Chi-square value representing the fit (poor) between the model prediction and the data.

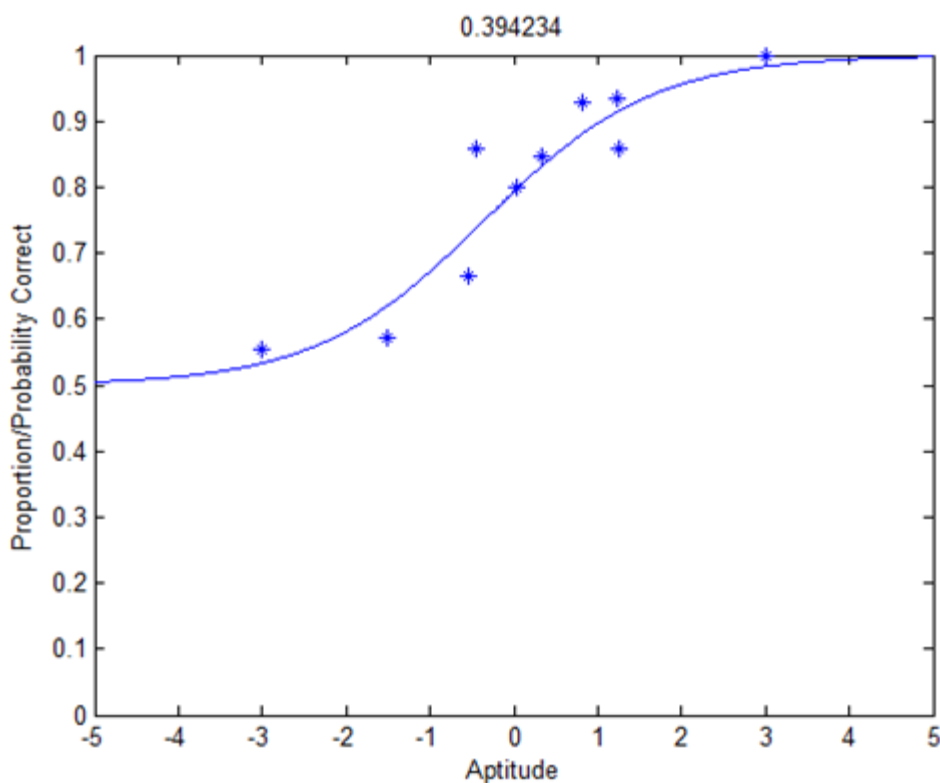


Figure C.4. Representative question (CW question ID: 2210) response curve from the post-test. The blue line is the IRT model fit, predicting the probability students will select a correct answer as a function of their 'aptitude'. The blue stars represent the average score for students in their aptitude group. The number at the top (0.394234) gives the Chi-square value representing the fit (good) between the model prediction and the data.

Figure C.5 presents a response curve for a post-test question (CW question ID: 2230) where IRT did not fit well. The curve looks similar to the results shown in Figure C.3. The question asked, "How many double bonds must be formed to write the correct Lewis Structure for NCl_3 ?" Obviously each of the three chlorine atoms form a single bond with the nitrogen, therefore the answer is zero. This question was brought up in a meeting and it was hypothesized that students might be confusing chlorine with some sort of nitrogen-carbon-iodine molecule because of rendering issues in

BlackBoard™. The question has since been updated to use a molecule that is less susceptible to misinterpretation.

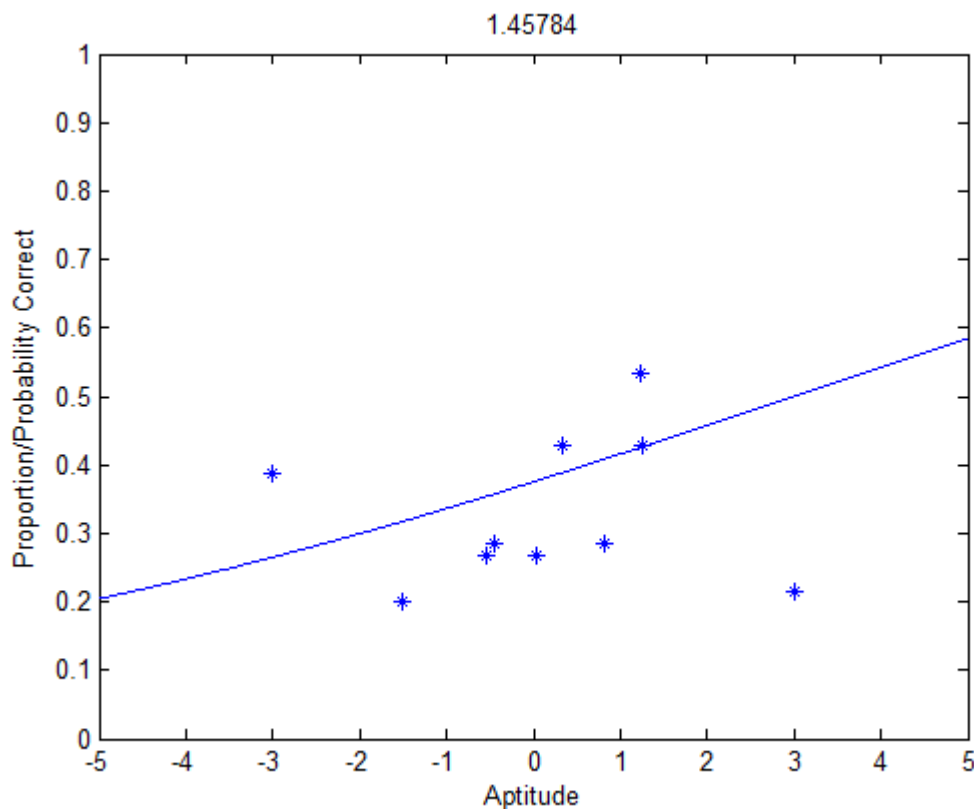


Figure C.5. Flagged question (CW question ID: 2230) response curve from the post-test. The blue line is the IRT model fit, predicting the probability students will select a correct answer as a function of their ‘aptitude’. The blue stars represent the average score for students in their aptitude group. The number at the top (1.45784) gives the Chi-square value representing the fit (bad) between the model prediction and the data.

C.5 Conclusion and Implications for the Future

In the CBEE study, I find it interesting that the erroneous IRT model was able to be fit better than with the corrected model. Furthermore, it is interesting that irrespective of the model or the goodness of fit, there was always a higher number of

questions with low guessing parameter (c_i) in the set of questions where there was a significant written explanation effect on multiple choice answer selection. The association of c_i with the likelihood of guessing and the ‘focused’ nature of private interactions discussed in Chapter 3 leads me to believe there is still potential for using IRT to simultaneously characterize concept questions and the intervention using them.

From the results of the CH study, I propose using IRT on the results from post-tests as a way to measure the quality of chemistry concept questions. Questions where student responses do not fit the model after instruction should be flagged for further scrutiny. I do not recommend that IRT be used to characterize the questions used before instruction because the random results make sense. If none of the students have seen the material yet, it is reasonable to conclude that they are likely responding by randomly guessing. Therefore, grouping them into aptitude groups is also random. Consequently, it makes sense that the response curves would look like those in Figures C.3 or C.5. After instruction, students appear to respond characteristically, unless the question has flaws.

I have three hypotheses for why IRT worked better for CH than for CBEE. First, the effectiveness could be attributed to the much larger number of students responding. IRT usually relies on a much higher number of responses than either of these analyses used. Second, the effectiveness could be attributed to the written explanation intervention creating a significantly different response pattern. A different response pattern could explain why the model fit better with the dyslexic error, see Equations C.1 and C.2. Finally, my choice of using an additional aptitude parameter

could have been a less optimal way of accounting for variability due to the intervention. Perhaps it would have been better to have separate difficulty, discrimination, and guessing parameters to account for the intervention.

C.6 References

- 1 "Top Hat Monocle," Top Hat Monocle, [Online]. Available: <http://www.tophatmonocle.com/>. [Accessed 31 5 2012].
- 2 "LectureTools," LectureTools, [Online]. Available: <http://www.lecturetools.com/>. [Accessed 2 6 2012].
- 3 "Turning Technologies," [Online]. Available: <http://www.turningtechnologies.com/>. [Accessed 2 1 2013].
- 4 B. Brooks and D. Gilbuena, "Preliminary Development of the AIChE Concept Warehouse," in *2012 Annual Conference & Exposition for the American Society for Engineering Education*, San Antonio, TX, 2012.
- 5 M. Koretsky and B. Brooks, "Student Attitudes in the Transition to an Active-Learning Technology," *Chemical Engineering Education*, vol. 46, no. 1, pp. 41-49, 2012.
- 6 M. Smith, W. Wood, W. Adams, C. Wieman, J. Knight, N. Guild and T. Su, "Why Peer Discussion Improves Student Performance on In-Class Concept Questions," *Science*, vol. 323, no. 5910, pp. 122-124, 2009.
- 7 C. Singh, "Impact of peer interaction on conceptual test performance," *American Journal of Physics*, vol. 73, no. 5, p. 446, 2005.
- 8 B. J. Brooks and M. D. Koretsky, "The Influence of Group Discussion on Students' Responses and Confidence during Peer Instruction," *Journal of Chemical Education*, vol. 88, no. 11, p. 1477-1484, 2011.
- 9 M. Nelson, M. Geist, R. Miller, R. Streveler and B. Olds, "How to Create a Concept Inventory: The Thermal and Transport Cocept Inventory," in *Annual Conference of the American Educational Research Association*, Chicago, Illinois, 2007.
- 10 J. Wang and L. Bao, "Analyzing Force Concept Inventory with Item Response Theory," *American Journal of Physics*, vol. 78, no. 10, pp. 1064-1070, 2010.
- 11 S. Willoughby and E. Gustafson, "Technology talks: Clickers and grading incentive in the large lecture hall," *American Journal of Physics*, vol. 77, no. 2, p. 180, 2009.
- 12 M. James, "The effect of grading incentive on student discourse in Peer Instruction," *American Journal of Physics*, vol. 74, no. 8, p. 689, 2006.
- 13 E. Mazur, *Peer Instruction*, Upper Saddle River, NJ: Prentice-Hal, 1997.
- 14 Thackray, K., & Lerner, M. M. (2013). Program Overview and Assessment Methodology for Online Instruction in Chemistry at Oregon State University.