Towards a Model of Educational Transformation:

Documenting the changing educational practices of professors, institutions, and students in introductory physics

by

Chandra Anne Turpen

B.S., University of California, Santa Cruz, 2004

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Physics
2010
This thesis entitled:

Towards a Model of Educational Transformation: Documenting the changing educational practices of professors, institutions, and students in introductory physics

written by Chandra Anne Turpen

has been approved for the Department of Physics.

____________________________________________________________
Noah D. Finkelstein

____________________________________________________________
Katherine K. Perkins

____________________________________________________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

HRC protocol #0205.21 and #1005.43
Abstract

Turpen, Chandra Anne (Ph.D., Physics)
Title: Towards a Model of Educational Transformation: Documenting the changing educational practices of professors, institutions, and students in introductory physics
Thesis directed by Associate Professor Noah D. Finkelstein

While research-based curricula and instructional strategies in introductory physics are becoming more widespread, how these strategies are implemented by educators is less well understood. Understanding classroom implementation of these strategies is further complicated by the fact that they are being used beyond the institutions at which they were developed. This thesis examines how educational innovations are taken up, take root, and transform educational practice. Data is analyzed from two case studies in educational change at the University of Colorado: the use of Peer Instruction (PI) and the use of the Tutorials in Introductory Physics. Our research studies on PI establish that 1) professors’ actual practices involving the use of PI differ strikingly, thus exposing students to different scientific practices, 2) variations in classroom practices create different classroom norms, and 3) students perceive PI classrooms differently in ways that are associated with corresponding PI implementation. Investigations into the use of the Tutorials in Introductory Physics (Tutorials) reveal that focusing purely on individual faculty members’ experiences does not fully capture the complexity of the change processes associated with Tutorials adoption. Although individual faculty members play important roles in the adoption and institutionalization process, other changes occur simultaneously throughout the educational system (i.e. shifts in internal and external funding, as well as expanding partnerships between the physics department, other STEM departments, the School of Education, and other university programs). By examining faculty within the situations that they work, we have found that structural changes in how institutions operate are coupled with changes in how individual faculty members’ teach their courses. These findings call into question the common assumption of dissemination approaches that focus solely on individual faculty members’ adoption and individual use of curricular materials and suggest that approaches to educational change might be more successful by coordinating and addressing multiple levels of the educational system simultaneously.
Acknowledgements

This work has been supported by the National Science Foundation (NSF No.0410744 and NSF No.0448176), the AAPT/AIP/APS (Colorado PhysTEC program), the Center for the Integration of Research, Teaching, and Learning—CITRL (NSF No.0717768), and the University of Colorado at Boulder. First and foremost, I would like to acknowledge my thesis advisor, Noah Finkelstein for his continued support and phenomenal mentoring. My persistence in the PhD program in the physics department and my completion of this thesis project is largely due to my rewarding professional relationship with him. I am thankful for the thoughtful feedback from my thesis committee: Noah Finkelstein, Valerie Otero, Kathy Perkins, Steven Pollock, and Patricia Rankin. I am incredibly grateful for the many CU physics faculty members who were willing to examine their classroom practices and open their classrooms to researchers. Without their willingness to participate in education research and reflect on their teaching practices, this thesis work could not be done.

The continuing thoughtful feedback from the physics education research group at Colorado has been integral to the evolution of this research project and my own thoughts on teaching and learning physics. Steve Pollock has consistently been a critical eye and thoughtful contributor to my work. Many of the ideas expressed in this thesis were co-constructed in informal discussions with Noah Podolefsky, Lauren Kost, and Jessica Watkins. These colleagues were essential in fleshing out early vague thoughts into formal research constructs. I cannot imagine a better community of thinkers with which to work. I also greatly appreciate the innovative and pioneering research conducted by Charles Henderson and Melissa Dancy; their prior work has paved the way for new research questions to be addressed in Physics Education Research and in this thesis.

My development as education researcher in physics has been significantly improved through invaluable collaborations with professors in University of Colorado, School of Education including Margaret Eisenhart, Steven Guberman, Susan Jurow, Ben Kirshner, and Valerie Otero. Their guidance has expanded my knowledge of research designs, educational theories, educational psychology, analytical strategies, and qualitative methodologies.

Lastly, I would like to thank my wonderful family. My husband, Eli, has been my support along the way, from editing papers and chapters to late night debates about philosophy and education. This thesis only exists due to the phenomenal support and encouragement from my family: my husband (Eli Quinn), my mom (Carol Conger), my dad (Tom Turpen), my brother (Jeremy Turpen), Steve and Ann Quinn, Dave Dunn and Renee Neugent. My little ones (Laura Turpen, Colleen Turpen, Kathryn Plant, Carl Plant, and Willow Parvati) have been continued motivation to improve science education.
CHAPTER 1: Introduction

I. Motivation

II. Literature Review: Overview of Prior Research
   A. Variation in Student Learning through the Use of Research-based Curricula
   B. Research into University Educators’ Beliefs about Teaching and Learning
   C. Change Literature: Research and Theory

III. Conceptual Framework
   A. Overview: Socio-cultural Theoretical Approach
   B. Mediation and the Social Origins of Individual Mental Functioning
   C. Cultural Historical Activity Theory (CHAT)
   D. Cultural Psychology and Frames-of-Context Perspective

IV. Overview of Thesis Project
   A. Purpose of Thesis Project and Research Questions
   B. Methodological Approach
   C. Overview of Dissertation

References (Chapter 1)

CHAPTER 2: Variations in Physics Professors’ Implementation of Peer Instruction

I. Introduction

II. Background
   A. Prior Research
   B. Studying an intervention: Peer Instruction

III. Bridging Classroom Norms and Observable Practices

IV. Description of Setting and Methods

V. Data
   A. Clicker question set up stage
   B. Clicker question response stage
   C. Clicker question solution discussion stage

VI. Data Analysis
   A. Summaries across Dimensions of Practice
   B. Case studies illustrating a classroom norm

VII. Conclusions

References (Chapter 2)

CHAPTER 3: The Construction of Different Classroom Norms during Peer Instruction: Students Perceive Differences

I. Introduction

II. Background
   A. Description of Peer Instruction

v
## Chapter 5: Understanding the Students’ Perspectives on the Tutorials

### I. Introduction

### II. Prior Research into Students’ Experiences in Tutorial Learning Environments

- A. Research into Students’ Attitudes and Beliefs about Physics and Learning Physics
- B. Student Anxiety & Motivation within Cognitive Conflict Driven Learning Environments
- C. Students’ Responses to Midterm and End-of-term Evaluations in Tutorial Learning Environments

### III. Description of Setting and Methods

- A. Description of Survey Methodology and Courses Surveyed
- B. Variations in Course Characteristics between Phys1 & Phys2
- C. Statistical Methods for Analysis of Student Survey Data

### IV. Longitudinal Analyses of Students’ Broad Perceptions of Tutorials

- A. Identifying Semester-to-Semester Variations in Students’ Responses to Survey 1 Questions
- B. Examples of Semester-to-Semester Variations on Survey 1 Questions
- C. Examining Semester-to-Semester Variations in Student Background Factors
- D. Examining the Influence of TAs
- E. Examining the Influence of the Lecture Professor on Students' Survey Responses
- F. Examining the Influence of the Back-up Professor across Three Phys2 Semesters
- G. Summarizing the Influence of Instructors on Students’ Perceptions of Tutorials
V. Eliciting Students’ Finer-grained Perceptions of Tutorials with Survey 2
   A. Student-Student Interactions 150
   B. Student-TA Interactions 151
   C. Coordination of Course Components 154
VI. Study Limitations and Directions for Future Study 155
VII. Conclusions 156
References (Chapter 5) 159

CHAPTER 6: Conclusions 162

BIBLIOGRAPHY 170

CHAPTER 2 APPENDICES 189
   Appendix A: Instructor Attributes and Characteristics of Instructor Practices by Stage 189
   Appendix B: Descriptions of Dimensions of Practice 191
   Appendix C: Classroom Observation Rubric 198
   Appendix D: Classroom Observation Rubric User’s Guide 199

CHAPTER 3 APPENDICES 202
   Appendix A: Survey questions used to document students’ perceptions of Peer Instruction 202
   Appendix B: Kruskal-Wallis p-values for set of three semesters of data 206
   Appendix C: Semester-by-Semester Pair-wise Comparisons via Mann-Whitney Test 207

CHAPTER 5 APPENDICES 208
   Appendix A: Tutorial Survey Questions for Survey 1 and Survey 2 208
   Appendix B: Statistical Tests and Associated P-Values for Survey 1 Questions 212
   Appendix C: Summaries of Variations in Student Background Factors 215
Chapter 1: Introduction

I. Motivation

Calls for science education reform echo across the world. In hopes of adding guidance and vision to these efforts, the United States produced two influential publications: the National Science Education Standards by the National Research Council [1] and the Benchmarks for Scientific Literacy by the American Association for the Advancement of Science’s Project 2061 [2]. In these documents, scientists and science educators call for science classrooms to more authentically parallel the practices of scientists emphasizing inquiry-based learning, model construction, conceptual understanding, problem solving, critical thinking, and experimentation. These calls have culminated in recommendations for Science, Technology, Engineering and Mathematics (STEM) education improvement from the National Academy [3] and even Congressional Action (Pub. Law 110-69) [4]. Of the Academies’ recommendations, at least half call for the transformation of STEM in higher education.

Nationally, science departments at the university level are beginning to transform instruction from a traditional view of knowledge reproduction which views students as passive recipients of knowledge (transmissionist, teacher-centered) to a more recently accepted view that students actively construct knowledge based on their prior knowledge and experiences (constructivist, student-centered) [5,6]. Over the last few decades, many research-based curricula have been developed which draw from a student-centered, constructivist approach. Throughout these developments, there has yet to be significant documentation of how university professors adopt and implement new curricula which incorporate a constructivist approach [7].

Science education researchers have begun developing curricula to address these new demands of and goals for science education. In many scientific disciplines—including introductory physics—demand for new teaching materials and methods has also stemmed from the development of assessments which have exposed the lack of conceptual understanding students attain from traditional transmissionist science instruction [8]. Many believe that these conceptual surveys have played a major role in convincing university physics professors that there is a need for changing how physics courses are taught at the university level.

In the context of university-level introductory physics, a number of instructional strategies and associated materials have been produced based on research into teaching and learning (e.g., Peer Instruction [9], Ranking Tasks [10], Interactive Lecture Demonstrations [11], Cooperative Group Problem Solving [12], Workshop Physics [13], Just-in-Time Teaching [14], Tutorials in Introductory Physics [15], and Active Learning Problem Sheets [16]). At the University of Colorado at Boulder (CU), physics professors commonly use Peer Instruction (PI) [9], one of the most widely used instructional strategies [17], and the Tutorials in Introductory Physics (Tutorials) [15], one of the most extensively researched set of physics instructional materials [18].

Designers of curricula have historically studied how students interact with their instructional materials while ignoring complex contextual features of the classroom environment in which curricula may eventually be implemented. For example, significant research has been conducted on how students interact with the University of Washington Tutorials curriculum [18], but we are unaware of any information that has been gathered concerning how professors interact with and implement these materials.

At the beginning of this thesis project, the physics department at the University of Colorado was at a crucial junction concerning department-wide change. In this physics
department, an individual faculty member teaching a course holds a great deal of control over what gets taught in the course and how it gets taught. Some physics professors involved in the physics education research (PER) group used this freedom to reform classroom practices, such as adopting the Tutorials. However, in order for departmental practices more broadly to shift towards more student-centered and constructivist approaches, additional faculty members must be willing to try out these educational innovations and use them effectively.

When this thesis project began in 2005, the use of Peer Instruction by physics professors at CU in the introductory physics courses was beginning to become common practice. However, it was unclear at the time how Peer Instruction was being implemented or used by faculty at CU. At the same time, the 2005-2006 academic year was the first year in which physics professors from outside the physics education research group implemented the Tutorials curriculum. It was unclear how these early adopters of the Tutorials would react to the changes associated with the curriculum. We hypothesized that many departmental faculty members would look to these early adopters from traditional research disciplines to decide whether the curriculum was worth implementing in the future. Research has shown that peer opinion is influential in the adoption of innovations [19]. We became interested in documenting how professors with limited knowledge of the curriculum and limited professional support would implement this curriculum and develop opinions about the curriculum based on these experiences.

In order for research-based science curricula to support wide-scale educational change, the nature of implementing curricula in specific, complex educational contexts must be better understood. Many subtle contextual features contribute to the form that a particular educational innovation takes in the classroom. However, individual faculty members control many of the major decision points concerning the establishment of classroom norms and practices; for this reason they are the main focus of this research project. Understanding educational change requires an awareness of how implementation impacts the form that educational innovations take in the classroom [20]. Therefore, interactions between educators and instructional tools as well as educators and students within a working classroom need to be better understood. This thesis conducts two parallel case studies on faculty adoption of Peer Instruction and the Tutorials in introductory physics courses at the University of Colorado at Boulder (CU). The primary focus in these research studies is on professors’ implementation practices surrounding these two tools. However, attention is paid to factors that shape (and are shaped by) professors’ implementation such as broader institutional structures and variations in student engagement and perceptions.

II. Literature Review: Overview of Prior Research

This thesis draws from research on variations in student performance associated with research-based curricula, studies of university professors’ beliefs about teaching and learning, and models of changing and institutionalizing course transformations in undergraduate STEM. Variations in student learning through the use of research-based curricula are reviewed because many researchers have suggested that variations in student learning outcomes may be due to how the curricula are enacted [8,20,21]. Prior research into university professors’ beliefs about teaching and learning are described in order to summarize useful findings and discuss why the investigations in this thesis do not approach the research problem from a beliefs perspective. Models of change and institutionalization of course transformations in undergraduate STEM are reviewed to motivate the particular conceptual framework adopted in this thesis. Each of these areas of research will impact the studies included in this thesis. The reviews of literature
A. Variation in Student Learning through the Use of Research-based Curricula

Research studies that have examined the adoption and use of pedagogical innovations in undergraduate physics have tended to examine student conceptual learning through pre/post multiple-choice content assessments. A well-cited example of this type of research is the 1998 paper by Hake [8], which showed consistently higher learning gains for courses where the professors reported using “interactive engagement methods” versus traditional lecture format classes. However, a closer look at these data reveals that there is still significant variation in student learning gains within interactive engagement classrooms; the bottom decile of IE courses achieve average normalized learning gains ($<g>$) [22] ranging from about 0.16 to 0.24 while the top decile of IE courses achieve average normalized learning gains ranging from about 0.60 to 0.64 [8]. Average normalized learning gains are used to compare courses, since this measure largely controls for variations in students’ incoming pretest scores. For this reason, Hake hypothesizes that large variations in average student learning gains may be due to “course-to-course variations in the effectiveness of the pedagogy and/or implementation” [Ref. 8, p. 66]. Hake’s study establishes that there is large variation in average student learning gains across IE courses, but leaves unanswered exactly what variations in pedagogical practices may exist between these classrooms that may affect student learning.

Other researchers in physics education have documented variation among student learning gains for various implementations of the same curriculum. In a study of five university physics courses implementing Workshop Physics [13] across different institutions and different instructors, researchers at the University of Maryland found that average normalized learning gains for these courses ranged from 0.39 to 0.57 [23] as measured by the Force Concept Inventory (FCI) [24]. Similarly, in a study of thirty introductory physics courses implementing Peer Instruction across eleven different institutions, researchers at Harvard University found that average normalized learning gains for these courses ranged from 0.19 to 0.55 (estimated from [25]) as measured by the FCI. For comparison’s sake, the learning gains documented with these curricula have been super-imposed onto the Hake plot shown in Figure 1. Such significant variation in student learning suggests possible differences in how these curricula are implemented, and calls for a characterization of faculty practices and a framework for measuring similarities and differences in instructors’ implementation of interactive engagement techniques.
At CU, PI and the Tutorials curricula are used in tandem in the introductory calculus-based physics courses (Phys1 and Phys2). The student learning outcomes that are measured for these courses is therefore a composite effect of both curricula over the entire course. The Tutorials curriculum was first used by PER researchers at CU in 2003 in the Phys1 course. These early implementations of the Tutorials at CU were carefully studied to document the impact the curriculum had on student conceptual learning gains [26]. The student conceptual learning gains documented in the early implementations ($<g>$=0.63) using this curriculum have been impressive compared to national averages [8] (see Figure 1). The complexities of curriculum transfer became of particular interest as the curricula began to be used by faculty with different levels of teaching experience and expertise with the curricula. As of Spring 2009, researchers at CU have documented the impact of the Tutorials [15] in recitations and Peer Instruction in the lectures on student learning over twelve semesters and found a range of average normalized learning gains from 0.32 to 0.64 (see [21] with more recently collected data added) as measured by the Force and Motion Conceptual Evaluation (FMCE) [27]. From these data, we can see that even within a given institution using the same instructional approach, a broad range of student conceptual learning gains is measured.

Throughout this dissertation, more extensive literature reviews of the impacts of Peer Instruction and Tutorials will be discussed as they become relevant to particular research studies. Variations in student learning when clickers (broadly defined) and Peer Instruction are used will be discussed in more detail in Chapter 2. Variations in student learning documented when the Tutorials in Introductory Physics are used will be discussed in more detail in Chapter 4. Overall, research has demonstrated the positive impacts of PER instructional materials for student conceptual understanding. However, there are still many unanswered questions about the range of impacts that have been documented through the use of the same curricula.
B. Research into University Educators’ Beliefs about Teaching and Learning

At this point, one might consider analyzing the variation in student learning outcomes in terms of characteristics of the professors who were teaching the courses. Many research studies have been conducted aimed at better understanding the beliefs and knowledge that educators bring with them to the classroom, assuming that the beliefs and knowledge of the educator largely dictate their actions in the classroom. We briefly describe some findings that have resulted from this research and then discuss why the investigations in this thesis go beyond this limited perspective to include examinations of faculty practices and institutional structures.

1. Research using the Construct of Beliefs to Understand Educators

The construct of beliefs has been commonly used in education research, yet it has remained largely ill-defined. Dewey (1933) described that beliefs are crucial for “it covers all the matters of which we have no sure knowledge and yet which we are sufficiently confident of to act upon and also the matters that we now accept as certainly true, as knowledge, but which nevertheless may be questioned in the future” (Dewey cited in Ref. 28, pg. 313). Dewey argues that educational research into individual’s actions should consider the individual’s beliefs and knowledge, and the intimate relationship between the two, for while beliefs inform action, beliefs may also be called into question.

Despite the historical presence of beliefs in education research, the community of educational researchers has yet to reach consensus about a meaningful and operationalized definition of beliefs [28]. Generally, the construct of beliefs is used to refer to the interpretive lens that people apply to phenomena they experience in order to make sense of experiences. On one side of this interpretive lens we can imagine all possible sensory inputs available from an event and on the other side of the interpretive lens, the actions or responses that a person enacts. Between these two sides lies the interpretive lens or system of beliefs of the individual, which necessarily must be inferred by the researchers [28].

The relationship between knowledge and beliefs has been either not specified or addressed through artificial distinctions (such as knowledge being related to objective fact and beliefs being related to judgments or evaluative cognitive components) [28]. Many of the same theoretical struggles that have been encountered in the context of modeling learners’ conceptual development have been mirrored in theories directed at understanding teacher development. It has often been assumed that teachers have robust theories of teaching, learning, or development (see Table A for examples) that guide their actions in the classroom. This has limited researchers’ abilities to identify inconsistencies and context-sensitivities in teacher thinking.
Table A: Categorization schemes for university academics’ conceptions of or beliefs about teaching proposed by various research teams (taken from Ref. 7).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Categorization Schemes</td>
<td>Dall’ Alba (1991) [See Ref. 29]</td>
<td>Seven Conceptions of Teaching:</td>
<td>Seven Understanding of Teaching:</td>
</tr>
<tr>
<td>1)</td>
<td></td>
<td>1) teaching as presenting information</td>
<td>1) imparting information</td>
</tr>
<tr>
<td>2)</td>
<td></td>
<td>2) teaching as transmitting information</td>
<td>2) transmitting structured knowledge</td>
</tr>
<tr>
<td>3)</td>
<td></td>
<td>3) teaching as illustrating the application of theory to practices</td>
<td>3) providing and facilitating understanding</td>
</tr>
<tr>
<td>4)</td>
<td></td>
<td>4) teaching as developing concepts/principles and their interrelations</td>
<td>4) helping students develop expertise</td>
</tr>
<tr>
<td>5)</td>
<td></td>
<td>5) teaching as building the capacity to be expert</td>
<td>5) preventing misunderstandings</td>
</tr>
<tr>
<td>6)</td>
<td></td>
<td>6) teaching as exploring ways of understanding from particular perspectives</td>
<td>6) negotiating meaning</td>
</tr>
<tr>
<td>7)</td>
<td></td>
<td>7) teaching as bringing about conceptual change</td>
<td>7) encouraging knowledge creation</td>
</tr>
</tbody>
</table>

In Pajares’ review article, “Teachers’ Beliefs and Educational Research: Cleaning up a Messy Construct,” Pajares relies heavily on the early descriptions of beliefs by Rokeach [28]. Rokeach strongly urged researchers who are interested in inferring the beliefs of individuals to investigate the many avenues in which an individual’s beliefs arise [32]. Rokeach claimed that researchers’ “…inference must take into account all the ways that individuals give evidence of belief: belief statements, intentionality to behave in a predisposed manner, and behavior related to the belief in question” (Ref. 28, pg. 315). However, many of the research studies in the context of higher education have tended to focus solely on interview and survey methods of eliciting professors’ stated or self-reported beliefs [7]. These criticisms will be discussed in detail in coming sections.

2. Apprenticeship of Observation

Lortie coined the term, “apprenticeship of observation” to capture the influence of the extended periods of time (for many almost 16 years) that pre-service teachers had spent in classrooms as students [33]. Lortie claimed that these intimate interactions between teachers and students from when pre-service teachers were students are extremely influential in the development of teachers’ ideas about teaching and learning which are well-established by the time pre-service teachers reach the university.

In the university classroom, the cultural and historical influence of past schooling experiences is particularly important since most professors have spent extensive time in academic settings and simultaneously receive little formal training as educators. Most physics faculty members have spent much of their lives participating in communities of practice [34] that revolve around education, either as students, teaching assistants, researchers, or instructors. These experiences occur within evolving social institutions which provide the foundation for
professors to informally or implicitly develop their own ideas about the nature of teaching and learning. As such, faculty members have spent even more years within formal schooling environments than the K-12 pre-service teachers studied by Lortie [33]. It is through a combination of these classroom experiences and the effects of the established norms and practices of the academic discipline that a professor establishes his/her teaching beliefs and practices. The apprenticeship of observation is one way in which individual professors now teaching with the Tutorials were influenced by broader cultural historical activities.

In thinking about the prior experiences of the professor involved in this study, we can think of many different trajectories that may influence professors. The professors in our study may or may not have practical experiences with 1) teaching at university-level courses, 2) teaching large enrollment courses, 3) teaching with particular PER instructional strategies or curricula, or 4) observing other professors using particular PER instructional strategies or curricula. The professors in our study may or may not have principled knowledge about 1) the PER instructional strategies or curricula, 2) the intentions behind the design of the specific curricula, or 3) the results and findings of education research through involvement in departmental brown bags, reading of articles, or informal discussion with local PER researchers. These aspects of professors’ prior experiences will be discussed whenever possible.

3. Research into the Beliefs of Physics Faculty

Physics education researchers have begun to examine the adoption and use of particular pedagogical innovations in PER by examining professors’ beliefs, values, or conceptions about teaching and learning [35,36,37,38,39]. These research studies, drawing from evidence gathered through interviews or surveys, develop frameworks for examining professors’ broad conceptions about teaching and learning [36], self-reported classroom practices [36], and beliefs and values about teaching and learning in the specific context of physics problem solving [38].

The research of Yerushalmi, Henderson, Heller, Heller, and Kuo resulted in a concept map of physics professors’ beliefs and values about the teaching and learning of problem solving [38,39]. They developed a model of the common mental structure around the topic of problem solving from a series of structured interviews with six college faculty from a given institution. They claim that:

“Although many physics instructors present their courses in a manner in which their goal appears to be to transmit information to students, this is not the case. Indeed, their unstated ‘learning theory’ could be characterized as extreme constructivist . . . Instead it is the result of a more complex interplay among their beliefs about student learning, professional values, and perceived external constraints” (Ref. 38, pg. 27).

From the developed concept map, internal tensions or contradictions were identified which could serve as useful starting points in professional development activities aimed at producing conceptual change. Yerushalmi, Henderson, Heller, Heller, and Kuo suggest that the findings of this work could inform curriculum development (i.e., the development of tools that faculty might actually use) and professional development over varying lengths of time focused on particular aspects of professors’ mental structures that it may be possible to influence which do not already align with educational research results.

Henderson and Dancy [35,36,37] found that although professors’ conceptions about teaching and learning physics align with non-traditional or alternative educational ideas, many
professors report that their practices did not always align with these conceptions due to situational constraints. That is, faculty practices are constrained more by structural considerations (such as expectations of content coverage, lack of instructor time, class size, or room layout) than by their beliefs about productive educational practices [37].

These research studies have contributed to a deeper and more nuanced understanding of how professors think about teaching and learning in higher education and has questioned the common assumption that professors’ beliefs about teaching and learning sit in conflict with the principles of research-based educational reforms [36,38]. None of these studies however used observations of the professors’ teaching as evidence for establishing the beliefs of physics faculty. Without information about the professors’ actions and decision-making in concrete classroom situations, it is difficult to understand how faculty navigate their own thoughts about teaching and learning and their perceived external constraints within concrete teaching situations.

4. Critical Reviews of Research into Educators’ Beliefs in Higher Education

In Kane’s review article, she established that most of the research aimed at understanding university academics’ beliefs has relied solely on self-reported information [7]. Drawing from the work of Argyris and Schon [40], Kane describes the importance of combining an understanding of an individual’s espoused theories (stated beliefs) with his or her theories-in-use (those reflected in the actions of the individual) [7]. In documenting the beliefs of professors, most research in higher education has relied solely on interviewing and surveying faculty. Since these studies tend to rely on a single source of evidence, their claims cannot be strengthened or verified through the triangulation of multiple sources of data.

The literature points to the fact that an educator’s beliefs dramatically affect an educator’s actions, and therefore the implementation of a particular curriculum [28]. However, many of the research studies done on teaching and learning have looked exclusively at educators’ espoused beliefs about teaching and learning, while neglecting to study the educators’ theories-in-use. This is particularly true for those studies focusing on higher education [7]. This thesis project emphasizes the theories-in-use of professors (interpretations of observed classroom practices) while secondarily focusing on faculty members’ espoused beliefs (stated beliefs in interviews). The research studies contained in this thesis will focus on triangulating multiple sources of data to strengthen the arguments and claims made, drawing from observations of teaching, collection of course artifacts and documents, and interviews with faculty.

C. Change Literature: Research and Theory
1. Research-development-dissemination-evaluation (RDDE) Model

The “Research-development-dissemination-evaluation” [RDDE] model (Havelock, 1969, cited in Ref. 41, pg. 28) has been found to implicitly guide many change efforts [41]. In this model, one group of people engages in revising theory, conducting research, and developing prototype products. These products are then disseminated to another group of people, “targeted users,” where the use of the products is evaluated. Dissemination in this context typically refers to the “transfer of knowledge within and across settings” (Ref. 42, pg. 28). The knowledge that is being transferred may be educational research results, pedagogical practices, curricula, or instructional materials. Experienced researchers may balk at the use of the word ‘transfer’ in this definition of dissemination and as we will see transfer becomes problematic in the context of change theories as it has in the context of cognitive studies of learning.
Specific examples of the RDDE model can be seen in the NSF sponsored science curriculum development projects of the 1950’s & 1960’s, such as the Physical Science Study Committee (PSSC) or Biological Sciences Curriculum Study (BSCS) [43]. Driven by the post-Sputnik goals of improving the quality of science education and increasing the number of scientists, scientists engaged in developing science curricula for the K-12 public education system. These scientists therefore took responsibility for the research and development of new curricular materials. As Hutchinson and Huberman describe however, “the dissemination strategies used were based on the simple conviction that if the materials were good they would be used, and that their use would result (automatically) in further diffusion” (Ref. 42, pg. 29). The NSF and the U.S. Office of Education at the time attempted to encourage dissemination through the use of teacher institutes and demonstration projects, however, “it appeared that neither agency’s strategy resulted in robust and durable change at the classroom level” (Ref. 42, pg. 29).

Many criticisms of the RDDE model have been put forth. In many ways the RDDE model presents many of the same problems found in the transmissionist model of learning. As Hubbard, Mehan, and Stein suggest the RDDE model depicts a “causal arrow of change traveling in one direction—from active thoughtful designers who make plans for passive and compliant implementers who carry out the plans; that is complete predetermined goals and objectives of the design team” (Ref. 44, pg. 13). Embedded in this model is the implicit assumption that the educational researchers and/or curriculum designers are best suited to develop the solutions to educational problems which will then be used by educators without significant modifications. This model “casts the flow of knowledge as a one-way process and does not take into account the motivations, contexts, and realities of the intended recipients” (Ref. 42, pg. 27-28). In the linear RDDE model, the only relationship between potential users and the developers is a one-way interaction with knowledge or products transferring from the developers to the potential users such that the context within which the potential users work does not inform the research and development process.

Unfortunately, many aspects of the RDDE model can be seen as recently as 2005 in the change model presented in the NSF’s CCLI Program Solicitation [45]. The NSF’s change model articulates five project components: 1) Conducting Research on Undergraduate STEM Teaching and Learning, 2) Creating Learning Materials and Teaching Strategies, 3) Developing Faculty Expertise, 4) Implementing Educational Innovations and 5) Assessing Learning and Evaluating Innovations [45]. Although some nuance has been added to the RDDE model by representing it as a cyclic process (as shown in Figure 2), the model proposed by the NSF retains many of the failings associated with the previous RDDE model. For example, the model still depicts program components as being linked through a one-way interaction and project components are only shown to interact with neighboring project components. This model constrains the relationships between project components such that two-way feedback and co-evolution of these components is neither recognized nor permitted.
Later in the 2005 CCLI program solicitation, the NSF appears to recognize some of these concerns: “In addition to affecting the subsequent development of an innovation within one component, evaluation and assessment results from one component will influence the design of other components” [45]. However, the relationships between assessment/evaluation and other program components are not clarified in the document nor indicated in the diagram. Additionally, the NSF intended to fund between 3 and 4 times as many Phase I projects (which typically address only one program component) as Phase II/III projects (which typically address multiple program components) [45]. Although the monetary awards increase in each progressive phase, the vast majority of CCLI grants awarded are supporting the development of innovations within only a single program component. Furthermore, in the descriptions of Phase I projects, it is implied that these may lead to Phase II or III projects. Thus, implicit in these guidelines is the fact that early stages of research and development are still occurring largely isolated from the prospective users of these innovations.

The limitations of the RDDE model have influenced this dissertation project in two substantive ways: 1) in our analyses of local change processes, we will choose a theoretical framework that will avoid the common criticisms of the RDDE model and will approach researching change processes from a systemic view, and 2) the specific research-based instructional strategies studied here were developed through the support from the NSF’s CCLI program such that many of the implicit assumptions of the RDDE model may be embedded in these tools (e.g. Tutorials in Introductory Physics—NSF #0088840 and #0618185; Peer Instruction—NSF #9980802). By reflecting on the CCLI funding process, we can see that the history of development of these tools followed much of the traditional RDDE process in which the research and development occurred prior to and in isolation from the broader community of potential users of these innovations. A common approach to supporting change (and therefore use of these innovations) is to convince individual faculty that what they are doing in the classroom does not work, attempt to change their beliefs about teaching and learning, and send them away with new materials to use at their home institutions. This process neglects the
relationships that faculty have with other educational materials, other colleagues, and educational settings at their home institutions.

2. Faculty Adaptation of Curricular Materials

Through interviews with five physics faculty, purposefully chosen to represent “likely users of education research,” Henderson and Dancy found that the expectations of physics faculty differed significantly from those of educational researchers and curriculum developers [35]. The theoretical framework proposed by Henderson and Dancy to make sense of these divergent expectations frames change as involving two types of participants: 1) the instructors being interested in or being asked to change and 2) the ‘change agents’ which include curriculum designers and professional development providers. This framing continues to subscribe largely to a RDDE or dissemination model. Drawing from the work of Rogers [19], Henderson and Dancy describe four different change processes across which the relationships between the instructors and change agents vary (see Figure 3).

![Adoption-Invention continuum](Proposed by Henderson and Dancy in Ref. 35).

At the adoption end of the continuum, the change agent develops all of the materials and instructional strategies and the instructor uses them as they are designed. At the invention end of the continuum, the instructor develops all of the materials and instructional strategies with little influence from the change agent. Of the 20 identifiable instances of change identified in the interviews, 70% of them were categorized as reinvention or invention [35]. They found faculty modifications to curricula were frequent and were made largely independent from the assistance of change agents. Henderson and Dancy conclude their article with five recommendations for the research community: 1) Provide easily modifiable materials, 2) Disseminate and research findings in addition to curriculum, 3) Explicitly research the conditions for transfer, 4) View faculty as partners, 5) Acknowledge that change is difficult and support, rather than blame instructors [35]. In order to integrate many of these recommendations seriously into the common practices of researchers and curriculum designers, the community of educational researchers and curriculum developers would need to adopt a model of change that differed significantly from an RDDE model.

3. Literature Review of Change Strategies—Looking beyond RDDE

Henderson, Beach, Finkelstein, and Larson conducted an extensive interdisciplinary literature review of articles which identify change strategies for impacting STEM undergraduate instruction and present evidence of success (or lack of success) regarding different change strategies [46]. The review process began with a preliminary assumption that many different (and fairly isolated) communities engage in efforts to reform STEM undergraduate instruction
and research these processes: 1) Disciplinary-based STEM Education Researchers (SER), 2) Faculty Development Researchers (FDR), and 3) Higher Education Researchers (HER). Initial analyses from this project demonstrated that three recent reviews of change processes from these three traditions (SER—Seymour, 2002 [47]; FDR—Emerson & Mosteller, 2000 [48]; HER—Kezar, 2001 [49]) did not contain a single common reference or citation [46].

Upon further analysis, this interdisciplinary team was able to identify two guiding questions that were useful for categorizing change strategies present in the literature: 1) “What is the primary aspect of the system that the change strategy seeks to directly impact (individuals or environment and structures)? 2) “To what extent is the intended outcome for the individual or environment known in advance (prescribed or emergent)? Using these two dimensions, the authors proceeded to define four change strategies [46,50]:

- Disseminating: Curriculum and Pedagogy (Individual/Prescribed),
- Developing: Reflective Teachers (Individual/Emergent),
- Developing: Policy (Environment & Structures/Prescribed)
- Developing: Shared Vision (Environment & Structures/Emergent)

These four change strategies were found to well-represent the vast majority of articles contained in their literature review. These categories were also useful in demonstrating that not only were the SER, FDR, and HER communities not referencing each others’ research, but these communities also tended to adopt particular change strategies. The SER community was found to most likely adopt the disseminating curriculum and pedagogy change strategy while the FDR community was most likely to adopt the Developing Reflective Teachers strategy. The SER community was involved in authoring fewer than 4% of the articles in either of the Environment & Structures categories. The authors proceed to describe the strengths and weaknesses of these four change strategies as well as suggest that this framework will be useful to researchers in all communities as they reflect on the current work and the direction of future work. They claim that this framework “can elucidate areas of focus that all too often remain as tacit, immutable components of our education system” (Ref. 51, pg. 8). This research tells us as SER researchers that we are particularly inclined to focus our change efforts on individuals at the expense of focusing attention on the surrounding environment and structures.

Based on the preliminary synthesis work done by Henderson, Beach, Finkelstein, and Larson, four conclusions were drawn: 1) of the research articles examined, more researchers focus on the scale of individuals than on larger units of analysis such as departments, institutions, etc.; 2) different disciplines tend to draw from different change strategies; 3) few research articles connected their work to broader change literature; 4) due to the limited explicit attention to change strategies and their associated justification, there is little strong evidence presented for the success or failure of particular change strategies [46,50,51].

The results of this synthesis of research on change have framed this thesis project in multiple ways. Particular attention has been paid to drawing from broader change literature. This thesis examines change processes through two carefully designed case studies directed at understanding the broad implementation of two research-based curricula. A theoretical framework for these investigations was carefully chosen to work across levels of the educational system, rather than focusing on the individual in isolation. We focus on documenting the daily routines that physics professors develop around these new tools with attention paid to concurrent changes occurring at broader levels of the educational system (i.e. the department or institution). The particular theoretical lens adopted for the research studies in this thesis addresses many of the concerns raised by the literature review of Henderson, Beach, Finkelstein, and Larson by
studying faculty members in their local context, paying attention to environment/structures surrounding faculty, and working to explicate the associated relationships faculty members have with various tools, artifacts, and people at their local institutions.

III. Conceptual Framework
A. Overview: Socio-cultural Theoretical Approach

This dissertation will draw broadly from a socio-cultural perspective which “sees context and cultural practices as the fundamental units within which cognition has to be analyzed. Human mental functioning is seen as emerging from and located within social practices” (Ref. 52, pg. 38). Drawing from the work of Cole [56] and Vygotsky [53], learning is viewed as the internalization of social norms and practices as a result of an individual’s action mediated by the use of cultural tools in social activity. In some chapters, more recent interpretations of this perspective will be used such as cultural historical activity theory (CHAT) which draws attention to specific relationships between individuals and communities in addition to interactions between individuals and tools [54,55]. In order to study the process by which educational innovations are taken-up, take root, and come to transform educational practice, we examine changes across multiple scales of the educational system. I draw on work from Cole [56], Finkelstein [57], and Hubbard et al. [44] to distinguish the scales of this system: university institution, physics department, course level, class activities and individuals engaged in tasks. Approaching this thesis project from a perspective of mediated action is a particularly important contribution to the PER field as it draws from a theoretical perspective broader than the cognitivist perspective which predominates the field [58]. This theoretical approach allows for investigation into the daily work of educators in the classroom context as they negotiate new routines, rather than focus on historical divisions between the individual and the environment.

B. Mediation and the Social Origins of Individual Mental Functioning

The socio-cultural perspective developed largely based on the work of Vygotsky. Three ideas fundamental to a socio-cultural approach derive from Vygotsky’s work: 1) tools (signs) mediate thought, 2) the development of all higher psychological processes are social in origin, and 3) the appropriate unit of human cognition is human action (which may be external or internal, may be carried out by a group, both small and large, or by individuals) [59]. Each of these ideas will be discussed in the following paragraphs using concrete examples.

In describing mediation, Vygotsky claimed that many interactions between individuals and the world around them do not occur through direct means but rather through indirect means. “Human thinking, [Vygotsky] said, is mediated by signs, just as human behavior in the world is mediated by tools” (Ref. 60, pg. 269). Indirect means of interacting with the world may include the use of physical tools that have been invented to serve particular functions in historically evolved human activity. For example, the act of tying a string to one’s finger to remind oneself of something is an example a mediator. Through human action upon physical environment (i.e. tying a string to one’s finger), objects (string and finger) are modified to better direct a person’s future thought and action (use of string to regulate or control memory processes). This relationship of a direct path and a mediated path are illustrated below in Figure 4.
Vygotsky focused on the developmental trajectories of human thought. In his general genetic law of cultural development, he claimed that, “every function in the [person’s] cultural development appears twice: first, on the social level, and later, on the individual level; first between people (interpsychological), and then inside people (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relations between human individuals” (Ref. 53, pg. 57). For example, a mother may tie a string to her child’s finger to remind him/her to pick up him/her younger sibling on his/her way home. Gradually, the need for this external reminder may no longer be needed. The child may internalize this external symbol, such that the child has learned to control his/her memory without relying on external aids.

Vygotsky also emphasized that “…mental and behavioral processes are aspects of a unitary system of human action-in-the-world. Planning, remembering, reasoning, and problem solving—cognitive acts in general—are not undertaken as isolated acts or ends in themselves; they are not functions of a mind cut off from a body. Rather they are indissolubly tied to behavioral actions, with both embedded in larger systems of activity whose purposes they serve” (Ref. 60, pg. 267-268). Socio-cultural research is not preoccupied with behaviors, but with action, its intention-based counterpart. As in the previously mentioned example, the act of tying a string to one’s finger is an attempt to modify the external environment in order to influence an individual’s future thoughts and actions. In the context of physics learning, a physics major may learn to “play a movie” or simulation of a physical event in his/her head prior to solving a physics problem. This cognitive resource that the physics major has developed may have its early roots in external forms such as drawing pictures of physical phenomena or experiences using computer simulations. The prior use of these external presentations in the process of solving physics problems (about questions that have evolved as important to physicists) are intimately tied to the future use of the cognitive resource towards particular ends. In this way, socio-cultural research focuses on social practices and changes in participation.
C. Cultural Historical Activity Theory (CHAT)

Building on the work of Vygotsky, cultural historical activity theory (CHAT) worked to emphasize how the thought and action of individuals developed through simultaneous interactions with tools and a community of other people during joint activity [55]. CHAT frames people as continually shaping and being shaped by social context. Associated with these theoretical developments, a heuristic representation was developed where the role of mediation is expanded beyond an emphasis on tools, to include other actors, division of labor, and rules of interaction simultaneously (see Figure 5). CHAT researchers take activity systems as the fundamental unit of study.

![Activity Triangle](Taken from Ref. 56 and Ref. 57).

CHAT researchers and theorists tend to focus on identifying contradictions or tensions that exist in an activity system, since these aspects of the system are theorized to be likely to change. Therefore, this theoretical perspective, “is of immense interest to us because it has shown to be fruitful for both analyzing data recorded in real classrooms and designing change when trouble and contradictions become evident in these cultural settings” [Ref. 55, pg. 188].
This theoretical perspective assists us in explicating particular relationships to focus on during observations of ongoing collective classroom activity. As we will see in chapters 2 and 3, it is useful to describe the roles and associated responsibilities of participants in the classroom in order to understand the particular scientific practices that students are (or are not) given opportunities to engage in.

D. Cultural Psychology and Frames-of-Context Perspective

In Cole’s book Cultural Psychology, he describes the work of Vygotsky and his students as follows: “The central thesis of the Russian cultural-historical school is that the structure and development of human psychological processes emerge through culturally mediated, historically developing, practical activity” (Ref. 56, pg. 108). Cole builds on the work of Vygotsky and develops a rich description of culture and the role of context in activity which will be revisited in Chapter 4.

Cole argues against the common use of context as a backdrop or an empty container waiting to be filled by individuals and their activities. He claims that particularly for studying learning, context must be taken as a more complex and substantial element of the learning process. To develop this model of context, Cole draws on “the Latin root of the term, contextere, which means to weave together” (Ref. 56, pg. 135). He develops this model further and describes context and task as mutually constitutive, suggesting that context and task interact and define each other. Context can be thought of as “the interconnected collection of factors that frame the structure and meaning of human actions” [Ref. 61, pg. 198]. In the definitions provided, context is multi-faceted consisting of the task being worked on, artifacts and tools used during this work, the personal histories of the participants, the social and physical setting. Cole proceeds to discuss how multiple layers of context influence (and are influenced by) a given task and how interactions across these contextual levels define activities that take place. The idea of multiple nested co-constituting levels of context will be particular helpful in understanding the Tutorials. Cole argues the most significant influences on a given frame are the frames immediately above and below.

Finkelstein first applied Cole’s conception of multiple layers of context to understand student learning in a physics course about teaching and learning physics [57]. This course integrated three course components: physics content (electricity and magnetism), theories of teaching and learning and practical teaching experiences creating new kinds of learning experiences for students. In another study, Finkelstein and Pollock also showed the frames-of-context perspective to be useful in understanding the replication of the Tutorials at CU [26]. Replication studies have shown that subtle contextual features of communities and educational environments can have dramatic impacts on the effectiveness of educational innovations [26]. A frames-of-context model helps us to broaden our perspective on what might influence Tutorial implementation.

This structure allows us to focus on appropriate frames at the appropriate times. This thesis focuses primarily on three levels: the institutional/departmental level, the course level, and the task level associated with concrete classroom actions. For example in the context of the Tutorial studies, we will focus primarily on the institutional/departmental level and the course level. For the Peer Instruction studies, we will focus primarily on the course level and the task level (classrooms actions that constitute a PI episode).
IV. Overview of Thesis Project

A. Purpose of Thesis Project and Research Questions

This dissertation responds to calls for systematic study of educational change across multiple scales. We describe two case studies based on the adoption and use of two different research-based curricular tools for teaching introductory quantitative physics—*Peer Instruction* and *Tutorials in Introductory Physics*. The goal of this dissertation is to examine how educational tools are taken up, take root, and come to transform educational practice in physics at institutions different from those that developed them. In this thesis, I seek to understand 1) how local communities of educators and students negotiate new daily routines through the use of new educational tools and 2) how organizational structures shift to support or impede these efforts.

In the last eight years, the CU Physics Department has dedicated significant energy and resources to the adoption of existing research-based curricular tools and has achieved significant success [21,26]. Through implementing and modeling student-centered instruction associated with these reforms, the expectation of program leaders is that the quality and quantity of STEM students, researchers, and future K-12 STEM educators will increase and that STEM departments will become critical actors in the recruitment and preparation of future K-12 educators [33,62]. The CU physics department’s efforts—and associated documentation of those efforts—make it uniquely situated to serve as a case study in institutional change to meet the national calls for STEM educational transformation. In order for national STEM education reforms to achieve large scale impacts, many detailed studies of curriculum implementation must be done across many contexts to better understand their efficacy and change processes. The case-studies presented here are an attempt to address these large-scale questions within a particular local context.

This study investigates the use of two specific curricular innovations, *Peer Instruction* [9] and *Tutorials in Introductory Physics* [15], directed at transforming different aspects of introductory physics courses. One of the innovations, *Peer Instruction* (PI), was intended by the inventor to modify existing large lecture courses by increasing student engagement through conceptual content questions that are answered through peer discussion. The second innovation, *Tutorials in Introductory Physics* (Tutorials), was targeted at reforming the recitation sections that commonly accompany large lecture introductory courses. The adoption of this curriculum required a restructuring of the recitations: shifting content emphasis from computation to conceptual understanding; shifting student activity from watching, listening, and transcribing to actively discussing, reasoning, and problem solving; shifting student interactions from individual work to group work; and shifting the role of the educator from a source of answers to a source of guiding and focusing questions. In this way, the two innovations sought to change different aspects of large enrollment introductory physics courses without challenging the broader structural characteristics of the educational system. Despite the availability of data documenting student content mastery through the use of these curricula, barriers to use of these tools nationally are not well understood.

Early models of curricular diffusion assumed that the innovations would be disseminated and used as they were designed [20]. This model of dissemination has been challenged and significant work has been done to demonstrate that new users inherently adapt and even reinvent
these tools to fit their specific needs [35]. By examining professors’ enacted use of these curricula, processes of adaptation or reinvention can be better understood.

In this thesis, we focus on the enacted use of two PER curricula. The nature of these two reforms calls for slightly different research focuses in each case. Peer Instruction activities are embedded within a given lecture class period while lecture is only one of many activities that make up the course. Although the PI activities will be our primary focus, we will examine neighboring contextual levels: examining how PI is situated in the lecture class and the course as well as investigating the task level focusing on particular student and professor actions that occur within a given PI episode. The implementation of the Tutorials however is more complex. There are at least two critical activities that directly influence Tutorial implementation: 1) The recitation sections where the Tutorials are worked on by students with the support of educators and 2) The Tutorial training meetings where faculty prepare the Tutorials educators to teach particular Tutorial activities. Since the use of Tutorials spans multiple course activities and requires simultaneous structural changes to the educational system within which they are embedded, a broader scale of focus is used.

We focus on the implementation of curricula because this is where we can gather evidence about instructors’ theories-in-use and about the process of adapting curricula. In our investigations we focus on documenting physics faculty’s actual classroom practices rather than relying solely on self-reported descriptions of their practices. The evolving adoption process of these two curricula allow for slightly different research questions to be asked. At the time this thesis began many faculty already had prior experience using PI. Therefore in the context of PI, we examine how a diverse set of faculty members implement PI and the possible impacts on students due to varying implementation. In the context of the Tutorials, research studies were able to capture the early decision making process associated with adopting the Tutorials. We can therefore document in-situ, the justifications given by faculty for initially deciding to try using the Tutorials. We are also able to examine associated changes in institutional and departmental structures which support or impede the use of Tutorials. Although implementation of Tutorials is complex, we begin by looking for differences in students’ perceptions of Tutorials to see if possible implementation differences may impact students’ reported experiences in these courses.

Research Questions

1. Do the daily classroom routines of a traditional lecture vary significantly from the classroom routines surrounding the use of Peer Instruction and the Tutorials by professors at CU? Are the students engaging in doing physics in new ways?
2. Are there significant differences amongst CU physics professors’ implementations of Peer Instruction? What are the impacts of these differences on how students are observed to be engaged in doing physics?
3. How do students perceive courses using Peer Instruction and the Tutorials? Do students perceive the use of the educational tool and the classroom in different ways depending on the tool’s use in classroom activities?
4. How has the CU-Boulder physics department created a sustained use of the Tutorials? What changes occurred throughout the educational system that allowed for the sustained use of the Tutorials?
5. What justifications do physics professors give for initially using the Tutorials?
B. Methodological Approach

The overarching methodological framework for this thesis project draws from the ethnographic traditions of anthropology, education, sociology, and more recently PER. Through extensive engagement of the researcher within a community, ethnographic research aims to build models based on both insider and outsider perspectives. The researcher’s goal is to make explicit models of the implicit meaning systems of the community. These models are influenced by the insiders’ interpretations of events, but due to the researcher’s ability to both bring an outside perspective and a reflective approach to the system, the researcher is in a unique position to identify overarching patterns that can give sense to the patterns of the community. In this way, the researcher’s interpretation of the values, unspoken in the practices of the community, is important and can help to make explicit cultural elements that are largely implicit to the participants [63]. Following from these traditions a combination of qualitative and quantitative methods are used. Data are gathered through classroom observations, audio-recordings of classroom events, collection of documents and course artifacts, semi-structured interviews with physics professors, and student survey responses. More detailed discussions of the methods used for particular studies are described in following chapters.

C. Overview of Dissertation

I briefly describe the structure of this thesis and the particular research studies that will be examined therein. Throughout these coming chapters, conceptual framing and additional research results from the literature are discussed as they pertain to specific research studies.

In Chapter 2 (which addresses research questions 1 and 2), the classroom practices surrounding the implementation of Peer Instruction (PI) are examined across six professors’ introductory physics classrooms. While educational reforms in introductory physics are becoming more widespread, how these reforms are implemented is less well understood. Through observations of classroom practice, we find that professors’ classroom practices during Peer Instruction differ strikingly. A framework is described for capturing instructional choices and resulting variations in enacted instructional practices during Peer Instruction. Based on these observations, we find that there are a variety of scientific practices that are supported and modeled during the use of Peer Instruction. In all of the classrooms studied, students were found to be trying out and applying new physical concepts and discussing physics with their peers. However, there were large discrepancies in students’ opportunities to engage in formulating and asking questions, evaluating the correctness and completeness of problem solutions, interacting with physicists, identifying themselves as sources of solutions, explanations, or answers, and communicating scientific ideas in a public arena. Detailed case studies of two professors demonstrate how these variations in classroom practices may, in aggregate, create different classroom norms, such as the relative emphasis on student sense-making vs. answer-making during Peer Instruction.

In Chapter 3 (which addresses research question 3), variations in instructors’ implementation practices during Peer Instruction (PI) are summarized and linked to different norms of classroom interactions. We describe three classroom norms integral to Peer Instruction: 1) faculty-student collaboration, 2) student-student collaboration, and 3) the emphasis placed on sense-making vs. on answer-making. Based on interpretations by an observing researcher, three different PI classrooms are placed along a continuum representing a set of possible norms. These interpretations are then checked against students’ perceptions of
these environments from surveys collected at the end of the term. We find significant correspondence between the researchers’ interpretations and students’ perceptions of Peer Instruction in these environments. We find that variation in faculty practices can set up what students perceive as discernibly different norms. Concrete classroom practices are described that appear to encourage or discourage these norms.

In Chapter 4 (which addresses research questions 1, 4 and 5), three coupled research studies on Tutorial implementation are presented: 1) Systematic examination of Tutorial adoption and institutionalization at CU, 2) An interview study examining individual faculty members’ early Tutorial adoption decisions, and 3) Modeling Tutorial implementation from a frames-of-context perspective. Drawing on an in-depth case study approach, these coupled studies contribute to a more nuanced model of educational change processes. Results from the first study draw attention to the fact that purely describing the changes that individual faculty members’ experience would not fully capture the complexity of this change process. Individual faculty members played important roles in the adoption and institutionalization process, but other simultaneous changes throughout the educational system cannot be understated. These simultaneous changes throughout the education system included shifts in external funding, shifts in the dedication of departmental resources, shifts in support from university administrators, expanding partnerships across STEM departments, expanding partnerships between the School of Education and School of Arts and Sciences, and expanding partnerships between the physics department and other university programs across campus. By broadening our lens and examining faculty within the situations that they work, we see that structural changes in how things are done are coupled with changes in how individual faculty members’ engage in teaching their courses. Results from the interview study of early Tutorial adopters suggest that 1) some initial individual dissatisfaction with current instructional methods may be critical in decisions to try something new, 2) locally collected data may be particularly compelling for other faculty members, and 3) support and resources offered for the new methods may be persuasive. The third analysis from a frames-of-context perspective demonstrates that adoption of the Tutorials required changes throughout a complex, nested educational system. This descriptive analysis suggests that a model of change (or claims about change) requires simultaneous attention to multiple levels of the educational system and how these levels interact. Based on frames-of-context analysis and our preliminary observations in Tutorial recitations, TA/LA Tutorial Training Meetings and associated course lectures, we expect that there is significant variation in how the Tutorials are implemented and integrated into the course.

In Chapter 5 (which addresses research question 3), we proceed to investigate whether students’ survey responses on end-of-term surveys about the utility and enjoyment associated with the Tutorials vary from semester-to-semester. This investigation is motivated by the observed differences in Tutorial implementation which may influence how Tutorials are perceived by students. Through analysis of student survey responses, significant semester-to-semester variations in students’ survey responses about the Tutorials are found. For example, one semester only 25% of students report that Tutorials are useful for their learning while another semester as many as 50% of students are reporting that Tutorials are useful for their learning. Although students are sometimes found to be quite favorable about the utility of the Tutorials, students are often quite negative about the enjoyment associated with the Tutorials. We find that students consistently respond more favorably about working with their peers in Tutorials than they do about the Tutorials more broadly. We demonstrate that variations in student background factors are unlikely to account for variations in students’ perceptions of Tutorials on a semester-
to-semester basis. After carefully examining particular semesters of the second semester physics course where one of the two co-teachers didn’t change, we found that the secondary professor can influence students’ perceptions of Tutorials. We find evidence that instructors’ decisions in implementing Tutorials (likely in either the primary or secondary roles) may influence students’ perceptions of Tutorials.

This thesis concludes with Chapter 6 where results are summarized and connections across research studies are described. Here we look across the examinations of Peer Instruction and Tutorials in the introductory physics courses. We summarize the extent to which the integration of these instructional strategies has led to fundamental changes in how students engage in learning physics as compared to traditional lecture courses. The complexities of change processes are summarized drawing from concrete examples of classroom routines.
References (Chapter 1)


[22] Average normalized learning gain is defined as ratio of the actual average gain (%post - %pre) to the maximum possible average gain (100 - %pre).


Chapter 2: Variations in Physics Professors’ Implementation of Peer Instruction

I. Introduction
Consider two representative vignettes of interactive engagement in a large-enrollment introductory physics course at the University of Colorado at Boulder:

Professor Green’s Classroom. Professor Green displays a ConcepTest [1] for his students to answer. The students discuss the question with their peers and enter their answers on individual handheld devices. 75% of the students answered correctly after the peer discussion. The professor requests an explanation from the students. One student contributes a clear and concise explanation. The professor paraphrases the student’s explanation, agrees that it is the correct solution, and moves on.

Professor Red’s Classroom. Setting up the clicker question in a similar way, Professor Red displays a ConcepTest, the students respond after discussing with their peers and the professor, and enter their answers individually. 75% percent of students answered correctly after peer discussion. The professor requests an explanation from the students, and many students respond, each giving the reasoning behind his/her answer. After a student speaks, the professor repeats the student’s idea so that the other students can hear the idea. Most of the contributing students comment or build on previous comments made by fellow students, arguing and debating with each other. After the students seem to have made a good amount of progress on their own, the professor displays his solution on a PowerPoint slide and walks through it quickly.

These two vignettes of classroom practice are easily recognizable as Peer Instruction [1], one of the more widespread innovations in introductory college physics courses [2]. Both professors appear to be actively engaging their students; they employ the same tool (personal response systems), present the ConcepTests similarly, etc. Both professors demonstrate certain hallmarks of interactive engagement (IE), but we also note significant variation in their implementations. In Green’s class, one student contributes a correct explanation, and the professor moves on; while in Red’s course more students are engaged in public discussion and debate about the reasoning process and the correctness and completeness of the ideas presented. Noting these and related differences, we seek to describe how variation in faculty practices impact student learning. In this chapter, we begin to address this question by developing a system for describing and measuring classroom practices that contribute to the construction of different classroom norms [3,4,5,6,7], i.e., different roles and rules of using personal response systems and Peer Instruction in these classrooms. Through an investigation of six, large-enrollment lecture courses that use personal response systems (or, “clickers”) as a primary tool for IE, we show how differences in instructors’ practices can be delineated and measured. Then, we discuss how these pedagogical differences result in different roles for students and instructor as well as different rules for the use of clickers in the classroom. We find that variations in educators’ practices result in disparate opportunities for students to practice conceptual reasoning [1,8,9], skills at “talking physics” [10,11], agency [12,13,14], and scientific inquiry [8,15,16].
II. Background
A. Prior Research

In K-12 educational research, the idea of a “teacher proof” curriculum (a curriculum independent of instructor effects) has been largely debunked [17,18,19]. Researchers have documented variations in teacher practice in the context of K-12 classroom; however, little has been done to document educational practices in the context of higher education. What little work has been done has tended to document broad distinctions such as reformed or not reformed, interactive engagement or not interactive engagement, student-centered or not student-centered. While useful, these categorizations are too coarsely grained to establish research metrics for distinguishing specific implementations and to inform the specific instructional choices of educators. Research on curriculum and instruction implementation has highlighted the importance of and lack of attention to the roles and role relationships necessitated by curricular change [20]. For example, if in one curriculum students are accustomed to the teacher being the primary source of information in the class, and in another curriculum students are accustomed to experiments conducted by the students as being the primary source of information, an organizational change in the roles (of the teacher and students) and the relationships between the participants must occur in significant ways. Fullan and Pomfret’s review of research also stated that organizational aspects of curricular change are usually not evaluated and left implicit in discussions of the curricular design, if addressed at all [20]. The research presented in this chapter provides a metric for documenting and discussing these organizational changes associated with the use of Peer Instruction. Here, we present a fine-grained set of characteristics of instructional practice that are both research-based and theoretically grounded.

Research studies that have examined the adoption and use of pedagogical innovations in undergraduate physics have tended to examine (1) student learning through pre-/post multiple-choice content assessments or (2) instructors’ perspectives on the use of physics education research (PER)-based approaches in teaching through surveys or interviews. A well-cited example of this first type of research is the 1998 paper by Hake [21], which showed consistently higher learning gains for courses where the professors reported using “interactive engagement methods” versus traditional lecture-format classes. However, a closer look at these data reveals that there is still a significant variation in student learning gains within interactive engagement classrooms; the bottom decile of IE courses achieve average normalized learning gains [22] ranging from about 0.16 to 0.24 while the top decile of IE courses achieve average normalized learning gains ranging from about 0.60 to 0.64 [21]. Hake hypothesizes that large variations in average student learning gains may be due to “course-to-course variations in the effectiveness of the pedagogy and/or implementation” [Ref. 21, p. 66]. Hake’s study establishes that there is large variation in average student learning gains across courses but leaves unanswered exactly what variations in pedagogical practices may exist between these classrooms that may affect student learning.

Other researchers in physics education have documented variation among student learning gains for various implementations of the same curriculum. In a study of five university physics courses implementing Workshop Physics [23] across different institutions and different instructors, researchers at the University of Maryland found that average normalized learning gains for these courses ranged from 0.39 to 0.57 (Ref. 24) as measured by the Force Concept Inventory (FCI) [25]. Similarly, researchers at CU studied five introductory courses implementing the Tutorials in Introductory Physics [26] in recitations at the same institution and found a range of average normalized learning gains in different courses [from 0.45 to 0.64 (Ref.
as measured by the Force and Motion Conceptual Evaluation (FMCE) [28]. Such significant variation in student learning suggests possible differences in how these curricula are implemented; calling for a characterization of faculty practices and a framework for measuring similarities and differences in instructors’ implementation of interactive engagement techniques.

Another avenue of research, focused on the adoption and use of particular pedagogical innovations in PER, has examined professors’ beliefs, values, or conceptions about teaching and learning [29,30,31,32,33]. These research studies, drawing from evidence gathered through interviews or surveys, develop frameworks for examining professors’ broad conceptions about teaching and learning [30], reported classroom practices [30], and beliefs and values about teaching and learning in the specific context of physics problem solving [32]. This work has contributed to a deeper understanding of how professors think about teaching and learning in higher education and has questioned the common assumption that professors’ beliefs about teaching and learning sit in conflict with the principles of research-based educational reforms [30,32]. Furthermore, the work of Henderson and Dancy [29,30,31] found that although professors’ conceptions about teaching and learning physics align with nontraditional or alternative educational ideas, many professors report that their practices did not always align with these conceptions due to situational constraints. That is, faculty practices are constrained more by structural considerations (such as expectations of content coverage, lack of instructor time, class size, or room layout) than by their beliefs about productive educational practices [31].

Building on and complementing these prior research programs, the thesis work presented in this chapter investigates the actual practices of professors within similar situational constraints who implement the same pedagogical technique. We focus on faculty use of Peer Instruction and develop a framework and measurement tool for describing the differences and similarities in how professors conduct the same instructional activity. In this present work, we compare six university faculty members implementing Peer Instruction in six different introductory physics courses and document similarity and variation in faculty practices through classroom observations, audio recordings of lecture classes, and interviews with these faculty members. These studies result in an observational protocol for documenting professor’s implementation of Peer Instruction, measurements of observable aspects of faculty practice that can vary, often significantly, and a framework for documenting variation along thirteen dimensions of practice. We also show how these differences in practices surrounding the implementation of PI impact the opportunities students have to engage in various scientific practices and create different classroom norms (cultures). In the next chapter, we will examine how these practices and norms are associated with student perceptions in these environments.

B. Studying an intervention: Peer Instruction

One of the primary interactive engagement techniques used in large-scale lectures at the University of Colorado at Boulder is Peer Instruction [1]. According to Mazur, who developed the method, Peer Instruction is a pedagogical approach in which the instructor stops lecture approximately every 10–15 minutes to pose a question to the students. These questions or ConceptTests are primarily multiple-choice, conceptual questions in which the possible answer options represent common student ideas. Mazur describes the Peer Instruction process as follows [1,34]:

General format of ConceptTest
(1) Question posed
(2) Students given time to think
(3) Students record or report individual answers
(4) Neighboring students discuss their answers
(5) Students record or report revised answers
(6) Feedback to teacher: Tally of answers
(7) Explanation of the correct answer

By implementing this questioning process as described, Mazur presents compelling evidence that Peer Instruction methods improved his students’ ability to complete both conceptual and traditional computational physics problems [1]. The Mazur research group has also investigated conceptual learning gains across eleven separate higher education institutions including thirty introductory physics courses in which Peer Instruction has been implemented [35]. According to this study, professors who implement Peer Instruction in this format on average achieve average normalized learning gains of 0.39±0.09 as measured by the FCI [35]. These results provide evidence that implementing this pedagogical technique can result in substantial student conceptual learning.

For the purposes of this chapter, we have collapsed the questioning format above into a three stage process which was found to be common across the implementations of Peer Instruction by the professors observed in this study. We use the following stages: the Clicker Question Set Up Stage (Mazur steps 1–2), the Clicker Question Response Stage (Mazur steps 3–5), and the Clicker Question Solution Discussion Stage (Mazur steps 6–7). These stages will be described in more detail in the data section.

Furthermore, the present study examines clicker questions (CQs) more broadly than focusing on solely ConcepTests or conceptual questions. Within this terminology, Peer Instruction (PI) as described by Mazur is a particular subset of clicker use, and ConcepTests are a subset of CQs. We see CQs as part of a broader domain of questions posed in class which are mediated by the technology of clickers [36,37,38]. We seek to characterize faculty use of Peer Instruction which requires a slight broadening of the definition of Peer Instruction (PI) based on results of faculty adaptation.

III. Bridging Classroom Norms and Observable Practices

We take classrooms to be cultural systems which are constituted by norms of behavior that arise out of the repeated use of shared practices [3,7]. Instructors and students make choices (implicit or explicit), which, in collection, establish a microculture with specific norms and expectations of the participants [3,6,7,39]. These microcultures are important because they are tightly coupled to the new understandings about physics, the nature of learning, and the nature of physics that participants develop as part of the course [7,39]. In order to connect classroom norms to specific, pedagogically relevant, observable classroom practices, we describe two finer-grained scales of practice which in combination, over time, make up the classroom microculture: 1) observable characteristics of practice and 2) dimensions of practice (DoP) which are both further explained below. Collections of observable characteristics of practice (such as the professor leaving the stage) make up DoPs (such as faculty-student collaboration), and combinations of DoPs make up classroom norms (such as a classroom highly valuing faculty-student collaboration). We begin by describing the DoPs. We then present observable characteristics of practice that were documented in six classrooms implementing PI and link these observable characteristics to the associated DoPs. In the data analysis section, we connect the observable characteristics and DoPs to classroom norms through two methods: summaries of stand-out collections of practices in six classrooms and case studies of a single PI episode in two different classrooms.
As we began our observations of several courses implementing Peer Instruction, we initially found it difficult to describe and measure the dramatic differences that we saw. We sought to create a framework to guide the documentation and discussion of these differences. We particularly focused on practices that we hypothesized would influence student learning outcomes of content knowledge, epistemological beliefs about science [40,41], attitudes about learning science [42,43,44,45], and scientific abilities [16]. The first draft of our DoPs was based on preliminary classroom observations and a review of the literature along three themes: research into professors’ beliefs about teaching and learning physics [29,30,31,32,33,46], descriptive studies of classroom response system implementation [1,34,35,47,48,49,50], and broad educational research results [51,52,53,54,55]. We then began an iterative process of revising and modifying our DoPs on the basis of additional classroom observations, aspects of practice presented in the literature, and the utility of these dimensions in documenting the practices in the learning environments we were studying. Based on our collection of descriptive field notes [56,57,58] and a synthesis of literature on interactive engagement, we were able to frame and organize pedagogical considerations surrounding the implementation of Peer Instruction along 13 dimensions. This literature is described in more detail in Appendix B along with the detailed descriptions of the DoPs. These dimensions direct our classroom observations by guiding us to look at particular practices that have been demonstrated to impact student learning in other classroom contexts. In this way each DoP is based on broad educational research into student learning and our observations of PI practices in particular. These dimensions help us to focus on aspects of the classroom norms, obligations, and expectations about the participant roles and the rules of interaction [3,4,5,6,7] that surround the use of clickers.

We have organized these DoPs into two sets. The first set of dimensions involves how the professor communicates what the classroom participants will and will not be doing during this activity—his or her expectations of students. These decisions are negotiable boundaries placed on the activity by the professor in advance of the students actually attempting to solve the problem. We call this first set of dimensions “defining the academic task.” The second set of dimensions describes student-professor interactions during the academic task. The use of clickers is further defined and negotiated based on the nature of these interactions. Another set of dimensions could be developed to describe the details surrounding student-student interactions [59,60,61,62,63]; however, this set is beyond the scope of this study. The dimensions that we have identified are summarized in Table B. The dimensions described below are not completely independent but, rather, form a set of overlapping considerations that instructors (and students) manage in the classroom while implementing Peer Instruction. These DoPs are designed to, on the one hand, link to observable choices and actions that faculty make, and, on the other hand, lead collectively to establishing norms and expectations in the classroom. The purpose of the DoPs is to help us link classroom norms to specific, pedagogically relevant, observable classroom practices and choices of faculty.
### Table B: Summary of Dimensions of Practice.

<table>
<thead>
<tr>
<th>Dimensions of Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Do I integrate or coordinate the use of clickers within broader course activities and course evaluation? If so how?</td>
</tr>
<tr>
<td>(2) What do I want students to learn and what do I want to learn from the students?</td>
</tr>
<tr>
<td>(3) As I pose the question, how do I convey to the students what this question is about and what the students should be doing?</td>
</tr>
<tr>
<td>(4) Do I allow, support, or encourage student discussion during the CQ?</td>
</tr>
<tr>
<td>(5) Given the nature of this question and what I expect students to be doing, how long should this take?</td>
</tr>
<tr>
<td>(6) Should I walk around the room?</td>
</tr>
<tr>
<td>(7) Do I need to listen to students’ ideas and reasoning and what are the benefits of listening to students’ ideas?</td>
</tr>
<tr>
<td>(8) What kinds of interactions should I have with the students during the CQ?</td>
</tr>
<tr>
<td>(9) Do I build on students’ prior knowledge and in what ways?</td>
</tr>
<tr>
<td>(10) Should students voice their understanding and reasoning during class? If so when and how should this happen?</td>
</tr>
<tr>
<td>(11) How do I respond to the CQ results when there is a split response amongst the students?</td>
</tr>
<tr>
<td>(12) Is the information that I am gathering useful in determining what happens next? If so how is this information useful?</td>
</tr>
<tr>
<td>(13) Do I want to know where students got to? Where did the students get to?</td>
</tr>
</tbody>
</table>

### IV. Description of Setting and Methods

All courses observed in this study were large-enrollment introductory undergraduate physics courses with average lecture attendance ranging from 130 to 240 students. Pseudonyms have been chosen to assure the professors’ anonymity: Yellow, Green, Blue, Purple, Red, and White. Five of the six courses studied were courses required for science-based degree programs. The other, Purple’s course, was an elective course for non-science majors. The lead instructors for these courses varied from tenured professors to temporary instructors. Two of the six instructors observed, Green and White, were novices with respect to the use of Peer Instruction and clickers in large-enrollment courses. Both also happened to be temporary instructors with no experience teaching large-enrollment courses. Three of the instructors observed, Blue, Red, and Purple, were active members of the physics education research group at the University of Colorado at Boulder. It is also important to mention that prior to this research study Blue mentored Yellow as he learned to use clickers and Peer Instruction in his own large-enrollment...
introductory physics courses. These course and professor attributes are summarized in Appendix A, Table I for reference.

Although all of these educators used the language of Peer Instruction to describe their practices, none of them implemented Peer Instruction exactly as described by Mazur. Each of these professors used an electronic classroom response system to collect and tally the students’ votes. These systems do allow students to change their answers while the voting time is still open. The most notable variation between these professors’ practices and Mazur’s description is that none of the faculty observed in this study had an explicit “silent” phase of the CQ where the students came to an answer individually first. We observed significant student discussion in all classes. For most questions, we found that the noise in the room is very limited at the beginning of the CQ, and then the noise level quickly rises. We hypothesize that students were spending some fraction of the CQ response stage to think independently even though students were not asked to commit to an answer individually prior to peer discussion. In this way, the use of the term “Peer Instruction” by physics faculty and our use in this thesis should be loosely interpreted.

Ethnographic research methods [56,57,58] were used for this study. Through extensive engagement of the researcher within a community, ethnographic research aims to build models based on both insider and outsider perspectives. The researcher’s goal is to make explicit models of the implicit meaning systems of the community. These models are influenced by the insiders’ interpretations of events, but due to the researcher’s ability to both bring an outside perspective and a reflective approach to the system, the researcher is in a unique position to identify overarching patterns that can give sense to the patterns of the community. In this way, the researcher’s interpretation of the values, unspoken in the practices of the community, is important and can help to make explicit cultural elements that are largely implicit to the participants [64]. Following from this tradition, the data sources for this study are qualitative and quantitative, including interviews with each professor, audio recordings of a subset of lecture periods, daily electronic records of CQs asked along with student responses, broad descriptive field notes, focused observational field notes, and student survey responses surrounding clicker use and classroom dynamics.

For the six courses that constitute the focus of this study, field notes of two distinct types were collected: descriptive narrative field notes and focused observation rubric data. The first type of field notes, collected primarily at the beginning of the semester, were broad narrative descriptions of the instructor’s actions, interactions between students, and also interactions between the instructor and the students [65,66]. These preliminary field notes informed the creation of an observation rubric for better collecting aggregate patterns of interaction with primary emphasis being placed on the instructional choices of the professor. See Appendix C for the observation rubric and Appendix D for a user’s guide to accompany this rubric. Multiple researchers, two from the PER group at Colorado and one from another institution, used the rubric and its associated instructions and provided formative feedback on the instruction guide. In subsequent reliability studies conducted with an additional researcher, reliability of 80% or greater was achieved on all items, averaging 96% agreement overall. This observation rubric along with narrative field notes was designed to collect information relevant to the 13 DoPs.

The observation rubric was used as a descriptive and analytical tool in the observations of an additional 6–10 class periods that constituted at least 20% of the class periods for each course. Data from the observation rubrics were used to quantify and compare characteristics of instructor practices across courses. After comparing these aggregate data, the initial descriptive field notes
were revisited along with the audiotapes of the class periods to infer how these variations in practices were contributing to the norms and expectations of these classroom communities.

At the end of the semester, semi-structured interviews were conducted with each of the instructors who participated in the study [56,67]. The topics discussed included differences between a traditional lecture and an interactive engagement lecture, descriptions of what an engaged student would be doing in an introductory physics lecture, descriptions of what the professor and his/her students do during a typical CQ, purposes of clicker use, and the importance of various student activities associated with CQs such as articulating reasoning in the lecture. Additional information about each instructor’s classroom practices was available for all courses through the daily electronic data collected by the clicker software program, and additional artifacts such as course syllabi were collected through the course web pages.

V. Data

The thirteen DoPs, presented in Table B, frame the data collection and case studies of six professors’ practices. We begin by presenting these data organized by the chronological stages presented in the background section: Clicker Question Set Up, Clicker Question Response, and Clicker Question Solution Discussion. The Clicker Question Set Up stage includes the professor’s introduction and framing of the clicker question for the students, in addition to decisions made prior to class, such as the types of clicker questions to be presented. The Clicker Question Response stage is the time that the students are given to construct their answers to the clicker question and communicate their responses to the professor through an electronic response system. This response stage includes both silent and group response times, if these are offered. The Clicker Question Solution Discussion stage is the whole class explanation and solution discussion phases. It is the time that the class or professor spends constructing a public solution to the clicker question. We will revisit these data in the analysis section, triangulating multiple observable characteristics [68], to describe classroom norms for each professor.

A. Clicker question set up stage

Some instructor decisions, concerning the set up of the CQ, occur prior to any class period. In order to understand on a broad scale the norms and expectations surrounding clicker use in these various classroom communities, we analyzed the explicit expectations laid out in the course syllabi. The professors in all courses gave reading assignments to their students, but Just-in-Time Teaching (JiTT) (Ref. 69) was not used in any of the courses; therefore, students were not directly held responsible for completing these assignments. Although professors were occasionally observed using Interactive Lecture Demonstrations (ILDs) [70] and/or Socratic questioning, in all courses studied PI was the vastly dominant interactive engagement technique [71]. The course syllabi outline for the students how clickers will be used and also how students will be evaluated for their participation. Four of the six professors in this study, White, Blue, Purple, and Red, had a mandatory fraction of students’ course grade based on clicker points. In Blue’s course, clickers accounted for 1% of the students’ grade, and clicker points were awarded based on correctness of the response. In Red’s course, 15% of students’ grade was based on clicker participation, and CQs were rarely graded based on correctness (only for occasional reading quizzes). In White and Purple’s courses, clicker participation was 5% of the students’ grade, and the CQs were not graded based on correctness. In Purple’s course, students could also be awarded extra credit clicker points which were graded based on correctness, and these extra credit points could replace up to 12% of students’ total exam grade. In Yellow and Green’s
courses, CQs were graded based on correctness, and clicker points were only awarded as extra credit points to replace up to 10% of the student midterm exam grade. These grading policies give one perspective on the relative value of clicker participation and the emphasis on the correctness of the CQs. For example, if students are only awarded extra credit for clicker participation, they may think that the clicker activities are not very important, since it is not a required component of their grades. In a similar way, if students’ grades depend on the correctness of their answer, they may perceive the focus of clicker activities to be about getting the correct answer, rather than why the answer is correct.

Additionally, the role of clickers in the course is framed by the type of lecture resources made available for students. All of the courses involved in this study provided students with copies of CQs and CQ answers after the class period was over. Except for Yellow, all professors provided short explanations of the CQ solutions in the resources made available for the students. Three of the six professors, White, Purple, and Red, provided lecture notes with these CQs placed in sequence with the presentation of other relevant information. These lecture resources provide one piece of evidence about the degree to which CQs were embedded in the course and the degree to which the explanation to the CQ was emphasized.

Other messages concerning the role of clickers within the course are established implicitly through repeated patterns of classroom practice. Many additional patterns of practice will be discussed throughout the next two clicker implementation stages, but we will describe here some broad class-level patterns such as the average number of CQs per hour and the average fraction of students getting the CQs correct after peer discussion.

Using data collected through the clicker software, we were able to calculate the average number of CQs asked per hour of class using data from the entire semester, shown in Table C. From these data we find that Green asked the fewest number of questions per hour of class, averaging about 3.2 questions. Yellow, Blue, Purple, and Red asked a moderate number of questions per hour of class ranging from about 5 to 6.5. White asked the largest number of CQs per class averaging about 8.2 questions per hour of class.

<table>
<thead>
<tr>
<th></th>
<th>Yellow</th>
<th>Green</th>
<th>White</th>
<th>Blue</th>
<th>Purple</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQs/hr</td>
<td>5.9 ± 0.5 a</td>
<td>3.2 ± 0.1</td>
<td>8.2 ± 0.6</td>
<td>6.5 ± 0.4</td>
<td>5.0 ± 0.3</td>
<td>5.3 ± 0.4</td>
</tr>
</tbody>
</table>

aIn all tables the error stated is the standard error on the mean, unless otherwise stated. Over the term between 26-34 hrs of class time was analyzed which covered between 108-213 CQs from each course.

Similarly, we were able to calculate the average fraction of students getting the CQ correct. These data are summarized in Table D. Based on these calculations, we see that Blue has the highest average percent of correct student answers at about 76%, and White has the lowest percent correct at about 64%.

While the average fraction of correct responses do not vary dramatically, the distributions are found to be skewed toward higher percent correct and do vary by professor. As a result, in the classes of Blue and Red, half or more of the questions asked received 80% correct or greater. Another way of presenting this data is to look at the fraction of questions where at least 70% of students got the question correct. In Blue’s class, at least 70% of students answer correctly most of the time (71% of CQs), while in White’s class this does not occur most of the time (41% of CQs). This analysis shows that students are more commonly successful at answering CQs
correctly in Blue’s class as compared to White’s. The degree to which students are correctly responding to the CQs varies from course to course as evidenced by both the average and median values of average fraction of correct student answers.

| Table D: Percent of students getting the clicker question correct on average. |
|-------------------------------------------------|--------|--------|--------|--------|--------|--------|
| Percent correct                                | Yellow| Green  | White  | Blue   | Purple | Red    |
| Median percent correct                         | 69% ± 2%| 68% ± 2%| 64% ± 3%| 76% ± 1%| 70% ± 2%| 72% ± 4%|
| No. of CQs                                     | 69%   | 72%    | 65%    | 82%    | 75%    | 80%    |
| No. of CQs                                     | 197   | 117    | 56a    | 190    | 146    | 29a    |

*For these courses, the correct answer to the CQs needed to be entered manually by the researcher. Therefore, the correct answers were only entered for a subset of the CQs.*

During the clicker question set up stage, the instructor must also determine which types of CQs to ask. Through our analysis and observations, we broadly distinguish types of CQs: logistical questions and content questions. As described in more detail in Appendix B, logistical questions were questions used to poll students’ perspectives or opinions about the course. For the observed period, Green and White almost never (less than 2% of the time) asked logistical questions, while Yellow, Blue, Purple, and Red occasionally asked logistical questions (6–12% of the time). We also found that within the category of content questions, one of the professors observed, Red, occasionally used CQs to give graded reading quizzes.

Many educators and researchers have proposed various schemes for categorizing types of content-based CQs [1,72,73,74,75,76]. We limit our analysis of question type to three coarse-grained categories: Recall, Algorithmic, and Conceptual Questions. Descriptions of these categories can be found in the descriptions of the DoPs in the Appendix B and most closely resemble the categories used by Towns [73]. The results of a categorization of a random subset of CQs are shown in Table E. These results show that all of the courses are primarily (65–85% of the time) asking conceptual questions. Additionally, Red, White, and Yellow were found to be asking a relatively higher fraction of recall questions compared to the other professors. However, half of the recall questions (2 out of 4) asked by Red were reading quizzes.

| Table E: Fraction of Clicker Questions that were: Logistical, Recall, Algorithmic, or Conceptual. |
|-------------------------------------------------|--------|--------|--------|--------|--------|
| Fraction of Logistical CQs                      | Yellow | Green  | White  | Blue   | Purple |
| Fraction of Recall CQs                         | 0.06   | 0.20   | 0.06   | 0.69   |        |
| Fraction of Algorithmic CQs                    | 0.04   | 0.11   | 0.08   | 0.79   |        |
| Fraction of Conceptual CQs                     | 0.02   | 0.15   | 0.06   | 0.64   |        |
| Fraction of Recall CQs                        | 0.10   | 0.06   | 0.06   | 0.77   |        |
| Fraction of Algorithmic CQs                    | 0.06   | 0.09   | 0.06   | 0.79   |        |
| Fraction of Conceptual CQs                     | 0.12   | 0.24   | 0.00   | 0.64   |        |

Another classroom practice surrounding clickers that was not observed to vary significantly from question to question was the role of student-student collaboration in answering the CQ. In this way, it appears that whether student talk was allowed or not allowed during the CQs was set up early in the semester and not negotiated by the professor and the students on a question-by-question basis. In all of the classrooms observed, students were most often allowed and encouraged to discuss with their peers in constructing their CQ answer. In most cases, students were observed to spend a small amount of time at the beginning of the CQ quietly thinking before consulting with their peers. However, students were not expected to commit to
an answer individually by voting as described in Mazur’s original descriptions of Peer Instruction [1]. Red was the only professor observed to occasionally ask questions that were intended to be answered individually, and these questions were reading quiz questions. Out of the 38 CQs observed, Red asked that the CQ be answered individually only twice. These data from this stage are summarized in Appendix A, Table II.

To summarize, all professors are choosing similar types of CQs (Dimension of Practice No. 2), focusing primarily on conceptual knowledge in their question statements. Due to the nature of the questions being posed in class, students in all these courses are given the opportunity to try out and apply new physics concepts on their own. Another important similarity across these courses is that student discussion is allowed, encouraged, and does occur in all of these courses. In this way, the level of student-student collaboration (DoP 4) is comparable across these courses, and students are given similar opportunities to practice discussing physics with their peers. Although CQs are integrated into the lecture of each of these classes by asking questions throughout the evolution of the lecture, the extent to which CQs are integrated with the rest of the course and the course assessment varies from course to course (DoP 1). For example CQs have a different role in Yellow’s course as compared to Red because in Yellow’s course CQ solution explanations are not provided or embedded into the online lecture notes, and there is no mandatory fraction of the students’ grade that is dependent on their CQ participation or responses. In Red’s class, however, CQs and CQ solution explanations are embedded into the online resources, and a large mandatory fraction of the students’ grade is based on their clicker participation, which helps to place greater emphasis on the clicker activity.

B. Clicker question response stage

During the time interval where students were constructing an answer to a CQ, what did the professors do? Based on our observations, individual physics professors spent this time differently. One significant difference was the extent to which the professor left the stage area of the lecture hall and walked around the classroom among the students. The first column of data in Figure 6 shows the fraction of observed CQs where the professor left the stage area of the classroom. The professors also varied in how they interacted with the students during this CQ response time. The fraction of the observed CQs where the professor answered students’ questions (where the students initiated a question) or discussed with the students (initiated by the instructor) varied, as shown in columns two and three of Figure 6, respectively. We intentionally coded answering student questions and discussing with students distinctly. For example, if a student raised his/her hand and the professor approached the student to address his/her question, this interaction may evolve into a discussion, but these details were not always available to the researcher. The initiating event was a question by the student. This interaction would be coded as answering student question (but not discussing with students). These coding procedures may result in us underestimating the amount of discussion that occurred between faculty and students in classes where the professor commonly answered student questions.
Figure 6: (Color) The percentage of observed CQs where the professor was observed to participate with the students during the response time by leaving the stage (column 1), answering student questions (column 2), or actively discussing with the students (column 3). The error bars shown are the standard error on the proportion.

From Figure 6, we see that Yellow, Green, and Blue almost never leave the stage (less than 15% of the time). Based on our observations, Yellow and Green not only chose to stay in the stage area, but they chose to have very limited interactions with students during the response time. Each of these professors would specifically distance him or herself from the activity the students were engaged in. Green would stand in a far back corner of the stage, removing himself as an object of focus during that time, physically distancing himself from the students. Yellow often spent this time organizing his lecture materials and overhead projector sheets for the explanation period. He would also pace around the podium area as he monitored the clicker votes coming in. Although Blue did not often leave the stage (10% of the time), he occasionally engaged in listening to students in the first few rows of the class.

On the other hand, we see that Purple, Red, and White usually leave the stage (more than 50% of the time). These professors would walk among different areas of the classroom during questions, sometimes as far as the furthest back rows of the lecture hall. These professors would sometimes listen to student conversations without participating, and at other times they would discuss with students or answer student questions. We also see that Purple and Red are much more likely to be answering student questions during the response time, while Yellow, Green, Blue, and White are engaged in answering student questions during the CQ response time for less than a quarter of all CQs. Similarly, Yellow, Green, and Blue rarely (less than 20% of the time) discuss with students during the CQ response time. Red discusses with students very
often, approximately 80% of the time. Purple and White discuss with students moderately often, approximately 40–50% of the time.

Another important task of the professor during the CQ response time is to determine the length of time that students should be given to complete the CQ. The length of time students are given to respond provides additional evidence concerning the type of tasks the students are engaged in and also the amount of conversation between students that is expected. We first compare the average length of time given for students to respond to CQs as shown in Figure 7. From this comparison, we see that in general Green and Red give their students the most time to respond to the CQ, averaging about two and a half minutes. We see that Yellow, Blue, and Purple fall just below this upper group, on average giving their students just over two minutes to complete the CQs. Finally we see that White’s students are given the least amount of time to respond, averaging approximately one and a half minutes.

![Figure 7: (Color) Average time given for students to respond (seconds). The error bars shown are the standard error on the mean.](image)

By looking at the distribution of clicker timing data, we can begin to describe further differences between these professors’ practices. The time elapsed for each question, as captured by the clicker software, is binned into thirty-second intervals and shown in Figure 8 for two professors, Red and Yellow. These professors were chosen to represent one professor with a relatively high average time and one professor with a moderate average time, relatively.
From the comparison of these distributions, we see that while Yellow’s data appear as a Gaussian skewed toward lower time intervals, Red’s timing data appear fairly flat over the time ranges of half a minute to four and a half minutes. These variations in distributions inspired us to compare the standard deviations of the timing data by course. We found that the Yellow and Green had the smallest standard deviations, 59±3 and 61±4 s, respectively. White’s standard deviation was just greater than Yellow and Green at 69±3 s. The three highest standard deviations were Blue, Red, and Purple at 89±5, 103±6, and 116±7 s, respectively [77]. These data suggest that the tasks that Blue, Red, and Purple ask their students to complete during a CQ are more varied than in the classrooms of Yellow, Green, and White (although other explanation may be possible). These data from this stage are summarized in Appendix A, Table III.

In summary, we see that professors are creating or reducing spatial boundaries (DoP 6) between themselves and students to varying degrees. For example, Blue leaves the stage only 10% of the time, while Purple leaves the stage almost 80% of the time. This simple act can open new possibilities for students to interact with physicists. We can also see that faculty-student collaboration (DoP 8) varies across these courses. In one course, Blue only answers student questions during CQ about 20% of the time and discusses with students only 15% of the time. This is contrasted with Red’s course for example where student questions are answered 60% of the time and faculty-student small group discussions occur approximately 80% of the time. Students in these classrooms are given different opportunities to practice formulating and asking questions. Similarly, opportunities for the instructor to model discussion and justification practices are varied depending on the frequency and type of faculty-student collaboration.

C. Clicker question solution discussion stage

After the students had finished responding to the CQ, how did the professor conduct the explanation phase of Peer Instruction? As a preliminary metric, we used data collected from the observation rubric and the audio files to calculate the average amount of class time spent discussing the CQ solution, see Table F.
Table F: Average time spent discussing the CQ solution (minutes:seconds, N=number of questions).

<table>
<thead>
<tr>
<th></th>
<th>Average time spent discussing solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow</td>
<td>3:04 ± 0:18 (N=25)</td>
</tr>
<tr>
<td>Green</td>
<td>3:16 ± 0:30 (N=22)</td>
</tr>
<tr>
<td>White</td>
<td>1:10 ± 0:10 (N=29)</td>
</tr>
<tr>
<td>Blue</td>
<td>2:28 ± 0:18 (N=25)</td>
</tr>
<tr>
<td>Purple</td>
<td>3:26 ± 0:35 (N=28)</td>
</tr>
<tr>
<td>Red</td>
<td>3:42 ± 0:26 (N=25)</td>
</tr>
</tbody>
</table>

From these data we can see that White spends the least amount of time discussing the CQ solution, averaging about 1 min and 10 s. Blue spends about two and a half minutes discussing the solution, while Yellow, Green, Purple, and Red all spend over three minutes. Overall, the time spent discussing the CQ solution varies by as much as a factor of 3.

We have also identified two characteristics of this discussion that vary across professors: whether incorrect CQ answers were addressed and whether students actively contributed to the explanation of the CQ solution. The first column of data in Figure 9 shows the fraction of observed CQs where incorrect answers were discussed during the description of the solution. The second column of data in Figure 9 shows the fraction of the observed CQs where student explanation(s) of the CQ solution were heard during the whole class solution discussion.

![Figure 9: (Color) The percentage of observed CQs where the wrong answers were discussed and student explanations were heard. The error bars shown are the standard error on the proportion.](image-url)
From Figure 9 we can see that Purple and Red usually discuss incorrect CQ answers, and that they do so more frequently than the rest of the educators observed. Although Green and Blue are discussing incorrect options about the same fraction of the time, in Green’s class these occasional explanations of the incorrect options originated from the students, while in Blue’s class the occasional explanations of the incorrect options originated from the professor’s explanation of common student difficulties. In Purple and Red’s courses, the discussion of incorrect answer options was commonly originated from both the students and the professor. The second column of data in Figure 9 also shows that Green always uses student explanations when explaining the CQ solution. Purple and Red usually use student explanations in the construction of the CQ solution, while Yellow, Blue, and White rarely (less than 20% of the time) use student explanations.

The number of student explanations usually heard in class for a given question also fluctuated from course to course. When students were asked to contribute explanations of the CQ solution, Yellow on average hears from 2.2±0.2 students, Green: 1.4±0.1, Blue: 1.3±0.3, Purple: 2.3±0.5, and Red: 2.4±0.4. White was not included in this analysis since student explanations were only used once. We see that Blue and Green primarily hear from only one student concerning the correct answer, when student explanations are used, while Yellow, Purple, and Red usually hear from at least two students. This is characteristic of practice where Yellow, Purple, and Red’s classrooms place more emphasis on discussing the incorrect answers and associated student reasoning.

In order to get a further sense of how often students offer their explanations in each course, we calculated the average number of student explanations that are heard in every hour of class. We find that students most frequently speak publicly in the classrooms of Purple, Green, and Red which heard from an average of 4.2±0.5, 4.6±0.6, and 4.8±1.3 students per hour, respectively. Professor Yellow hears a moderate number of student explanations at 2.4±0.6 students per hour. Professor Blue and White hear from a relatively low number of students per hour at 0.6±0.4 and 0.1±0.1, respectively. These data from this stage are summarized in Appendix A, Table IV.

We find that faculty-student collaboration (DoP 8) was even more varied now that we have examined the CQ solution discussion stage. For example, in Green’s class students were always contributing their own descriptions of the CQ solution, while in Blue’s class student explanations of the solutions were heard only 10% of the time. Similarly, we see that the use of student voice (DoP 10) varied from course to course. Students in these courses were given different opportunities to practice communicating in public. Even when student explanations were heard, how many different students were heard varied. We also see that Purple, Green, and Red’s classes averaged the most number of students sharing (~4.5), while in Blue and White’s classrooms less than a single student contributed a solution publicly in each class period. Students were given different opportunities to practice identifying themselves as sources of solutions, explanations, or answers.

VI. Data Analysis
A. Summaries across Dimensions of Practice

Observable characteristics of practice vary from professor to professor, and we can see how different combinations of these characteristics over time, in aggregate, create different classroom norms and expectations surrounding the implementation of Peer Instruction. We use both the characteristics of practice described in the data section and the DoPs, which frame the
observations, to summarize the standout classroom norms surrounding clicker use. These descriptions are a narrative summary of the data presented in the previous section with augmentations from relevant field notes. We describe a few examples of how these collections of DoPs combine to create classroom norms; however, these are not an exhaustive examination of norms.

We are particularly interested to investigate if student professor interactions are modified from traditional patterns through the use of Peer Instruction. The research literature has documented a prevalent educator-student dialogic pattern of interactive engagement, referred to as an IRE/IRF interaction sequence [78,79,80]. In these interactions between the professor and the students, the educator initiates (I) an exchange, usually by asking the students a question; then the student(s) respond (R), and the educator gives evaluation (E) or feedback (F) based on the responses. Within these traditional patterns of interaction, the educator holds all of the responsibility for deciding which questions will be asked and also for determining the correctness and completeness of the responses provided [81]. Within this pattern of interaction, students are not given opportunities to practice formulating and asking questions, evaluating the correctness and completeness of problem solutions, or identifying themselves as sources of solutions, explanations or answers. Different educational environments break out of this traditional IRE pattern to varying degrees and, therefore, provide students with different opportunities to engage in these scientific practices.

While we emphasize characterizations that distinguish faculty practice, we note that each of these courses is considered a successful educational environment, engaging students and leading to student learning. All professors in this study asked at least a few CQs interspersed throughout the lecture (See Dimension of Practice 1 and data from Table C). All of the professors in this study asked conceptual questions the majority of the time (DoP. 2; Table E). All of the professors allowed and encouraged student-student collaboration during the CQ response stage, and the students exhibited significant discussion in all courses (DoP. 4). All of these uses of Peer Instruction resulted in new opportunities for student-student interactions and increased emphasis on conceptual understanding when compared to a traditional lecture. These instructors may be learning about the use of these tools, so these summaries of their practices represent a snapshot in time.

Blue. During both the CQ response time and the CQ solution discussion, there were few interactions between the faculty and the students (DoP. 6, 8, and 10; Figure 6 and Figure 9). In this way clickers were primarily used in an IRE format, and responsibility for determining the completeness and correctness of the solutions was not shared with the students. It is interesting that although this professor understood common student ideas that may lead to incorrect answer options, these common missteps were not usually discussed explicitly with the students. When the professor discussed the correct solution, the professor did build on students’ prior knowledge in his explanation as well as emphasize the concepts and big ideas that were being used in the solution (DoP. 9).

Yellow. There are many similarities between the standout clicker practices of Yellow’s and Blue’s classrooms as may be expected because Blue mentored Yellow in teaching a prior version of this course. It was the norm for the professor to have limited interactions with the students, and the students had a limited role in the construction and evaluation of the public solution (DoP. 6, 8, and 10, Figure 6 and Figure 9). The professor’s explanations were usually quite detailed, consisted of asking students questions at key substeps, and clearly illustrated the big conceptual steps on an overhead projector. Similar to Blue, clickers were used to foster new
Clickers were primarily used in a traditional IRE/IRF interaction sequence. 

Green. Although Green’s classroom looked similar to those of Yellow and Blue during the CQ voting, Green’s classroom used more student voice during the CQ solution discussion. During the CQ solution discussion, the professor usually requested student explanations (DoP. 7, 8, and 10, Figure 9). However, many times the professor would state that “the majority of people chose ‘a’” and then ask if “someone could give a quick motivation for response ‘a’” without discussing incorrect answer options (Figure 9). In this way, the professor usually gave the correct answer away early in the public discussion and only heard from one student who usually contributed a fairly complete and correct explanation (DoP. 10 and 11). Therefore, clickers were used to somewhat modify faculty-student interactions beyond the traditional IRE format; however, students were not given more responsibility for determining the correctness or completeness of the solution description. In this way, faculty-student collaboration was not significantly modified.

White. The CQ voting time looked very different in this course. White would usually walk around among the students (DoP. 6, Figure 6), sometimes answering student questions, and occasionally discussing with the students (DoP. 7 and 8, Figure 6). However, these interactions with students were usually brief (DoP. 8 and 10, Figure 7). The professor almost never requested student explanations during the solution discussion (DoP. 7 and 8, Figure 9). Students, therefore, had very little voice or role in the final construction of the answer solution (DoP. 10). Notably, White had the lowest percent of students answering correctly and the shortest time for response and solution description (Table D, Figure 7, and Table F). In this way, the use of clickers in this class was primarily for quick, check-for-understanding questions (DoP. 13) in which the professor was attempting to gauge if students sufficiently mastered a topic or not, but the use of clickers was not treated as a significant opportunity to involve students in significant sense-making (DoP. 12 and 13).

Red. During the introduction of the CQ, the professor would usually explicitly state to the students, “I am interested in your reasoning” or “I’m going to ask for you to explain your answer” (DoP. 3). While the students were responding to the CQ, the professor would wander around the room, answering questions or discussing with students by asking groups of students, “What do you guys think?” or “How are you all doing here?” (DoP. 6, 7, and 8, Figure 6). The professor would usually get a chance to interact with two to four different groups of students during the CQ response time (DoP. 7 and 8). During the CQ solution discussion, the professor would then ask students to contribute their explanations publicly in the whole class discussion (DoP. 7 and 8, Figure 9). The professor would usually hear from multiple students and would usually ask clarifying questions of the students as they described their solution (DoP. 8, 10, and 11). The professor would often follow one student’s explanation with a phrase like, “Does anyone want to retort?” In this way, the professor made a space for students to actively disagree with each other in a respectful way. Red’s classroom did establish significantly different forms of faculty-student collaboration.

Purple. Purple’s classroom looked similar to Red’s during both the CQ response stage and the CQ solution discussion stage. When introducing the CQ, Purple would remind the students to “try to convince yourself why the other answer options are wrong” or “what are some easy ways that other students might get the wrong answer” (DoP. 3). During both the CQ response stage and the solution discussion, Purple often collaborated with the students: walking around the room, answering student questions, and discussing with various groups of students
(DoP. 6, 7 and 8, Figure 6). Professor Purple usually asked students to contribute explanations of the CQ solution (DoP. 7 and 8, Figure 9). As the student contributed an explanation, the professor intermittently interrupted and asked other students if that first idea made sense, or repeated what the student said for the rest of the class (DoP. 8 and 10). In this way, clickers were used to change faculty-student interactions beyond the traditional IRE format. The professor usually heard from multiple students and verbally encouraged students to think about different ways to get to or think about the solution (DoP. 10 and 11).

Based on our observations, there are a variety of scientific practices that we, as educators, value and that students can gain experience with through the use of Peer Instruction:

(i) To try out and apply new physical concepts
(ii) To discuss physics content with their peers
(iii) To justify their reasoning to their peers
(iv) To debate physical reasoning with their peers
(v) To formulate questions and ask questions
(vi) To evaluate the correctness and completeness of problem solutions
(vii) To interact with physicists
(viii) To begin to identify themselves as sources of solutions, explanations or answers
(ix) To communicate in a public arena

While not traditionally assessed, there are a variety of practices such as those described above that we, as science educators, value for our students. Our studies demonstrate the potential for Peer Instruction to support the development of these scientific practices; however, the availability of opportunities for students to engage in scientific practices depends upon the specifics of PI implementation. In all of the classrooms studied, students were found practicing the first four items in this list. In other instances, there were large discrepancies in students’ opportunities to engage in the remaining five practices. The large discrepancies in students’ opportunities to engage in the last five practices will be further illustrated in the following case studies from Red and Green’s classrooms.

B. Case studies illustrating a classroom norm

Now that we have summarized some differences that exist on the scale of the course, we can demonstrate one utility of the dimensions by identifying key differences between professors’ implementation of a single conceptual CQ. Collections of varying DoPs accumulate to create differing norms within the classrooms—differing roles and rules for professor and student participation. We present case studies of a typical conceptual CQ from Red’s classroom and another from Green’s classroom. The case studies below draw from audio data, observational notes, and clicker software data.

1. Green CQ Case Study: Calculus-based Introductory Physics 2

This is the second CQ of the class which, begins about twenty-six minutes into the fifty-minute class period. Prior to posing this CQ, the professor has briefly discussed the domain model of magnetism and has described permanent magnets. A student asks a question about particular materials and their magnetic properties, which the professor addresses. The professor then says, “It’s time to go to the next chapter… electromagnetic induction. [pause].” The professor then begins writing the title of the chapter on the board. The professor continues, “I think that this is something that you can actually understand based on what we have done before. So I will start with asking a question on it before I have really started the chapter.”
The professor puts up the CQ (shown in Figure 10) and describes the question to the students: “So here I have a magnetic field going into the board and then a conducting ball, a metal ball, is moving through the magnetic field… moving to the right. And if you remember now that a conductor has lots of valence electrons that can move around inside the conductor then you should be able to determine what will happen with this ball when it moves through this field. And there are options there, that it will be polarized in different directions or that it will not be affected at all.” The professor spent about 40 s on his introduction to the CQ. After about 30 s in the voting, the professor asks, “Is it pretty clear what’s happening here? If there is anyone that thinks that anything in this question is not clear please raise your hand.” The noise level rises as the students begin to discuss. Not one of the students raises a hand. The professor replies, “Okay, good.”

During the CQ voting time, the professor stands by a door that is located at the very back left corner of the stage. He paces around the front of this doorway for most of the CQ voting time. Then the professor walks to the podium and checks the incoming clicker votes. Meanwhile, there seems to be a significant amount of discussion occurring among the students.

The professor warns the students, “Okay, 20 more seconds.” A little bit later the professor says, “Last few votes. Okay, I’ll stop it there.” The voting time lasted about 2 min and 30 s. The professor displayed the voting results (A: 72% B: 17% C: 4% D: 2% E: 5%). The professor says, “Most people thought that it would be polarized for sure and that it would be polarized and… that it has a net positive charge on the top and a net negative on bottom. Can somebody explain how they determined that?” Pause. One student raises his hand to offer an explanation. The professor calls on the student by name, “Joe.” The student explains, “Well in the ball the positive charges are moving to the right, so it’s an effective current to the right. With a B-field into the board, so the positive charges would be deflected by a force trying to push them up. And the negative charges are moving to the right, but it’s an effective current for the negative charges to the left and B-field into the board, so the

![Metal ball in \( \vec{B} \)-field](image-url)  

Figure 10: (Color) Screen shot of a conceptual CQ in Green’s classroom (correct answer: A).
force on the negative charges would be pointing down to the bottom of the ball. Does this make sense? [pause]. Yeah, it does make sense.” The students laugh at this comment. The professor continues, “But is it completely true though? …Both of these things? Or is it just one of these things that is true.” The students respond in murmurs, “Only one, the second one.” The professor continues, “Yeah, we usually think of the nuclei, the positive charges, in a conductor as being fixed and it is electrons that move around, but it is perfectly fine to think of positive charges moving as well. We can’t see positive charges are not moving around. But if we measure it, it will look like the positive charges have moved. Since usually… we will now consider current as well, which is like positive charges moving. So, it will be convenient to think of the positive charges moving as a result of force. Excellent. So I guess that I gave away that that was the correct response.” The solution discussion period lasted approximately 3 min.

The professor continues into a discussion of how this question is an example of electromagnetic induction using examples from demonstration equipment.

2. Red CQ Case Study: Calculus-based Introductory Physics 3

This is the second CQ of the class which begins about 10 min into the 50 min class period. The question is preceded by a discussion of what the work function is and the range of values of initial kinetic energy that the electrons could reasonably have. The professor has used a representation of a well with balls stacked in it along levels and that these balls are given kicks by photons. The professor has walked through an energy conservation argument for this exact physical situation when blue light interacts with the metal.

The professor puts up a clicker question (see Figure 11). The professor says, “Enough of me yammering. Electrons can have a large range of energy and equal chances of absorbing a photon. Okay. So umm, if I come in with higher energy light, initially you have blue light shining on a metal and if you change that frequency to violet light, at the same number of photons per second okay… So I’ve increased the intensity, but I have the same number of photons coming in per second, but the energy in the violet photons is… bigger or smaller?” The students call out answers, mostly saying bigger. The professor continues, “Bigger, okay. What happens to the number of electrons coming out?” He says, “So get into your discussion groups and chit chat.” The introduction of the question lasts about 50 s and shortly after, the students begin to discuss with each other.

![Figure 11: (Color) Screen shot of a conceptual CQ in Red’s classroom (correct answer: C).](image-url)
The professor starts the time for the clickers and wanders around the front of the room. He talks to a student in the front row. It is not obvious if he has initiated this interaction. Then he moves to a student in the second row on the other side of the room and he is heard asking a group of students, “What do you guys think?” The professor continues to engage in a discussion with this group of students.

After 2 min and 50 s the professor says, “Everybody in? Okay, Three, two, one. [CQ is closed with student responses: (A: 0%; B: 17%; C: 74%; D: 8%; E: 0%)] Okay, we might have an all time high in attendance. Okay, if we keep doing this do you know how many students we’re going to have at the end of the semester? An infinite number. [The students laugh.] That’s kinda cool. Students from all other universities are going to be piling into this class. So, I heard a bunch of great reasons. All of the reasoning was in essence correct that I heard; it’s just that some of the reasoning was incomplete. So, someone want to give a quick stab at what’s up?” The professor points to one of the students and says, “Yep.” The student says, “I said that more electrons got kicked out… because the photons have greater energy they are going to knock out more electrons from deeper inside the metal than they would have before.” The professor responds, “Okay does everybody agree that the purple or violet has greater energy than the blue? Okay, so then your argument is… if you got more energy then you can scoop down into the metal deeper, because the length of that arrow is longer, right? Okay… Do you want a [indiscernible candy name] or a chocolate?” After a student contributed a response, the student is tossed a piece of candy. The professor asks, “Okay… Does anybody want to retort?” The next student that speaks is inaudible on the recording, but the professor paraphrases the student comment to the rest of the class as follows, “Aaaha, so it could kick off. But wait a sec, there is enough from the blue to dig off from the top. Okay, so it could…” A student interrupts, “But don’t all the electrons have an equal probability of getting hit?” The professor says, “Aaaha, But photons aren’t very smart. They don’t know what ones they’re going to go for. So they all have equal probability. It’s not like there’s this hand guiding it.” A student asks another question, “I thought that there was always one photon kicking out one electron.” The professor responds, “Yes, One photon always interacts with one electron, but we don’t know which electron.” A student asks, “Just those top electrons?” The professor responds, “No, it could be any of those electrons.” Another few students speak. After the students seem to have made a good amount of progress on their own and have brought forward some of the key ideas, the professor displays his solution on a PowerPoint slide and walks through it fairly quickly. The solution discussion stage lasts about 5 min and 20 s. After an explanation of the question, the professor discusses typical photon energies and typical energy scales for work functions of different metals.

3. Comparative analysis of Red and Green case studies

Both professors are asking conceptual type CQ, and both Red and Green introduce their CQs in similar ways (DoP. 2 and 3). They both read the CQ out loud, but, rather, than just reading the question verbatim, they elaborate as they describe the question, reminding students of relevant ideas that have been previously discussed (DoP. 3). In both classes there is significant student-student discussion as evidenced by the noise level on the audio recording during the voting time (DoP. 4). Students in these two classrooms are given similar opportunities to discuss physics content with their peers. The professors spend a similar amount of time introducing the CQ (DoP. 3 and 5). In this way, the professors conduct the moment-to-moment set up of the CQ very similarly.
During the CQ response stage, the professors also give the students a similar amount of time to respond to the CQ (DoP. 5, Figure 8). However, the professors participate in the CQ response time differently (DoP. 6 and 8, Figure 6). Green stands at the front of the stage for the entire question while Red leaves the stage and actively discusses with students (DoP. 6 and 8, Figure 6). Red inquires and listens to what the students are thinking during this voting time (DoP. 7 and 8, Figure 6). We see that a similar fraction of students are getting the CQ correct in each of these cases. Students in these various courses are given different opportunities to practice interacting with physicists. Students in these classrooms are given different opportunities to practice formulating and asking questions. Similarly, opportunities for the instructor to model discussion and justification practices vary depending on the prevalence of faculty-student collaboration.

The most significant differences between Green and Red become apparent during the CQ solution discussion stage. Although both professors elicit student responses (DoP. 10), Green and Red spend significantly different amounts of time discussing the solution, Green: ~3 min and Red: ~5.5 min (DoP. 5). In addition to the differences in time spent, the types of participation from the professor and students vary during the solution discussion (DoP. 8, Figure 9). In Green’s case, only a single student explanation was elicited, and this student’s explanation was clear and correct (DoP. 10 and 11, Appendix A, Table IV). Following this correct student explanation, the professor communicated the correctness of this explanation and did not elicit additional student comments, although more than 25% of the students had answered the question incorrectly. In Red’s case, we see that multiple students contribute explanations, and some correct and some incorrect ideas are presented publicly (DoP. 10 and 11, Appendix A, Table IV). In this example, the student explanations build on fellow students’ answers (DoP. 11). Furthermore, each student contribution includes reasoning for his/her answer. In Red’s class, students are responsible for evaluating the correctness and completeness of the problem solution proposed by their peers. Students in these classrooms are given different opportunities to practice identifying themselves as sources and evaluators of solutions, explanations, or answers.

These differences result in different kinds of faculty-student collaboration (DoP. 8) and differences in the use of student prior knowledge (DoP. 9). Additionally, these differences in implementation contribute to varying degrees of emphasis on reasoning and sense making. It appears that although students do have a significant amount of voice in Green’s class (DoP. 10), the students who are contributing are usually contributing a clear and correct explanation to the CQ. Flawed student reasoning is not voiced equally in this class even on questions where there is a significant fraction of students incorrectly answering the CQ. Since incorrect ideas are not as likely to be shared, the importance of reasoning and sense making in this class is reduced. It is the answer that is predominantly valued.

Red’s course, on the other hand, further emphasizes the importance of reasoning through the professor’s management of disagreement among his students (DoP. 11). Because it was fairly uncommon for professors in our sample to foster discussion and debate among their students, it is worth describing how fostering debate and discussion was achieved in this specific case. From this case study we can see how Red encouraged student-to-faculty dialogue by asking clarifying questions (DoP. 8 and 10). Red also structured student-to-student dialogue during the solution discussion usually by positioning students in a way such that they should respond to or comment on another student’s contribution (DoP. 4 and 8). In this way, the professor structured the students’ interactions with other students such that they are debating, providing alternative explanations, arguing, defending, challenging, or clarifying each others
ideas. Students in Red’s class were given opportunities to practice communicating in public and defending and clarifying their scientific ideas.

VII. Conclusions

Although many professors talk about Peer Instruction and its implementation similarly in interviews, we have found that there are significant differences in professors’ classroom practices that combine over time to have significant pedagogical implications. Students are found to engage in different scientific practices in class which has significant pedagogical implications; having students engage in scientific practices in-class is essential for formative feedback by the professor and student peers to a given student on the execution of these practices according to community norms. If students are not engaged in these scientific practices in class, there are few opportunities for the students to receive feedback on these process-oriented learning goals.

We have identified observable and quantifiable aspects of practice which vary from classroom to classroom. Prior research has shown that faculty practices are constrained more by structural considerations (such as expectations of content coverage, lack of instructor time, class size, or room layout) than by their beliefs about productive educational practices [31]. In this investigation, we find that instructors within similar structural or situational constraints are making different instructional decisions. People seem to be negotiating the “same” constraints differently. These results suggest the need for a more detailed account of how instructors use their knowledge of educational innovations and situational constraints to arrive at practical decisions in the moment-to-moment demands of the classroom.

Differences in observable practices can be grouped along dimensions to illustrate the potential implications of small-scale classroom practices. We find that variation in teacher practice results in disparate opportunities for students to practice conceptual reasoning [1,8,9], skills at talking physics [10,11], agency [12,13,14], and scientific inquiry [8,15,16]. Based on our observations, there are a variety of scientific practices that students can gain experience with through the use of Peer Instruction. In all of the classrooms studied, students were found trying out and applying new physical concepts and discussing physics with their peers. However, there were large discrepancies in students’ opportunities to engage in formulating and asking questions, evaluating the correctness and completeness of problem solutions, interacting with physicists, identifying themselves as sources of solutions, explanations, or answers, and communicating scientific ideas in a public arena. Our investigation has uncovered possible benefits of particular implementations of Peer Instruction that are yet to be explored and assessed. The assessment of students’ facility with these scientific practices is a fruitful direction for future research.

Ultimately, these different classroom practices, over time, contribute to the construction of different local classroom norms and communicate different values to students. The case studies of Red’s implementation and Green’s implementation of PI demonstrate how in practice these professors place different degrees of emphasis on sense-making in the classroom. The following chapter will define particular norms of PI and investigate how students from these classrooms perceive the norms of the classroom and the use of PI differently.
Peer-reviewed Publications based on this work:


References (Chapter 2)


[22] Average normalized learning gain is defined as ratio of the actual average gain (%post-%pre) to the maximum possible average gain (100-%pre).


[36] We acknowledge that there is an on-going debate in the literature as to the impacts of the particular technology used (i.e., electronic response systems, colored cards, or simple hand-raising...)[See Refs. 37, 38]. We do not, however, debate the different affordances of these technologies here.


None of the instructors were observed to consistently ask questions in a sequences progression as suggested by Meltzer *et al.* [Ref. 50].
[77] Where the uncertainty in the standard deviation is given by \( \Delta \sigma = \frac{\sigma}{\sqrt{2(N-1)}} \) as described in J.R. Taylor, *An Introduction to Error Analysis 2nd Ed.* (University of Science Books, Sausalito CA, 1997).
Chapter 3: The Construction of Different Classroom Norms during Peer Instruction: Students Perceive Differences

I. Introduction

Based on observations of multiple Peer Instruction (PI) classrooms, presented in the prior chapter, we found that implementation practices of Peer Instruction [1] can vary widely from classroom to classroom [2]. Motivated by the striking differences between how students were engaging in PI classrooms, we developed a systematic tool for documenting differences in instructors’ fine-grained instructional practices during PI (described in Chapter 2). This research showed 1) how these practices can be documented and 2) how differences in PI implementation provide students with different opportunities to engage in scientific practices such as asking questions, evaluating the problem solutions of others, and justifying their reasoning. Although the research presented in the prior chapter showed that different PI practices provide different opportunities for students to engage in scientific practices, this work did not show that these differences in PI implementation made a difference to students. This research study follows up by addressing student perceptions of variation in classroom norms. We link the differences in PI implementation practices to different student perceptions of PI. How students perceive PI within a given course is important to understand because it will likely influence how students engage (or do not engage) in the course and PI activities particularly. How students engage in the course may in turn affect what scientific practices students engage in throughout the term and what they will learn in the course.

In order to link PI implementation practices to students’ perceptions of PI, it is necessary to aggregate particular fine-grained PI implementation practices into factors that would have meaning to students. Towards this end, we show how researchers can construct descriptions of classroom norms from collections of observed practices. We argue that collections of fine-grained classroom practices over time lead to the construction of classroom norms: a shared meaning system that gives sense or coherence to the community’s collective activity. We focus on three particular classroom norms integral to PI: faculty-student collaboration, student-student collaboration, and emphasis on sense-making vs. answer-making. The first two of these norms are particularly important because they are mechanisms by which student thinking is made visible and available for formative feedback from the professor or fellow peers. Understanding the emphasis on sense-making versus answer-making in the class allows us to describe the kinds of tasks or activities students are engaged in during PI. After presenting definitions of these norms in terms of concrete classroom practices, we then investigate whether students’ perceptions of these norms are consistent with the researchers’ inferred norms, by surveying students’ perceptions of classroom norms. We see strong consistency between the researchers’ inferred classroom norms (what we believe to be going on in class) and students’ perceptions of key elements of these classroom norms. These results suggest that how faculty members (and students) engage in class establishes different rules and roles (norms) for participants. We show that what instructors do in the classroom correlates with what students perceive as valued in the activity of PI.

II. Background

A. Description of Peer Instruction

According to Mazur, Peer Instruction [1] is a pedagogical approach in which the instructor stops lecture periodically to pose a question to the students. These questions or
ConcepTests are primarily multiple-choice, conceptual questions in which the possible answer options represent common student ideas. Mazur describes the Peer Instruction process as follows [1,3]:

1. Question posed
2. Students given time to think
3. Students record/report individual answers
4. Neighboring students discuss their answers
5. Students record/report revised answers
6. Feedback to teacher: Tally of answers
7. Explanation of the correct answer

Peer Instruction is one of the primary interactive engagement techniques used at the University of Colorado, Boulder (CU) in their large enrollment physics courses [2]. The instructors of these courses use an electronic classroom response system (“clickers”) to collect and tally the students’ votes. Questions that are asked using the electronic response system are called Clicker Questions (CQ). The most notable variation between CU physics professors’ practices and Mazur’s description is that faculty rarely have an explicit "silent" phase (Steps 2 and 3) of the CQ where the students come to an answer individually first [2]. The rest of the PI process is implemented as described above. In all classes observed, students discuss the CQ and then report their answers. Significant student discussion occurs in all classes observed [2]. In this way, the use of the term "Peer Instruction" by physics faculty and our use in this thesis includes slight variations on the general format described above.

B. Research Study Purpose and Research Questions

The purpose of this research study is 1) to describe how differences in PI implementation practices lead to the construction of discernibly different classroom norms and 2) to investigate whether differences in PI implementation practices result in different student perceptions of PI norms. Understanding types of interaction (components of classroom norms) that occur during PI is important because patterns of interaction constrain which responsibilities fall to the shoulders of different participants. These interactions also provide emergent resources that can be utilized by faculty and students in shaping how the class proceeds. We address the following research questions:

1) Can differences in PI norms be delineated by using definitions of norms that are tightly linked to observable classroom practices?
   Associated analytical questions--
   • What types of faculty-student interaction are happening?
   • What roles and associated responsibilities are available to the classroom participants (faculty or students)?
   • What instructional practices foster or constrain student-student collaboration?

2) Do students notice differences in PI norms? And are these perceptions associated with what the researchers observe to be going on?

III. Defining Classroom Norms
A. Broad Theoretical Approach

We consider classrooms to be cultural systems which are constituted by norms of behavior that arise out of the repeated use of shared practices [4,5]. As other sociocultural
researchers have claimed, “Every continuing social group, such as members of a classroom or workplace, develops a culture or set of social relationships that are peculiar and common to its members” [Ref. 6, 110]. Instructors and students make choices (implicit or explicit), which, in collection, establish a microculture with specific norms and expectations of the participants [4,7,8]. These microcultures can be described by the everyday activities of the participants, their ways of talking and interacting with each other, and their selective use of tools during their ongoing activity [9,10]. In our classrooms, students come to understand these microcultures in parallel with developing an understanding of physics content.

For the purposes of this research study, we take learning to be the internalization of social norms and practices. This definition captures both how students come to know a physics topic (e.g. the scientifically accepted definition of Newton’s Second Law) and how students may simultaneously come to know how science is done (e.g. what counts as justification for a scientific answer). The emphasis in this definition on the everyday activities of the classroom is helpful for understanding how students may be learning things in physics classrooms that physics instructors are not explicitly intending to teach, often referred to as the hidden curriculum [11]. In the context of physics learning, investigating classroom microcultures is important because the practices of the classroom are tightly coupled to the new understandings about physics, the nature of learning, and the nature of physics that students develop as part of the class [8,12].

For example, consider a student working on completing a ConcepTest about Newton’s Second Law while discussing with his/her peers. In completing this task, the student may develop an understanding of Newton’s Second Law, but simultaneously the student may develop an understanding about how discussion with others is useful for clarifying scientific ideas (See Figure 12). In this way the classroom practice of talking with your neighbor during class can simultaneously support students’ understanding of a physics topic and their understanding of classroom norms.
B. What Are Classroom Norms And How Are They Constructed?

Norms are shared meanings or interpretations about the roles and rules of a social activity [4]. These shared meanings are largely implicit, but provide a degree of coherence and stability to local shared activity. Norms allow participants to coordinate their activity and share some common sense of ‘What is it that’s going on here?’ [13]. These shared meanings must be inferred and, as a practical matter, a researcher may choose to start with fairly low inference interpretations and move towards more high inference interpretations throughout the investigation. In the example shown in Figure 12, one might say that classroom practice of talking to your neighbor implies that talk is allowed (a low inference norm), or one might say that the classroom practice of talking to your neighbor implies that discussion with other is useful for clarifying scientific ideas (a high inference norm). Each of these claims for a norm would require different evidence to support each claim.

Consider the following concrete example. If a physics professor were to walk into one of the primary lecture halls in the CU physics department, he or she may see one of the following things happening. He or she may see students turning to their neighbors to discuss physics problems. Talking amongst your neighbors during an introductory physics course is a norm in all introductory physics courses at CU [2]. However, if the professor instead walked in during the weekly professional physics colloquium held in the same room, he or she would observe physics faculty and graduate students quietly listening to a speaker presenting PowerPoint slides. Talking amongst your neighbors during a physics colloquium is not the norm at CU (a low inference norm). Although the setting of these activities is the same and these activities serve somewhat similar purposes, the norm varies for when talking is implicitly ‘allowed’ during these activities.

Norms are established through repeated engagement in social practices. Norms are socially negotiated and collectively agreed upon, although power and authority may not be
equally distributed amongst participants. These norms carry with them implicit value-sets of the culture (or microculture). Approaching classrooms as cultural systems (and describing their norms) helps to draw attention to the fact that teaching and learning “necessarily involves affording or constraining access to value-laden resources that affect the level and kinds of participation that individuals might achieve in a community” [Ref. 6, pg. 111]. For example, if communicating one’s scientific ideas is a valued practice in a scientific discipline and students are not given the opportunity to practice communicating their ideas in a science course, then we have constrained students’ access to this valued disciplinary resource.

Although describing patterns in social practices and classroom norms is useful, one must keep in mind that individuals are always improvising and modifying these systems. As Lemke says, “People are not slaves to the activity structure of their community. We do not just ‘follow the rules’—we use those rules as resources for playing the game according to our own strategies” [Ref. 14, pg. 9]. The actions and interpretations of these actions are always highly contingent on the context.

C. Classroom Norms in the Context of Peer Instruction

In studying Peer Instruction implementation, we have found a consistent three stage progression during a given PI episode. All PI episodes were found to begin with a whole-class discussion in which the CQ is posed, then the classroom participants proceed to work in small groups discussing the CQ, and finally conclude with a whole-class discussion of the CQ solution. In this way a PI episode consists of whole-class, small-group, and then whole-class discussion.

We distinguish between two kinds of small group discussions, one which consists of student-only small group work and another which consists of the educator and student(s) working together in a small group. Whole-class discussion in the context of PI occurs largely during the introduction of the CQ, the first stage in the progression, and during the public discussion of the CQ solution, the third stage in the progression. These three modes of participation (whole-class discussion, student-only small-group discussion, and student-educator small group discussion) were found to organize the vast majority of classroom interactions during Peer Instruction [15,16,17,18].

Within each of these modes of participation, we found multiple possible types of interactions that may occur. By specifying particular types of interactions that occur during these three modes of participation, we seek to codify the relations among participants, the corresponding roles of participants, and the normative expectations for appropriate behavior [19]. Each interaction consists of a series of actions or instructional moves, such as the instructor leaving the stage. We refer to these instructional moves as part of clicker use since they are applicable to a wide variety of pedagogical uses of clickers (more broadly defined than PI). Each instructional move or classroom practice may have multiple possible meanings (see prior discussion about low and high inference interpretations), however across the coordination of multiple practices over time particular meanings become more prevalent or preferred [20]. These prevalent meanings or interpretations make up the norms of the community.

Of the norms examined in this research study, two are explicitly about interaction: the norm of faculty-student collaboration and the norm of student-student collaboration. The PI norm of faculty-student collaboration can be understood by examining the types of interactions that occur during the whole-class discussions and the student-educator small group discussions. The PI norm of student-student collaboration can be understood by examining the classroom practices that shape student-only small group discussion. The prevalence of collaboration is
important to understand since it is a primary mechanism by which student thinking is made visible and available for formative feedback by fellow peers or the professor.

Based on the design of our research study, there are important differences in the depth and breadth of our characterization of these three modes of participation (whole-class discussion, student-only small-group discussion, and student-educator small group discussion). Through our research design, the whole-class discussion mode was always available to be documented and coded by the researcher since whole-class interactions occur in the public arena. However, the student small-group discussions occurred all over the classroom and the researcher only documented the pocket of student small-group discussions that were occurring in the immediate vicinity of the researcher. Similarly, in the classes where faculty-student small-group discussions occurred, only a handful of these interactions were available for the researcher to document and code, since they would only occasionally occur in the direct vicinity of the researcher. Small group interactions are also framed by the opportunities and explicit expectations created in the classroom which we describe based on observable practices. As a result, the PI whole-class discussion practices are described more completely than small-group discussion practices. In describing the constituent practices of these classroom cultures, we note that students may be engaged in other (unobservable) cognitive tasks; however, we focus only on the observable actions of students that are made public for feedback from either peers or the professor and therefore provide resources in the on-going collective activity of the participants.

We define three classroom norms that are of particular interest during the activity of PI: faculty-student collaboration, student-student collaboration, and emphasis on sense-making vs. answer-making. A given practice may contribute to multiple classroom norms which results in norms that are overlapping and co-constituting. We describe the common patterns of engagement of students and faculty in the classrooms of professors Yellow, Green, and Red. For full comparisons of the practices in these classes, please see the associated tables and figures in Chapter 2 (which are also contained in Ref. 2). These data are summarized here. The descriptions of these classroom cultures and their constituent practices allow us to compare these classrooms along each of these norms. We then present student survey responses on questions designed to probe student thinking about each of these norms. We show substantial correspondence between the researchers’ interpretation of the PI classroom norms and students’ perceptions of these classrooms. We see that particular collections of PI implementation practices may support students’ perceptions of these activities as valuing faculty-student collaboration, student-student collaboration, and sense-making.

IV. Review of Research on PI and Clicker Use

We review prior research literature relevant to understanding Peer Instruction, broader studies of clicker use, and impacts on students. This research falls into three categories: impacts of clicker use on students’ content knowledge, effects of grading incentives on student discussion practices, and students’ perceptions of the use of clickers. Reviewing the impacts of clicker use on students’ content knowledge shows the complex nature of clicker implementation and its possible relation to student learning. The diversity of results related to student content learning suggests the importance of better understanding clicker implementation. This research also shows that the impact of clicker use has been fairly limited to measures of students’ content knowledge neglecting other possible improvements in students’ ability to communicate or generate scientific explanations. Research into the impacts of grading incentives on student discussion practices during PI demonstrates how professors’ instructional practices can influence
how students engage in the course. In this research study, we will investigate additional instructional practices that influence student engagement and perceptions. Researchers have devised a variety of constructs and associated assessment instruments to better understand students’ own perspectives on their experiences learning science (and physics more specifically). These instruments include: the Maryland Physics Expectations (MPEX) Survey [21], the Colorado Learning Attitudes about Science Survey (CLASS) [22], clicker-specific surveys, and typical midterm or end-of-term course evaluations. These instruments all attempt to build an understanding of the meaning that students are making of physics instruction. Here we focus on reviewing clicker-specific surveys which are directly related to the measures of students’ perceptions that will be presented later in this chapter.

A. Impacts of Clicker Use on Students’ Subject-specific Content Knowledge

As recent literature reviews of clicker use (broadly defined) in higher education across a variety of disciplines have stated [23,24,25,26,27], there is contradictory evidence concerning the impacts of clicker use on student learning across a variety of disciplines. While multiple studies have shown statistically significant improvements in students’ course grades or examination scores through the use of clickers in their courses [28,29,30,31], others have shown no statistically significant improvement [32,33,34,35,36,37,38,39]. Studies show mixed results on the impact of clicker use on student learning. It is important to note that clicker use is very broadly defined in the literature reviewed in these articles and there is limited use of validated instruments to measure student learning.

However in the context of Peer Instruction, the specific way in which clickers are used is more specified and student learning has been more systematically studied through the use of validated instruments. Research studies correlating effects of PI with student learning show more consistent and promising implications. Crouch and Mazur demonstrated that the use of PI in both their algebra-based and calculus-based physics courses at Harvard University resulted in improved Force Concept Inventory [40] performance, in addition to improvement on other measures [41]. Fagen et al. also showed that the conceptual learning gains documented at Harvard were found to be consistent with the learning gains documented across a range of institutions and courses [42]. It is important to note that in Fagen’s study a specific criterion was applied to identify courses that implemented PI [Ref. 43, pg. 18].

The literature suggests a number of possible explanations for the conflicting research results regarding the impact of clicker use on student learning: 1) The rarity of valid and reliable measures of student learning across a variety of disciplines [25], 2) When comparing clicker and non-clicker courses, many additional pedagogical changes were made in the clicker courses that were not simultaneously changed in the non-clicker courses [26], and 3) The diversity of ways in which clickers were implemented in the experimental condition (clicker) classrooms [23]. The variation in student learning outcomes through the use of clickers is not well understood. In a review by Judson and Sawada, they claim however, that, “The only positive effects upon student academic achievement, related to the incorporation of electronic response systems in instruction, occurred when students communicated actively to help one another understand” [27]. This suggests that student discussion may be a critical element of PI. We use the term Peer Instruction (PI) loosely to describe what faculty at CU are doing with clickers, but when we are talking about particular actions associated with implementing PI we will often say ‘clicker use” since these are decisions about or elements of clicker use that may be applicable more broadly.
B. Effects of Grading Incentives on Student Discussion Practices

Recently several research articles have described the effects of grading incentives on student discussion practices during Peer Instruction [44,45,46,47]. We summarize the first two of these studies since they clearly link grading incentives to specific student discussion behaviors.

James studied how the assessment practices relating to CQs influenced the nature of conversations and degree of participation that occurred during PI [44]. This study was conducted in two astronomy courses for first year non-science majors taught by two different instructors using 3-5 questions per lecture of their own design. One instructor (in the high stakes course) had CQs count for 12.5% of the students’ grade, where incorrect responses were awarded one-third the credit earned for the correct response. The other instructor (in the low stakes course) had CQs count for 20% of the students’ grade and incorrect responses earned as much credit as correct responses. CQs were discussed amongst pairs of students in both courses.

Conversations between 12-14 pairs of students were recorded in three different classes of each course. These conversations were analyzed for ‘conversation bias,’ the difference between the fraction of all statements made by one partner and the fraction of all statements made by the other partner. If each person made the same number of statements, the conversation bias would be zero. The researchers found that in the high stakes classroom there was greater conversation bias and the students with more knowledge (indicated by higher course grade) tended to dominate the peer discussions. Students in the high stakes classroom were also more likely to vote in the same way as the other member of their discussion pair (only disagreed 8% of the time), whereas in the low stakes classroom student pairs disagreed quite often (37% of the time). They conclude that when there is a grading incentive that strongly favors the correct responses to CQs, the question response statistics may exaggerate the degree of understanding that actually exists and confound the ability of the instructor to make accurate pedagogical decisions based on student response feedback [44].

In a follow-up study, James et al. followed the same instructor in the same environment and studied the impact of only changing grading incentives [45]. They found that with low-stakes grading the conversation bias was reduced and that the fraction of CQs where the discussion pairs disagreed in their responses went up [45]. These results confirmed that grading incentives had significant influence on these student discussion practices. This study suggests that course grading practices have clear effects on student-student collaboration. This study provides an example of how concrete instructional practices can influence how students engage in PI activities associated with the course.

C. Students’ Perceptions of the Use of Clickers

In Judson and Sawada’s recent survey of clicker research [27], they compare assessments of early use of clickers from the 1960’s and 1970’s when use tended to be informed by a behaviorist pedagogical orientation, to more recent assessments of clickers from the 1990’s when use tended to be informed from more of a constructivist pedagogical orientation. Based on students’ positive perceptions of clickers documented in multiple research studies [37,38,39,48,49,50,51,52,53] across a variety of implementations, Judson and Sawada conclude, “Students will favor the use of electronic response systems no matter the nature of the underlying pedagogy” [Ref. 27, pg. 177].

The aim of more recent research studies has been to characterize students’ experience and engagement with clickers along particular dimensions to better understand why students find
clickers useful and enjoyable [54,55,56,57,58,59,60]. Some of these research studies have developed more sophisticated and robust instruments to measure students’ perceptions rather than relying on single or few item survey instruments [59]. Currently, there are calls for additional research to specifically examine students’ perceptions across varied pedagogical approaches and across varied student demographics [26]. A few recent research studies have begun to characterize whether students’ perceptions of clickers vary based on characteristics of the students themselves [55,59]. We summarize two of these more recent detailed examinations of students’ perceptions.

Graham et al. conducted a research study based on student survey responses aimed at answering the following research question: How does clicker use impact students who are least inclined to participate in class? [55]. This research study included 11 different instructors and a total of 688 students enrolled in nine different courses in the fields of chemistry, biology, physics, psychology, education, statistics, and marriage, family and human development. Graham et al. identified three different ways of identifying students as “at-risk for low participation” based on students’ self-reported orientation towards classroom learning in the survey: 1) students reluctant to share opinions in class, 2) students hesitant to ask questions in class, and 3) students who do not prefer courses where there is student participation. Reluctant students of type 1 and type 2 did not perceive clickers differently than their non-reluctant counterparts. Reluctant participants of type 3 were less likely than non-reluctant participants to view clickers as a helpful tool in the classroom. By also examining students’ perceptions by course, the researchers conclude that differences in clicker pedagogical practices may have a larger effect on students’ perceptions of its helpfulness than any of the psychological characteristics of the students [55].

MacGeorge et al. have developed a new instrument to measure students’ perceptions of clickers across multiple dimensions called the Audience Response Technology Questionnaire (ART-Q) [59]. Using their newly developed instrument, they surveyed students in three large (N>200) introductory courses (Communication, Forestry & Natural Resources, and Organizational Leadership & Supervision). Demographic characteristics (gender, ethnicity, and year in school) of students were not found to significantly correlate with student perceptions of clickers [59]. Similarly, in the forestry course, which had a variety of different majors from different schools on campus, there were no differences in students’ perceptions of clickers by school/major. The researchers call for subsequent work to examine whether students’ perceived benefit of clickers can be increased or lost if clickers are used in specific ways [59].

These wide-ranging lines of research into Peer Instruction all point to the need to investigate possible associations between PI implementation and impacts on students [26,61]. Building on prior research into student learning outcomes, student discussion behaviors, and students’ perceptions, we examine three introductory physics courses which all include substantial student discussion during PI activities. We examine correlations between clicker implementation practices and students’ perceptions across three courses. We are less concerned about differences in the student body of these courses because of the lack of dependence of students’ perceptions of clickers on student demographic characteristics that has been established in the literature. There are limited studies based on systematic classroom observations with two notable exceptions [50,62], and the research studies that have examined classroom practices have not compared a variety of implementations and associated impacts on students. This current work addresses this gap in the research. In this research study, we describe in detail how PI was
implemented across three classrooms and show that students perceive norms of clicker use in these classrooms differently.

V. Description of Setting and Methods
A. Description of Courses and Instructors’ Background

All courses observed in this study were large enrollment introductory undergraduate physics courses with average lecture attendance ranging from 130 to 230 students (See Table G). In the previous chapter [2], we presented case studies of six different physics faculty members implementing Peer Instruction. Each of these instructors was assigned pseudonyms to assure the professors’ anonymity: Yellow, Green, Blue, Purple, Red, and White. Complete descriptions of the characteristics of these six instructors can be found in Ref. 2. Here, we consider only three of the instructors and their associated courses: Yellow, Green, and Red. We are interested in comparing students’ perceptions of PI in these environments, and we focus on these three instructors since they were all associated with teaching the introductory calculus-based physics sequence. Thus, the student populations in these three courses are fairly similar, although Phys3 tends to have a larger portion of engineering students. Professor Yellow taught the first semester introductory mechanics course (Phys1). Professor Green taught the second semester introductory electricity and magnetism course (Phys 2). Professor Red taught the third semester introductory modern physics course (Phys 3). Both Yellow and Red are tenured professors while Green was a temporary visiting instructor. Green was a novice with respect to the use of PI and clickers in large-enrollment courses as well as inexperienced at teaching large enrollment courses. Of these three instructors, only Red is an active member of the Physics Education Research group at CU.

Table G: Course Characteristics.

<table>
<thead>
<tr>
<th>Course</th>
<th>Number of students enrolled in the course</th>
<th>Number of Lecture Sections</th>
<th>Average Number of Students Attending a Lecture Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phys1 (Yellow)</td>
<td>566</td>
<td>2</td>
<td>212 ± 5</td>
</tr>
<tr>
<td>Phys2 (Green)</td>
<td>339</td>
<td>1</td>
<td>227 ± 5</td>
</tr>
<tr>
<td>Phys3 (Red)</td>
<td>156</td>
<td>1</td>
<td>132 ± 2</td>
</tr>
</tbody>
</table>

B. Mixed-Methodology Research Study Design

In order to compare how Peer Instruction was implemented and to understand how participants made sense of PI in multiple classrooms, we used a mixed-methodology research design (as previously discussed in Chapter 2). A researcher conducted extensive ethnographic observations in each classroom to document observable classroom practices involved in implementing PI. These observations focused primarily on the practices of the instructor and the interactions between the instructor and the students and secondarily on the practices of the students. Audio recordings of these lecture classes were also collected. In addition to the collection of classroom documents (lecture notes, course syllabi, CQ data, etc) and field notes, the researchers also collected survey responses from students enrolled in each of these courses. Survey data are used to compare students’ perceptions of PI across different classrooms. We briefly describe the data collection of ethnographic observations, the student survey administration, and the quantitative methods used to analyze the student survey responses.
1. Ethnographic methods

Ethnographic research methods [63,64] were used for this study. Through extensive engagement of the researcher within a community, ethnographic research aims to build models based on both insider and outsider perspectives. The researcher’s goal is to make explicit models of the implicit meaning systems of the community. These models are influenced by the insiders’ interpretations of events, but -- due to the researcher’s ability to both bring an outside perspective and a reflective approach to the system -- the researcher is in a unique position to identify overarching patterns that can give sense to the patterns of the community. In this way, the researcher’s interpretation of the values, unspoken in the practices of the community, is important and can help to make explicit cultural elements that are largely implicit to the participants [65].

As discussed above in the theoretical framework section, social groups that are engaged in extended work together develop patterns of social relationships or culture. In order to understand the classroom cultures within which PI was embedded, we first collected descriptive narrative field notes which captured the professor’s actions, interactions between students, and interaction between students and the professor. These preliminary narrative field notes guided the design of an observational rubric for better capturing aggregate patterns of interaction in the classroom (for more details about this process please see Ref. 2).

As discussed in Chapter 2, an observation rubric which was developed in the early stages of this research study was subsequently used to collect data in an additional 10 class periods which constituted at least 20% of the class periods for the course. Observations of these courses captured between 29-51 CQs for each course (depending on the frequency of CQs per class period). For other claims about classroom practices, all CQs were analyzed using data from the clicker software program such as percent of students getting the question correct, etc. Unusual class periods, such as exam review days, or exam days, were excluded from our analysis.

Professors were asked to participate in the research study at the beginning of the semester and the observing researcher dropped into classes unannounced throughout the semester and sat in different areas of the classroom amongst the students. In many instances the professor was unaware of the researcher’s presence. On many occasions the researcher was also treated as a student by the surrounding students enrolled in the course (students would at times turn the researcher to discuss the CQ). These interactions as well as the low visibility of the observing researcher to the professor make us fairly confident that the presence of the research did not dramatically alter the environment begin observed (although some influence is unavoidable).

Due to the research questions and structure of data collection in this study, we have not studied student-student interactions in detail. A small number of student discussions were observed and recorded, but no recordings of student discussions throughout the classroom were collected. This approach restricts our ability to characterize student-student interaction and associated student responsibilities that were present in each of the classrooms.

The reliability of the in-person observations was checked through audio-recordings that were collected during the observations. These audio-recordings allowed the researcher to revisit CQ episodes. A subset of CQ episodes were transcribed to allow the researcher to reflect on and reconsider interpretations of CQ episodes after field notes were collected. Additionally, reliability studies were also conducted on the observation rubric with two other researchers (one from the Colorado PER group and one from an outside institution) who were trained on its use solely through reading the associated User’s Guide (available in Ref. 2). A reliability of 80% or greater was achieved on all items, averaging 96% agreement overall.
Following from an ethnographic tradition, the data sources for this study are both qualitative and quantitative, including: interviews with each professor (conducted at the end of the term), audio recordings of a subset of lecture periods, daily electronic records of CQs asked along with student responses, broad descriptive field notes, focused observational field notes through the use of a rubric [2], and student survey responses surrounding clicker use and classroom dynamics. For further discussion of the methods used for collecting and analyzing observational field notes, please see the discussion contained in Chapter 2 or Ref. 2.

2. Survey methodology

During the last two weeks of the semester, students in each of the classrooms studied were asked to complete an optional online survey where the students’ identity and associated responses would not be available to the instructor (i.e. only aggregate responses would be seen by the instructor). The survey was announced on two different days of class as well as posted at the top of the class website for each course. Students were awarded a small amount of extra credit points for completing the survey (affecting less than 1% of any student’s grade). The surveys were available online for approximately two weeks.

Survey questions were designed by the authors to target broad perceptions of the utility and enjoyment associated with PI and specific perceptions of a subset of three classroom norms associated with the use of PI. Some of these survey questions were designed based on typical end-of-course evaluations probing students’ broad perceptions of the utility and enjoyment associated with PI. Additional survey questions were designed by the authors to target a subset of student thinking about faculty-student collaboration, student-student collaboration, and the emphasis on sense-making. Each question was worded to inquire about PI or clicker use in a particular course specifically using student language found on prior long answer responses from prior semesters of survey data. The resulting survey included a mix of multiple-choice questions and open-ended short answer questions: approximately 25 Likert-scale items and about 5 open-response questions. For the complete survey used to collect data on students’ perceptions of PI, see Chapter 3, Appendix A.

Considering this was an optional survey, the response rates across the three courses were moderately high. The percent of students enrolled in each course that completed the survey are 57% of Yellow’s students (Nrespondents=323), 45% of Green’s students (Nrespondents=153) and 58% of Red’s students (Nrespondents=91). These moderate response rates increase the possibility that survey respondents represent a non-random sample and this adds additional uncertainty to the data. Past analyses of student surveys administered at CU have shown that students who complete the surveys tend to receive slightly higher grades in the course on average than students who do not complete the surveys [66]. Results that are found may only well-represent the higher performing students and care must be taken in generalizing to the entire population.

3. Methods for statistical analysis of student survey responses

A variety of statistical methods are used to identify course-by-course differences in student responses on Peer Instruction-specific survey questions. For survey questions that are categorical, but not rank ordered, a Chi-squared Test for Independence [Ref. 67, p. 204] is used to identify statistically significant variations across courses [68]. For survey questions which are categorical and rank ordered, a Kruskal-Wallis Test for k-independent samples [Ref. 67, p. 288] is used to identify statistically significant variations across courses [69]. A summary of the statistical test used for each question is given in Appendix B. When comparing the set of all
courses, p-values from 0.10 > p > 0.05 are taken to be marginally significant and p-values less than 0.05 are taken to be significant. On the questions which show significance or marginal significance, we compare semesters in a pair-wise fashion (using either a Chi-squared Test for Independence or a Mann-Whitney U Test [Ref. 67, p. 272], depending on whether the answer options are rank-ordered or not). When determining the statistical significance of the pair-wise comparisons, we will decrease the threshold for statistical significance depending on the number of pair-wise semester comparisons that are being made for that particular question. The resultant threshold values for significance and marginal significance are summarized by question in Appendix B.

C. Study Limitations

As mentioned above, there are multiple limitations to this study. With respect to the ethnographic data collection, the possible effects due to the presence of the observing researcher cannot be fully understood, since video-recordings of the classes were not collected. We also note that the make-up of student majors enrolled in these courses is not the same (with the Phys3 course having a larger portion of engineering majors). Although this might be a factor, prior research has shown no significant correlation between students’ perceptions of clicker use in a course and their declared major [59]. With respect to the survey methodology, the survey response rates pose a concern. Without all students taking the survey, the survey respondents represent a non-random sample of the student population enrolled in the course. This could skew our results. Lastly, extensive validity and reliability studies have not been conducted on the survey used in this study. Unfortunately, another relevant instrument, the ART-Q survey instrument did not exist when this study was conducted [59]. However, the survey questions asked in this survey were designed using language found in students’ long answer responses during prior terms. So although we cannot be certain that students are interpreting these questions consistently, we believe that the questions are phrased using common student language. These concerns inject some uncertainty to our claims and suggest that research studies aimed at replicating these findings should be conducted.

V. Data & Results: Comparing Three Introductory Physics Courses Along Three Norms, Students Perceive Differences

A. Norm 1: Faculty-Student Collaboration

1. Defining the norm of faculty-student collaboration

The classroom norm of faculty-student collaboration is of particular interest because it is a prevalent method of scaffolding student learning and a common mode of feedback between faculty and students (formative assessment [70,71,72]). Understanding faculty-student collaboration requires answering the following questions: What types of faculty-student interaction are happening? How often are different types of faculty-student interaction occurring? What roles and associated responsibilities are available to which classroom participants (faculty or students)? We define “high” faculty-student collaboration to be a classroom where there are many types of faculty-student interaction that occur often during class and in which students and faculty take on and move amongst a diversity of roles in the class. In contrast, we define “low” faculty-student collaboration to be a classroom where there are few types of faculty-student interaction, which occur infrequently and in which students and faculty take on but a few stable roles in the class. These definitions describe the extremes along a
particular dimension of classroom culture. Particular classroom norms are placed along a continuum representing a set of norms along this dimension. With these definitions and our observations of PI implementation practices [2], we can then characterize the classrooms of each of the professors that were observed.

In our prior work, we described observable aspects of faculty-student interactions that may occur during PI. Here we expand on this work, describing multiple types of faculty-student interaction (or modes of participation) and how these types of interactions limit the set of roles (or responsibilities) available to the participants (See Table H). In the prior section III.C., we described two broad modes of participation for faculty-student interactions: A) Student-educator small group discussions in which only a small subset of students have access to this exchange and can therefore be considered semi-private and B) Whole class discussions in which all students have access to this exchange and therefore occur in the public sphere. Within these two broad modes of participation, we detail types of faculty-student interaction and common, associated responsibilities of the faculty and students. The frequencies with which the various types of faculty-student interaction are employed, and the diversity of the types used within a classroom setting, contribute to establishing the norm of faculty-student collaboration.
### Small-Group Mode of Participation

| Type 1 | Faculty and student in close proximity: | When the instructor leaves the stage, the instructor is within earshot of the student conversations and the students can more easily get the instructor’s attention while students are engaged in collectively solving the CQ. Possible instructor responsibilities: listening to students talk, or scanning for students’ bids for attention. Possible student responsibilities: listening to a peer’s description of CQ solutions, criticizing or rebutting a peer’s physics ideas, describing their physics ideas to their peers, working the physics problem on paper, listening or talking about out-of-class topics, or bidding for the professor’s attention. |
| Type 2 | Faculty responds to student questions: | In this type of interaction a student raises his/her hand or calls out to the instructor as he/she passes by and the faculty member usually approaches the student to respond to the student’s inquiry in a more private setting. Possible instructor responsibilities: listening to students’ inquiry, or responding to the students’ inquiry. Possible student responsibilities: getting the instructor’s attention, presenting their question or inquiry, requesting clarification, requesting information, or requesting approval of their physical ideas. |
| Type 3 | Faculty discusses with students: | In this type of interaction, the professor approaches a student or group of students and instigates an interaction with an open-ended inquiring statement or question such as “what are you thinking about?” and engages in a discussion with students about the CQ. Possible instructor responsibilities: posing an open-ended question of students, listening to students’ reasoning, asking additional follow-up questions or clarifying questions, and possibly challenging students’ explanations. Possible student responsibilities: describing their physics ideas to the professor, responding to bids by the professor to elaborate or clarify their physical ideas, defend their physical ideas, modify their physical ideas, and articulate a physical argument. |

### Whole-Class Mode of Participation

| Type 4 | Faculty describes CQ problem: | In this type of interaction, the professor presents a problem for the class to consider. Possible instructor responsibilities: designing the problem to be considered, describing the problem to the students, and linking the current problem to prior material presented in class. Possible student responsibilities: listening to professor’s description of the problem and taking notes on material that is mentioned by the professor during the introduction of the problem. |
| Type 5 | Faculty describes CQ solution: | In this type of interaction, the professor presents a description of the CQ solution. Possible instructor responsibilities: completely describing the CQ solution. Possible student responsibilities: listening to the professor’s explanation and taking notes on the professor’s solution. (We note that it is quite possible that the students are engaged in other cognitive tasks; however, we focus only on the observable actions of students that are made public for feedback from either peers or the professor.) |
| Type 6 | Faculty and student(s) describe CQ solution: | In this type of interaction, the faculty and a student share responsibility for describing the CQ solution publicly, but the faculty member retains the majority of the responsibility for evaluating the CQ solution as well as designating speakers’ turns. Possible instructor responsibilities: requesting student explanations, nominating a student for contributing an explanation, listening to the student’s explanation, asking clarifying questions of the student, evaluating the student’s explanation, and offering a revised expert explanation. Possible student responsibilities: listening to other students’ explanations, offering their own explanations of the CQ solution, listening to the professor’s explanation, and taking notes on the professor’s solution. |
| Type 7 | Faculty and student(s) jointly describe and evaluate the CQ solution: | In this type of interaction the faculty member and students share responsibility for publicly describing and evaluating explanations of CQ solutions. A common rough indicator of this type of interaction is the inclusion of multiple students’ perspectives in the public description of the CQ solution, since it is in these occasions that students participate to some degree in evaluating CQ explanations. Possible instructor responsibilities: requesting student explanations, nominating students for contributing explanations, listening to the students’ explanation, asking clarifying questions of the students,
requesting other students to comments on their peers’ physics thinking, driving consensus in students’ explanations, and offering a short summary of an expert explanation. Possible student responsibilities: listening to students’ explanations, offering an explanation of one’s CQ answer, commenting on their peers’ physics thinking, debating/disagreeing with other students’ reasoning, listening to the professor’s explanation, and taking notes on the solution discussion.

Interaction Types 5-7 are mutually exclusive categories in our coding system. Any whole-class discussion of a given CQ solution was found to fall into only one of these categories.

These different types of interactions are important because they constrain which responsibilities fall to the shoulders of different participants. They are also important in that they create different resources which are then available for use by the participants during on-going classroom activities. For example, when a faculty member approaches a group of students discussing a physics concept, the faculty member has an opportunity to learn something new about students’ thinking that he/she was not previously aware of. This knowledge may simply inform how the instructor decides to proceed during the class period, or this talk can become a resource to be used in the immediate unfolding of the particular PI episode. The professor may privately ask the students if they would mind sharing their reasoning with the whole class. If the students agree, the professor now has an additional resource to use in leading a public discussion of the solution. We will return to discuss these emergent resources in the presentation of the final norm, emphasis on sense-making vs. answer-making.

2. Comparing three introductory physics courses based on implementation practices

Here, we examine variation in classroom practices with respect to the norm of faculty-student collaboration. We begin by describing typical interactions in a traditional lecture format course and the constraints that these interactions place on the responsibilities of the educator and the students. In a traditional lecture format, the professor is found at the front of the room -- usually in a clearly demarcated stage area -- where he/she controls the presentation of information and who is allowed to speak. In this format there are few permitted moves that allow for faculty and students to negotiate the meaning behind physical ideas. The responsibilities of both professor and students are rigidly set. The responsibilities of the professor include organizing the physics content, describing the physical ideas, and working example physics problems. The responsibilities of the students are limited to listening to the professor and taking notes on what the professor says or writes on the board. Student questions of the faculty are fairly infrequent and often clarifying in nature.

We place the traditional lecture format at low faculty-student collaboration. This is because there is only one primary type of faculty-student interaction and all other types of faculty-student collaboration occur infrequently, if ever. This means that the roles and responsibilities of the educator and the students are very clearly defined and the responsibilities that students take on do not change often from class to class or over the course of the semester. The location of a traditional lecture format is shown leftmost in Figure 13.

Low Faculty-Student Collaboration

High Faculty-Student Collaboration

Traditional Lecture

Figure 13: (Color) Continuum for the Set of Faculty-Student Collaboration Norms.
In order to describe the placement of Yellow, Green, and Red along this continuum, we code the frequency with which each type of faculty-student interaction occurred. A summary of this analysis can be found in Figure 14.

![Figure 14: (Color) Percent of clicker questions where the professor was observed to engage in each type of faculty-student interaction (Types of interactions are defined in Table A). The dashed line denotes that Types 5-7 at the rightmost side of graph are mutually exclusive categories.](image-url)

In Yellow’s class, the professor rarely leaves the stage (only 12% of CQs Type 1), rarely answers student questions (only 19% of CQs Type 2), rarely discusses with the students (only 8% of CQs Type 3), and rarely hears student explanations publicly (only 17% of CQs either of Type 6 or 7). Two percent of whole-class CQ explanations include the contribution of only one student, Type 6 as described in Table H. Fifteen percent of whole-class CQ explanations include the contributions of two or more students (Type 7). This means that the professor is usually conducting the description of the CQ solution entirely on his own (Type 5). Since faculty-student interactions Types 1-3 as described in Table H occur infrequently in Yellow’s class, we see that there are few opportunities for Yellow’s students to interact with the instructor within a small-group mode of participation. Similarly, since Type 5 is the highly prevalent mode of participation during the whole-class solution discussion, the students and faculty have fairly stable roles where students are rarely given responsibility over the solution description or the evaluation of the proposed solution description. These aspects of faculty practice provide evidence to support the placement of Yellow’s classroom at the lower end of the faculty-student collaboration continuum.
In Green’s class, the professor rarely leaves the stage (only 11% of CQs), rarely answers student questions (only 25% of CQs), never discusses with the students (0% of CQs), and always hears student explanations publicly (100% of CQs). Green usually hears from only one student during the whole-class public explanation of the CQ solution (67% of CQs). This student has usually offered a correct explanation. Green is usually quick to reveal correctness of student explanations. Since faculty-student interactions Types 1-3 occur infrequently in Green’s class, there are few opportunities for Green’s students to interact with the instructor within a small-group mode of participation. Similarly, since Type 6 is the dominant mode of participation during the whole-class solution discussion, the students and faculty have fairly stable roles where students are often allowed some responsibility over the solution description, but little responsibility over the evaluation of the proposed solution. Since the majority of CQ solution descriptions occur between the faculty and a single student, students rarely have opportunities to comment on or disagree with the physical reasoning presented. These aspects of faculty practice provide evidence to support the placement of Green’s classroom at the lower end of the faculty-student collaboration continuum.

In Red’s class, the professor usually leaves the stage (69% of CQs), usually answers student questions (63% of CQs), usually discusses with the students (84% of CQs), and often hears student explanations publicly (55% of CQs). Twenty-three percent of whole-class CQ explanations include the contribution of only one student. Thirty-two percent of whole-class CQ explanations include the contributions of two or more students. When Red requests student explanations, Red often hears from multiple students. Public debate and disagreement is supported and encouraged during these occasions. Red often withholds expert evaluation of answer correctness until consensus develops. Since faculty-student interactions Types 1-3 occur frequently in Red’s class, we claim that there are many opportunities for Red’s students to interact with the instructor within a small-group mode of participation. In Red’s class, each of Types 5-7 occur fairly often. During the whole-class solution discussion, the students and faculty have fairly flexible roles where students have a varying degree of responsibility over the solution description and the evaluation of the proposed solution depending on the CQ. These aspects of faculty practice provide evidence to support the placement of Red’s classroom at the higher end of the faculty-student collaboration continuum.

We therefore posit that student participants in Red’s course are more likely to perceive high levels of faculty-student collaboration during CQs than students from Yellow’s course and Green’s course. We proceeded to design four survey questions to elicit a subset of students’ ideas related to faculty-student collaboration during Peer Instruction.

3. Comparing three introductory physics courses based on students’ perceptions

To understand whether students perceived there to be relatively high or low value on faculty-student collaboration, we asked them four specific survey items that probed a subset of this classroom norm. For example, the first statement was: It is awkward to ask my professor questions during class. The students were given five answer options: Strongly Disagree, Somewhat Disagree, Not Sure, Somewhat Agree, and Strongly Agree. The distribution of student responses on this question is shown below in Figure 15.
From this plot, we can see that students from Red’s class are less likely than Green and Yellow’s students to think that it is awkward to ask the professor questions. In Red’s course we see that twice as many students are choosing disagree (somewhat disagree/strongly disagree) than are choosing agree (somewhat agree/strongly agree). Red’s students tend to choose higher (more favorable) categories than Green and Yellow’s on this question. We see in Table I that these differences are statistically significant.

We investigate whether this trend persists across an additional three survey questions. Question 16 asked, “How often do you raise your hand or ask questions in class?” (Never, About once a semester, About once a month, Nearly every week, and Nearly every class). Question 10 asked, “If my professor were to approach me in class during a clicker question, I would be comfortable discussing the content with my professor.” (Strongly Disagree, Somewhat Disagree, Not Sure, Somewhat Agree, and Strongly Agree). Question 12 asked, “How often do you speak directly to the professor during class?” (Never, Once or twice a semester, Once every few weeks, Nearly every week, and Nearly every class). All of these questions have answer options that are categorical and rank ordered. Since there were statistically significant differences between these three groups via Kruskal-Wallis Test on all questions, the courses are compared pair-wise using the Mann-Whitney U Test. The results of the pair-wise comparisons are provided in Table I.
Table I: (Color) Mann-Whitney U Test Results for Pair-wise Comparisons on Faculty-Student Collaboration Survey Questions. The arrows in the table indicate which group tends to yield higher (more favorable) responses for all p-values less than 0.1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yellow-Red</th>
<th>Green-Red</th>
<th>Yellow-Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q11 Awkwardness Asking Professor Questions</td>
<td>R p&lt;0.001*</td>
<td>R p=0.002*</td>
<td>p=0.629</td>
</tr>
<tr>
<td>Q16 Frequency Asking the Professor Questions</td>
<td>R p&lt;0.001*</td>
<td>R p&lt;0.001*</td>
<td>G p=0.032</td>
</tr>
<tr>
<td>Q10 Comfort Discussing with Professor</td>
<td>R p=0.033</td>
<td>R p=0.001*</td>
<td>Y p=0.034</td>
</tr>
<tr>
<td>Q12 Frequency Speaking with Professor</td>
<td>R p&lt;0.001*</td>
<td>R p&lt;0.001*</td>
<td>G p=0.024</td>
</tr>
</tbody>
</table>

We see from Table I that Green and Red differ significantly on all four questions, with Red’s students reporting more favorably on all questions. We also see that Yellow and Red vary significantly on three of the four questions, with Red’s students reporting more favorably on all questions. We do not find statistically significant differences in students’ perceptions of faculty-student collaboration between Yellow and Green’s courses. We conclude that Red’s students perceive there to be a higher value placed on faculty-student collaboration than do both Yellow and Green’s students. Yellow and Green’s students perceive faculty-student collaboration similarly.

These four survey questions show that students from these different courses notice some of the same differences in classroom norms as the research observer. The correspondence between the observing researcher’s interpretation of the norm of faculty-student collaboration and students’ perceptions regarding a subset of this norm supports the claim that concrete classroom practices associated with the implementation of PI can impact the way that students perceive faculty-student collaboration.

B. Norm 2: Student-Student Collaboration
1. Defining the norm of student-student collaboration

The classroom norm of student-student collaboration is of particular interest because it is a highly prevalent method of promoting active engagement [73] and a common strategy for making students’ thinking visible to the educator and other students allowing for formative assessment [70,71,72]. Understanding student-student collaboration requires answering the following questions: What types of student-student interactions are happening? How often are different types of student-student interaction occurring? What roles and associated responsibilities are available to students? We have not studied student-student interactions in detail in this study (as discussed in the methods section). We examine student-student collaboration by looking for faculty practices that constrain or allow for different amounts and kinds of student-student collaboration. We focus on the faculty practices of grading incentives, fraction of CQs where peer collaboration was allowed, prevalence of introductory comments encouraging peer collaboration, fraction of class time spent in peer collaboration, and the frequency and type of opportunities for faculty to model discussion practices with students both in small group and whole class modes of participation.
We define “high” student-student collaboration to be a classroom where there are low-stakes grading practices, significant opportunities for peer collaboration (i.e. peer collaboration usually allowed, relatively long CQ voting time provided, and relatively high percent of class time devoted to explicit peer collaboration), consistent explicit encouragement of peer collaboration (i.e. introductory comments supporting or encouraging talk amongst peers), and frequent opportunities for the instructor to model scientific discourse (i.e. prevalence of faculty answering questions and discussing with peers in small group interactions, and Type 6&7 interactions during whole class discussion). In contrast, we define “low” student-student collaboration to be a classroom where there are high stakes grading practices, few opportunities for peer collaboration (i.e. short length of CQ voting time, and low percent of class time devoted to explicit peer collaboration), little explicit encouragement of peer collaboration (i.e. few introductory comments supporting or encouraging talk amongst peers), and few opportunities for the instructor to model scientific discussion practices (i.e. few occasions of faculty answering questions and discussing with peers in small group interactions, and Type 6&7 of whole class discussion). These definitions describe the extremes along a particular dimension of classroom culture. Particular classroom norms are placed along a continuum representing a set of norms along this dimension. With these definitions and our observations of PI implementation practices [2], we can then characterize the classrooms of each of the professors that were observed.

In the previous section, we defined types of interactions between the instructor and students (see Table H). In small-group modes of participation, the instructor has opportunities to model discussion practices with students primarily in Types 2 & 3. In the whole-class participation format, the instructor has opportunities to model discussion practices during Types 6 & 7. In Type 6 interactions, the professor can model how one might ask clarification questions or ask a student for additional details about his or her thinking, modeling the kinds of follow-up questions students might ask each other during their discussions of their reasoning. However, with only one student sharing in the public discussion, this interaction format does not allow the instructor to model discussion practices that arise around disagreement or dispute. In Type 7 interactions, there is potential to model these discussion practices surrounding disagreement and dispute.

One important distinction within Type 7 interactions that is particularly relevant for modeling scientific discussion practices surrounding disagreement. All Type 7 interactions have some degree of student involvement in evaluating the public solution, but to varying extents. We found that in some classes, even though multiple student explanations were being heard in the public forum, these exchanges took place sequentially with the instructor interacting with a single student and then the professor interacting with another student and so on. In such instances, students were not explicitly commenting on the prior explanations of their peers (Type 7a—Instructor as non-mediator, no explicit student crosstalk). In this way, the instructor was not mediating student collaboration. In these situations, there is some implicit evaluation being done by students since the second student must evaluate whether what he/she has to say is significantly different enough from the perspective described by the prior student to justify describing his/her own thinking on the topic.

In other instances, an instructor was observed to support student collaboration and disagreement through playing a mediating role (Type 7b—Instructor as mediator, explicit student crosstalk). In these instances the professor would specifically position students to comment on or even actively disagree with their peers in a respectful way. For example, the professor would
say, “Would anyone like to retort?” after the first student explanation was heard. In this way the professor was opening the discussion for additional student disagreement and supporting student collaboration rather than closing the discussion. Type 7b supports the modeling of scientific discussion practices in a way that Type 7a does not. In these situations, there is significant explicit evaluation being done by students since they were found to be actively commenting on the physical reasoning presented by their peers.

2. Comparing three introductory physics courses based on implementation practices

Here, we examine variation in classroom practices with respect to the norm of student-student collaboration. We begin by characterizing student-student collaboration in a traditional lecture format. In a traditional lecture format course, there are usually few to no opportunities for students to test out their thinking without being graded. There are also few to no opportunities for students to talk with other students during the lecture; therefore, if peer collaboration is explicitly encouraged at all, it is expected to occur outside of the regular lecture time. Instructors rarely have opportunities to model discussion practices in traditional lecture courses since the instructor usually does all or most of the talking. We indicate that traditional lecture format would be placed at low student-student collaboration. On this continuum none of the instructors are at the high end of student-student collaboration because even in the class that is most encouraging of student-student collaboration, only approximately a third of class time is spent engaged in student-centered activities.

Yellow’s class had moderate-stakes grading practices because there was some evaluative emphasis placed on correctness, but only for awarding extra credit. In this way correct answers were awarded more extra credit points than incorrect answers, but incorrect answers receive some partial extra credit. In Yellow’s class, peer collaboration was allowed for all CQs, students were given about three minutes on average to complete a CQ with their peers (with 56% of CQs lasting more than 2 minutes), and approximately 30 percent of class time was explicitly devoted to peer collaboration. Based on these criteria, a moderate amount of opportunities for peer collaboration were available in Yellow’s class. Yellow began about half of his CQs with introductory remarks such as “Talk to your neighbor” or “Ask your neighbor.” With these fairly consistent introductory remarks, Yellow was providing explicit encouragement of peer collaboration in his class. In Yellow’s class there were few opportunities for the instructor to model scientific discussion practices both in the context of small group interactions and whole-class discussions. Yellow rarely created opportunities to interact with students in small groups (left the stage 12% of the time, answered student questions 19% of the time, and discussed with students 8% of the time). Similarly there were rarely opportunities to model discussion practices in the whole-class discussion, because student explanations were rarely heard (17% of the time). To summarize, Yellow’s course had moderate-stakes grading practices, moderate opportunities
for peer collaboration, and fairly consistent explicit encouragement of peer collaboration during PI, but few opportunities for the instructor to model scientific discussion practices in either small group or whole-class formats. These four aspects of faculty practice provide evidence to support the placement of Yellow’s classroom at the lower end of the student-student collaboration continuum.

Green’s class had moderate-stakes grading practices because there was some evaluative emphasis placed on correctness, but, as with Yellow, only for awarding extra credit. In Green’s class, peer collaboration was allowed for all CQs, students were given about three and a quarter minutes on average to complete a CQ with their peers (with 64% of CQs lasting more than 2 minutes), and approximately 17 percent of class time was explicitly devoted to peer collaboration. Green’s students and Yellow’s students have similar amounts of time to respond to CQs, however slightly more class time is spent on CQs in Yellow’s class. Based on these criteria, significant opportunities for peer collaboration were available in Green’s class, as in Yellow’s class. Green usually read the CQ out loud elaborating on what the diagrams represented, but rarely made introductory comments explicitly encouraging peer collaboration. Green made fewer remarks encouraging peer collaboration as compared to Yellow. In Green’s class (as in Yellow’s) there were few opportunities for the instructor to model scientific discussion practices in the context of small group interactions. Green rarely created opportunities to interact with students in a small group format (left the stage 11% of the time, answered student questions 25% of the time, and never discussed with students). In Green’s class there were some opportunities for the instructor to model productive student discussion practices in the context of whole-class discussions because student explanations were always heard. As mentioned in the prior section, Green usually only heard from a single student about his/her reasoning (67% Type 6). When Green did hear from multiple students, he was usually employing Type 7a of student-faculty interaction -- that is, he was generally not mediating or supporting disagreement amongst students. To summarize, Green’s course had moderate-stakes grading practices, moderate opportunities for peer collaboration, but few explicit remarks encouraging peer collaboration, and few opportunities for the instructor to model scientific discourse in a small group format. There were however moderate opportunities for the instructor to model productive student discourse in the whole-class format. These aspects of faculty practice provide evidence to support the placement of Green’s classroom near the lower end along the continuum of student-student collaboration.

Red’s class had fairly low-stakes grading since evaluative emphasis was rarely placed on correctness. For the vast majority of questions, all answer options were awarded an equal number of points independent of correctness. In Red’s class, peer collaboration was allowed for the majority of CQs (~95%), students were given about three and three quarters minutes on average to complete a CQ with their peers (with 55% of CQs lasting more than 2 minutes), and approximately a third of class time was explicitly devoted to peer collaboration. Based on these criteria significant opportunities for peer collaboration were made in Red’s class. Red began about half of his CQs with introductory remarks such as “Go ahead and discuss with your neighbors” or “Get into your groups and work this out.” With these fairly consistent introductory remarks, Red was providing explicit encouragement of peer collaboration in his class. In Red’s class there were frequent opportunities for the instructor to model productive student discussion practices in the context of small group interactions. Red often created opportunities to interact with students in a small group format (left the stage 69% of the time, answered student questions 63% of the time, and discussed with students 84% of the time).
Red’s class there were some opportunities for the instructor to model scientific discussion practices in the context of whole-class discussions because student explanations were usually heard (55% of the time). As mentioned under faculty-student collaboration, Red sometimes only heard from a single student about their reasoning (23% Type 6) and sometimes heard from multiple students (32% Type 7). When Red did hear from multiple students, he was usually playing a mediating role and supported public disagreement amongst students (usually Type 7b). To summarize, Red’s course had fairly low-stakes disagreement practices, moderate opportunities for peer collaboration, and fairly consistent explicit encouragement of peer collaboration during PI. Additionally, there were frequent opportunities for the instructor to model scientific discussion practices in small groups and only moderate opportunities for the instructor to model scientific discourse in the whole-class format. These aspects of faculty practice provide evidence to support the placement of Red’s classroom closer to the high end of the continuum of student-student collaboration.

3. Comparing three introductory physics courses based on students’ perceptions

To understand whether students perceived there to be relatively high or low value placed on student-student collaboration, we asked them four specific survey questions. For example, the first question was: How comfortable do you feel discussing the course content with your peers during clicker questions? (Not comfortable, Somewhat comfortable, Not sure, Somewhat comfortable, and Very comfortable). The distribution of student responses on this question is shown in Figure 17.

Figure 17: (Color) Example Question—Distribution of Student Responses to Comfort level discussing physics with student peers (Q14).

This plot shows that most students from all three classes report being either somewhat comfortable or very comfortable discussing with their peers. However, Red’s students were more likely to report being very comfortable discussing with their peers. Between 15-20% more
students in Red’s course chose “very comfortable” as compared to Yellow’s course and Green’s course.

We test whether Red’s students tend to choose statistically higher (more favorable) responses than Green and Yellow’s across a series of questions about collaborating with their peers. Four other survey questions were asked of the students (Q6a, Q6b, and Q7, and Q15).

Question 6 asked: “To what extent does your instructor usually encourage student-to-student discussion about clicker questions in class?” The students were given answer options that included: 1) Does not allow discussion, 2) Allows discussion, but does not encourage it, and a small fraction of students discuss, 3) Allows discussion, but does not encourage it, and a large fraction of students discuss, 4) Encourages discussion, and a small fraction of students discuss, 5) Encourages discussion, and a large fraction of students discuss. No students chose the first answer option in any of the physics courses studied, so this answer option was eliminated in the analysis. Then the remaining answer options were collapsed in two different ways for analysis. The first grouping, question 6a, was based on whether the instructor allowed discussion (options 2 & 3) or encouraged discussion (options 4 & 5). The second grouping, question 6b, was based on whether a small fraction of students discussed (options 2 & 4) or a large fraction of students discussed (3 & 5).

Question 7 asked: “When your instructor gives your class a typical clicker question and you are allowed to talk with others, what do you usually do? (Does not apply—we are usually not allowed to talk with other students, I rarely use a clicker in this course, I guess the answer and do not check with other students, I actively think about the question independently and arrive at an answer without speaking or listening to other students, I listen to other students’ answers and/or reasoning, and I actively participate in discussions with other students around me). For this question, less than 2% of students chose any of the first three answer options, so these options were deleted. The only answer options that were considered were the last three: independent, listen to others, and discuss.

Question 15 asked: “On average, what fraction of class time do students speak either with each other or to the professor?” The students were given answer options that ranged from less than 5 minutes out of a 50-minute class to more than 20 minutes out of a 50-minute class (with answer options in between in increments of 5 minutes). All of these questions have answer options that are categorical and rank ordered. There were statistically significant differences between these three groups via Kruskal-Wallis Test on all questions, except question 6b. Thus, the level of student discussion perceived by students in each of these three courses was statistically indistinguishable (See Appendix B, Q6b). The courses are compared pair-wise using the Mann-Whitney U Test on all questions, except for 6b. The results of the pair-wise comparisons are provided in Table J.
We see from Table J that Green and Red differ significantly on all four questions, with Red’s students reporting more favorably on all questions. We also see that Yellow and Red vary significantly on two of the four questions, with Red’s students reporting more favorably on three of the four questions. We find that Green and Yellow vary significantly on only one of the four questions, with Yellow’s students reporting more favorably on two of the four questions. We conclude that Red’s students perceive there to be a significantly higher value placed on student-student collaboration than do Green’s students. We conclude that Red’s students perceive there to be a somewhat higher value placed on student-student collaboration than do Yellow’s students. Overall, we do not see many significant differences between how Yellow’s students and Green’s students perceive student-student collaboration.

These five survey questions show that students from these different courses notice some of the same differences in classroom norms as the research observer. The correspondence between the researcher’s interpretation of the norm of student-student collaboration and students’ perceptions regarding a subset of this norm supports the claim that concrete classroom practices can impact the way that students perceive student-student collaboration.

C. Norm 3: Emphasis on Sense-making Versus Answer-making
1. Defining the norm of emphasis on sense-making versus answer-making

Researchers have distinguished two modes of student engagement in classroom science learning: answer-making and sense-making. Answer-making has been defined as a less productive framing of school science activities aiming to get students to know the scientifically accepted answers that scientists have developed to describe the natural world [74,75]. In this mode, students are usually trying to come to the explanation that they think the teacher wants to hear rather than coming to an explanation that makes sense to the student. Sense-making has been defined as a productive framing of school science activities in which the aim is to get students to build sensible and plausible models of the natural world that are intelligible to the students themselves [74,75]. Students as well as instructors play an active role in framing school science activities with one of these emphases. Researchers at the University of Maryland have done significant work characterizing evidence of students using a sense-making frame versus an
answer-making frame in small collaborative group problem-solving environments [75]. Here, we seek to focus on instructional moves, and to characterize sense-making versus answer-making as a norm of Peer Instruction activity.

In the context of Peer Instruction, we define a classroom with an emphasis on sense-making to be one where there are low-stakes grading practices, consistent explicit emphasis of sense-making or reasoning by the instructor (i.e., introductory comments reminding students to not just pick an answer but to be able to defend or explain why they chose that option, and CQ explanations, not just answers, integrated into the lecture materials made available to students), significant opportunities for student discussion of physical reasoning in small-group formats (i.e., relatively long CQ voting time provided and conceptual CQs frequently asked), significant opportunities for student discussion of physical reasoning in whole-class discussions (i.e., student explanations are usually heard in the public discussion of the CQ solution), and frequent opportunities for the instructor to model scientific discourse (i.e., prevalence of discussing incorrect answer options in the whole class discussion, multiple students are usually heard from in the public discussion, and instructor support and management of public disagreement amongst students). We define emphasis on answer-making to be a classroom where there are high-stakes grading practices, little explicit emphasis on sense-making or reasoning by the instructor, few opportunities for student discussion of physical reasoning in small-group formats or whole-class discussions, and few opportunities for the instructor to model scientific discourse. These definitions describe the extremes along a particular dimension of classroom culture. Particular classroom norms are placed along a continuum representing a set of norms along this dimension. With these definitions and our observations of PI implementation practices, we can then characterize the classrooms of each of the professors that were observed.

2. Comparing three introductory physics courses based on implementation practices

To begin with a familiar starting point, we describe the emphasis of sense-making versus answer-making in a traditional lecture format. As mentioned in the prior norm, traditional lecture courses generally use high-stakes grading since there are usually few to no opportunities for students to test out their thinking without potentially affecting their grade. There are also few to no opportunities for students to share their physical reasoning with other students or with the instructor during the lecture; therefore there are few opportunities for student discussion of physical reasoning in either small group or whole class discussion formats. Since the instructor usually does all or most of the talking in a traditional lecture, there is little opportunity for students to test out and receive feedback on their physical reasoning which provides few occasions for students and the instructor to negotiate meaning. For these reasons, a traditional lecture course would be categorized as at the emphasizing answer-making end of the continuum rather than sense-making end of the continuum.

Figure 18: (Color) Continuum for the Set of Answer-making versus Sense-making Norms.
Yellow’s class had moderate-stakes grading practices as described previously. In Yellow’s class, there was little explicit emphasis on sense-making or reasoning by the instructor, either through introductory comments reminding students to explain why they picked their answer or through the lecture materials made available to students. The CQs and associated correct answers were made available to the students but explanations for why these were the correct answer options were not provided. There were many opportunities for students to discuss their physical reasoning in small-group formats since the CQ voting time was relatively long and the majority of CQs were conceptual. However, students had few opportunities to discuss their physical reasoning in whole-class discussions. Student explanations were rarely heard in the public discussion of the CQ solution (only 17% of the time). Since the professor rarely interacted with students in small group formats and rarely heard student explanations in the whole-class format, Yellow had little exposure to student reasoning and limited opportunities to encourage student reasoning. Additionally, incorrect answers to the CQs and associated flaws in reasoning were rarely discussed during the public solution description (only 19% of the time). There were relatively few opportunities for the instructor and the students to negotiate meaning of physical ideas. Thus, Yellow’s classroom is placed towards the answer-making end of this continuum.

Green’s class, like Yellow’s class, had moderate-stakes grading practices. In Green’s class, there was some explicit emphasis on sense-making or reasoning by the instructor. Green, as with Yellow, rarely reminded students to explain their reasoning behind choosing a particular answer when introducing a CQ. However, CQ answers and associated explanations were made available to students in Green’s class. Similar to Yellow, there were many opportunities for students to discuss their physical reasoning in small-group formats since the CQ voting time was relatively long and the majority of CQs were conceptual. In Green’s class, students also had frequent opportunities to discuss their physical reasoning in whole-class discussions. Student explanations were always heard during the public discussion of the CQ solution (100% of the time). Green, as with Yellow, rarely interacted with students in small group formats. However, Green always heard a student explanation in the whole-class format. Green therefore had moderate exposure to students’ reasoning and a few opportunities to encourage student reasoning in face-to-face interactions. In Green’s class, incorrect answers to the CQs and associated flaws in reasoning were rarely discussed during the public solution description (only 25% of the time). There was a moderate amount of opportunities for the instructor and the students to negotiate meaning of physical ideas. These aspects of classroom practice provide evidence to support the placement of Green’s classroom near the answer-making end of this continuum.

Red’s class had fairly low-stakes grading since evaluative emphasis was rarely placed on correctness as described previously. In Red’s class, there was often explicit emphasis on sense-making or reasoning by the instructor. Red often reminded students to explain their reasoning behind choosing a particular answer when introducing a CQ, and often told students that they were going to be asked to offer their explanations to the rest of the class. Additionally, CQ answers and associated explanations were made available to students and integrated into the lecture notes for the class. There were many opportunities for students to discuss their physical reasoning in small-group formats since the CQ voting time was relatively long and the majority of CQs were conceptual. In Red’s class, students also had frequent opportunities to discuss their physical reasoning in whole-class discussions. Student explanations were usually heard during the public discussion of the CQ solution (55% of the time). Red often interacted with students in small group formats (either by answering questions or discussing with students) and he usually
heard student explanations in the whole-class format. Therefore, Red had high amounts of exposure to students’ reasoning and opportunities to encourage student reasoning in face-to-face interactions. In Red’s class, incorrect answers to the CQs and associated flaws in reasoning were often discussed during the public solution description (58% of the time). There were many opportunities for the instructor and the students to negotiate meaning of physical ideas. Red’s classroom is placed towards the sense-making end of this continuum.

3. Comparing three introductory physics courses based on students’ perceptions

To understand whether students perceived the emphasis of CQs in the course to focus on sense-making versus on answer-making, we asked them two specific survey questions. The first question was: In class, how important is it for you to articulate your reasoning either to your peers or during whole class discussion? (Not important at all, Not very important, Somewhat Important, Important, and Very Important). The distribution of student responses on this question is shown below in Figure 19.

![Figure 19: (Color) Example Question--Distribution of student responses to the importance of articulating their reasoning.](image)

From student responses to this question, we can see that Red’s students report articulating their reasoning as more important than do Yellow and Green’s students. We see that while Green’s students most likely report that articulating their reasoning is somewhat important, Red’s students most likely report that articulating their reasoning is important. Red’s students tend to choose higher (more favorable) responses than Green and Yellow’s on this question. We see in Table K that these differences are statistically significant when comparing Green and Red.

We investigate whether this trend persists across a second survey question: Knowing the right answers is the only important part of the clicker questions (Strongly Disagree, Somewhat Disagree, Not Sure, Somewhat Agree, and Strongly Agree). Both of these questions have
answer options that are categorical and rank-ordered. Since there were statistically significant differences between these three groups via Kruskal-Wallis Test on both questions, the courses are compared pair-wise using the Mann-Whitney U Test as described before. The results of the pair-wise comparisons are provided in Table K.

**Table K: (Color) Mann-Whitney U Test Results for Pair-wise Comparisons on Sense-making Survey Questions.** The arrows in the table indicate which group tends to yield higher (more favorable) responses for all p-values less than 0.1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yellow-Red</th>
<th>Green-Red</th>
<th>Yellow-Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q17</td>
<td>R p=0.067</td>
<td>R p&lt;0.001*</td>
<td>Y p=0.045</td>
</tr>
<tr>
<td>Q13</td>
<td>R p=0.002*</td>
<td>R p=0.001*</td>
<td>p=0.392</td>
</tr>
</tbody>
</table>

We see from Table K that Green and Red differ significantly on both questions, with Red’s students reporting more favorably on both questions. We also see that Yellow and Red vary significantly on one of the two questions, with Red’s students reporting more favorably on both questions. We do not find statistically significant differences in students’ perceptions of Yellow and Green. We conclude that Red’s students perceive there to be a higher emphasis on sense-making than do Green’s students. We conclude that Red’s students perceive there to be a somewhat higher emphasis on sense-making than do Yellow’s students. However, Yellow and Green’s students perceive the emphasis on sense-making similarly.

These two survey questions show that students from these different courses notice some of the same differences in classroom norms as the research observer. The correspondence between the observing researcher’s interpretation of the norm of sense-making and students’ perceptions regarding a subset of this norm supports the claim that concrete classroom practices associated with the implementation of PI can impact the way that students perceive the emphasis on sense-making.

VI. Discussion & Conclusion

We have defined continua representing a set of three norms of Peer Instruction: faculty-student collaboration, student-student collaboration, and emphasis on sense-making vs. answer-making. Through observations and analysis of classroom practices, we describe the norms associated with three different PI classrooms and place these classrooms along these continua. Students enrolled in these courses perceive differences in aspects of these classroom norms. We claim that concrete actions in the classroom lead to differences in classroom norms that are reflected in students’ survey responses.

The curious reader may question whether there is a simpler explanation for the differences in students’ perceptions. Red is the only physics education researcher and therefore may have a deeper understanding of the underpinning principles that make PI effective. We do not discard this possible explanation, but do not find it very useful. Defaulting to personal characteristics of Red does not help us to describe how Red’s knowledge is employed in the classroom. We seek a description of a plausible mechanism for how students come to different understandings about physics, the nature of learning, and the nature of physics that students develop as part of the class. We seek to describe the social practices (concrete interactions
between students and instructors) in each of these classrooms in hopes of better understanding how students come to different understanding of what PI is about and what is valued in their classes. We have drawn attention to concrete actions in the classroom that may support students in developing a sense that faculty-student collaboration, student-student collaboration, and sense-making are important in learning physics during PI.

Another possible alternative explanation for variations in students’ survey responses about PI might be that students’ perceptions of PI are dominated by students’ overall impressions of the course or instructor. We examined students’ responses to the university’s Faculty Course Questionnaires (FCQs) -- CU’s end-of-course evaluations collected through the university. Red’s instructor rating was 5.9 on a scale from 1 (low)-6 (high) while Yellow’s instructor rating was 5.35 and Green’s was 5.2. Although Red does have a statistically higher instructor rating, all three professors have instructor ratings that are quite high and are within a standard deviation (0.9) of the average for the natural science departments at CU (which is 5.0). The modest variation in overall course evaluation (all very positive) likely does not account for the qualitatively different perceptions of the specific classroom norms. In our larger set of student survey responses across six different courses [2], higher instructor ratings do not uniformly correlate with more positive student responses on PI survey questions. As above, the overall course evaluations do not provide a mechanism to explain why students have different perceptions – this overall course perception is a matter of perceived practices. Likely, we suspect, if anything, the particular PI practices and associated norms of valuing student discussion and supporting student engagement are components that lead to student perceptions of being valued and to higher course evaluations. Though, such claims are beyond the scope of this investigation.

Students’ prior experiences in physics courses and particularly physics courses that utilize PI will certainly influence the meaning that students make of PI in a given course. Enculturation of students over multiple semesters of using PI will lead to different expectations and possibly different perceptions of classroom norms. In some sense, that is what we argue within the span of a course. If however, the variation in student perceptions were solely a matter of prior experience with PI, we would expect to see a more dramatic shift in students’ perceptions between Phys1 (Yellow) and Phys2 (Green). More explicitly, in comparisons of two different Phys2 courses (Green vs. Purple) in Appendix C, it is clear that positive student perceptions are not simply a matter of how many semesters students have been exposed to clicker use. Students’ perceptions vary dramatically within the same Phys 2 course offered by different instructors. For more details about Purple’s PI implementation practices [76], please see Ref. 2. Since Purple’s students had taken the prior course, Phys1, with Yellow, we see that students’ perceptions of classroom norms surrounding PI can shift dramatically over the course of a single semester.

These studies demonstrate the association between particular sets of practices and student perceptions of classroom norms. Red creates a classroom of high collaboration between faculty and students – both during the smaller group discussions and during the whole class discussion, with more and varied opportunities for faculty-student discussion. Yellow and Green are more similar, where they provide selected modes and few opportunities for interaction. Notably, Green more often includes students in the whole class discussion as compared to Yellow. Students perceive these differences in norms, reporting greater comfort in interacting with Red by asking questions and discussing answers. Similarly Red is observed to engage in practices that are more likely to promote student-student collaboration; whereas Yellow and Green are remarkably
similar in their practices that do not promote student-student collaboration. No professor is observed to spend the majority of class time engaging in practices that focus on student-student collaboration. Again, students perceive these differences, with Red's students reporting more comfort and emphasis on speaking with each other than do Yellow or Green's students. Finally, Red's practices emphasize the role of sense-making more than answer-making in the PI episodes, and students perceive these differences, citing the importance of articulating reasoning over just getting the right answer. While Green is seen to engage in sense-making practices slightly more often than Yellow (with more whole-class discussion), they are both limited in their emphasis on sense-making, these modest differences between Yellow and Green are not observed in students’ responses.

This work suggests many avenues for future research. Based on our observations, there may be a wide variety of benefits to students associated with particular ways of using clickers. Researchers could begin to assess other possible benefits to students of particular implementations of PI such as changes in how students verbally communicate their scientific ideas, or changes in how students explain their reasoning associated with physics problem solutions. Attempts to replicate the findings of this research study across other classrooms would also be useful. Further investigations of PI implementation would be worthwhile to address the following questions: Is variation in PI implementation common within other institutions? Are variations in PI implementation across many institutions even broader than those identified in this study at a particular institution? At institutions where faculty closely collaborate in their teaching or spend a significant amount of time observing each other’s teaching, is there more consistency in PI implementation? Additional contrasting cases of PI implementation collected from a broader range of institutions would help in fleshing out a more complete understanding of clicker implementation and its associated impacts on students.

Prior research has called for a more detailed examination of classroom practices surrounding the use of clickers and associated impacts on students [61]. Professors’ PI practices do vary and PI is perceived differently by students in these classrooms. We show that collections of PI practices are associated with different classroom norms. In this research study, differences in classroom norms documented through classroom observations are associated with differences in how PI is perceived by students. We specify collections of classroom practices that appear to support students’ sense that PI is about faculty-student collaboration, student-student collaboration, and sense-making. Classroom norms may be specific to the personal preferences and learning goals of individual instructors, but the concrete practices articulated in this chapter may help instructors enact the goals (norms) they seek to encourage in their classrooms during Peer Instruction.

Peer-reviewed Publications based on this work:


References (Chapter 3)


[15] We note that these modes of participation have been referred to by other education researchers as participant structures (see Philips, 1983; Herrenkohl & Guerra, 1998) or participation frameworks (see Goffman, 1981).


[68] A Chi-squared Test is used to determine if we are justified in rejecting the null hypothesis which states that the row variables (instructor) and the column variables (possible answer options) are unrelated.
[69] A Kruskal-Wallis Test is used to determine if we are justified in rejecting the null hypothesis which states all of the populations are identical, versus the alternative that some of the populations tend to furnish greater observed values than other populations. This is analogous to an ANOVA for continuous data.
[76] In Ref. 2, Purple’s PI practices are described from a physics course for non-science majors, but observations of Purple in his Phys 2 course were found to be similar.
Chapter 4: Understanding the Tutorial Adoption, Implementation, and Institutionalization

I. Introduction

This chapter begins with a brief summary of what the Tutorials are, the structure of Tutorials use at the University of Colorado (CU), and prior research on the effectiveness of the Tutorials. Building on this foundation, additional research studies conducted on Tutorial adoption, institutionalization, and implementation at CU are presented. These studies include an analysis of Tutorial adoption and implementation at CU across multiple scales of the educational system, and an examination of instructors’ decision to adopt Tutorials. These research studies address the following research questions:

1) How did the Tutorials become established in the CU physics department?
2) How did individual instructors that were not involved in the initial Tutorials implementation at CU first decide to use the Tutorials in their courses?
3) How does Tutorial use vary across particular semester implementations?

Our analysis draws attention to the many simultaneous changes which occurred at multiple levels of the educational system that supported Tutorial adoption and implementation at CU as well as aspects of Tutorial use that have continually evolved. Physics instructors are seen as having a critical role in sustaining the use of the Tutorials since individual instructors maintain significant autonomy in academic settings. The critical role of faculty motivates examining instructors’ decision-making processes and reasons given for adopting the Tutorials through interviews with faculty immediately preceding their first semester teaching with the Tutorials. Following the results and discussion of these research studies conducted as part of this thesis project, a frames-of-context theoretical framework is applied to Tutorial use at CU in order to identify salient features of implementation. This framework is used to guide Tutorial research studies presented in following chapter. This analysis contributes to the development of a more nuanced model of educational change that moves beyond simplistic models of dissemination of curricular materials and a narrow focus on professional development of instructors.

II. Background

A. What Are The Tutorials?

The Tutorials in Introductory Physics (Tutorials) is a set of research-based curricular materials developed at the University of Washington (UW) to supplement a traditional large-enrollment introductory physics course at the university level [1]. The curricular materials include an activity workbook and associated homework book for students as well as an Instructor’s Guide [2] for professors with implementation suggestions, specific overview notes organized by activity, handouts, pretests for students, and suggested exam questions. The curricular materials were designed to be suitable for either a calculus-based physics course or an algebra-based physics course [Ref. 2, pg. viii].

The Tutorials were designed to modify the recitation section typically associated with large-enrollment introductory physics courses where students usually meet in smaller groups of 20-30 students. This research-based curriculum could therefore be used without dramatically changing the over-arching structure of a traditional course—large-enrollment lectures that meet three times per week, with a small group recitation meeting once per week. However, adoption of this curriculum does require changing the structure of the recitations: shifting content
emphasis from computation to conceptual understanding; shifting student activity from watching, listening, and transcribing to actively discussing, reasoning, and problem solving; shifting student interactions from individual work to group work; and shifting the role of the educator from a source of answers to a source of guiding and focusing questions.

Each activity in the Tutorials is based on extensive research into student thinking around a particular physics topic as well as considerable pilot testing of the activities. During the curriculum development process, the UW Physics Education Group has generally published multiple research articles (and many Ph.D. dissertations) on common student understandings and associated struggles encountered with particular physics topics including, for example, student thinking about velocity and acceleration in one dimension [3,4], or student thinking about introductory electricity [5,6]. In many of these articles, student pre/post performance data are presented on specific physical topics based on assessment questions designed at UW.

Each of the UW Tutorial activities is designed with the following structure: 1) elicit students’ initial thinking on the topic, 2) confront students with a situation either from an experiment that they conducted, based on logical reasoning, or based on self-consistency, and 3) demand that students resolve these conflicts [7]. As McDermott describes, “An instructional strategy that we have found effective for obtaining the necessary intellectual commitment from students is to generate a conceptual conflict and to require them to resolve it” [Ref. 7, pg. 296]. This model has generally been referred to as an “elicit, confront, resolve” approach. This process is strongly based on an accommodation model or cognitive conflict model of conceptual change [8,9] although others have presented theoretical and empirical problems associated with this model [10,11].

B. Typical Structure of Tutorial Use in Calculus-based Introductory Physics at CU

The CU physics department began using the UW Tutorials [1] in Fall 2003. A PER faculty first implemented these curricular materials in first semester calculus-based physics course (Phys1) which covers Newtonian mechanics, kinematics, energy, and momentum. In the seven years since that introduction, the Phys1 course has served approximately 575 students every semester. Students enrolled in the course enroll in one of two lecture sections that occur at different times such that the average number of students enrolled in a given lecture section is about 290. The Phys1 course consists of three 50-minute meetings in the lecture hall per week and an additional 50 minute small group recitation section. Each small group recitation section typically includes 25-30 students. Adopting the UW Tutorials curriculum predominantly impacted the small group recitation sections. In these environments, students now complete the Tutorials activities in groups of 2-5 students.

During the Fall 2004, the CU physics department began using the UW Tutorials [1] in their second semester calculus-based physics course (Phys2) which covers electricity, magnetism, and optics. The Phys2 course at CU, has over the last seven years, served approximately 400 students every semester. At the beginning of Tutorial use in Phys2 (2004), only one lecture section of Phys2 was offered in the spring semester and two lecture sections are offered in the fall semester. Beginning in 2008, two section of Phys2 were offered during both the Spring and Fall semesters. The average number of students enrolled in a given lecture section is about 250. Otherwise, the Phys2 course has the same basic structure as the Phys1 course described above.

Typically two faculty members are assigned to teach each of these introductory physics courses. Usually one faculty member, considered the primary instructor, is responsible for the
lecture portion of the course and the second faculty member, considered the “back-up” instructor, is in charge of course assignments and logistics such as the homework, exams, and grades as well as the Tutorial training meeting. An extensive Tutorial training meeting to prepare graduate teaching assistants (TAs) and undergraduate Learning Assistants (LAs) [12] to teach with the Tutorials in recitation sections replaced the short informal weekly meetings with instructional staff that tended to exist in prior traditional courses. Undergraduate LAs were CU’s solution to increasing the student-educator ratio required to implement small group work in recitations. These LAs are high-achieving undergraduates who have recently taken Phys1 or Phys2 and have an expressed interest in teaching. More details about the LA program will be discussed later in this chapter.

Other typical aspects of these introductory courses include 1) the use of clickers and Peer Instruction [13] during most lecture sections, 2) traditional end-of-the-chapter computationally-focused homework assignments completed through a computer-based system (e.g. Lon-CAPA, CAPA [14], or Mastering Physics [15]), 3) three mid-course exams and a comprehensive final, all scheduled outside of class time by the university, and 4) a Physics help room which was created in 2000 to allow students to work together on homework as well as to seek assistance from instructors and TAs [16]. With the addition of the Tutorials curriculum, students were also asked to complete online pretests on Tutorial specific topics prior to attending each recitation section and to complete homework associated with the Tutorial activities which was more conceptually oriented and required students to write out explanations of their reasoning.

III. Prior Research on Tutorial Implementation and the Effectiveness of Tutorials

A. Research on Tutorial Effectiveness at Other Institutions

To our knowledge, prior to the adoption of the UW Tutorials at CU (2003), there was only one research publication which described an attempt to replicate the student learning outcomes found through the use of the Tutorials at UW. Sabella investigated the impact of the UW Tutorials on student learning at Chicago State University [17] through three methods: 1) Force Concept Inventory (FCI) pre/post testing [18], 2) Open-ended pre/post testing on specific physics topics as conducted at UW, and 3) Interviews with a subset of students. This study first compared an algebra-based physics course taught with the Tutorials (N=7) to one taught traditionally (N=11) and found that there was little difference in student performance on the FCI. Next, pre/post data are presented on the specific topic of acceleration in one dimension for two courses using the Tutorials. The vast majority of students in both courses got the correct answer and explanation on the posttest question after completing the Tutorials, whereas none were able to on a similar question prior to the Tutorial. Further investigations into students’ thinking about acceleration in one dimension during interviews indicated that students’ understanding was not as robust as the authors would have hoped. The authors conclude that although students showed improvement on acceleration in one dimension, these results were not reflected in all of the topics covered by the Tutorials.

Around the same time as the CU replication studies, the United States Air Force Academy (USAFA) in Colorado Springs also conducted a replication study with the UW Tutorials [19]. In this study, tutorial students at USAFA were found to perform at the level of the honors students on conceptual final exam questions (on induced current and force) and were found to perform significantly better than control sections on three of four exam questions. Tutorial students also showed more confidence in writing physics explanations than control students. In contrast to the Chicago State study, this investigation suggests that students’
conceptual understanding was improved through the use of the UW Tutorials although only induced current topics were explicitly assessed.

In initial pilot studies of Tutorials use at the University of Cincinnati, Tutorial students consistently scored several percentage points higher than their non-tutorial peers on both conceptual and quantitative exam questions [20]. Sections that used the Tutorials were also found to have a lower withdrawal and failure rates. Following full adoption of Tutorials in their general physics course, students demonstrated significantly improved conceptual understanding (with Tutorials average normalized gain ranged from 0.46 to 0.57 as measured by the FCI [18], without Tutorials average normalized learning gains ranged from 0.18 to 0.24) [20]. In contrast to the Chicago State and USAFA studies, this research study provides compelling evidence that the Tutorials improved student conceptual learning across a breadth of topics.

Investigations at Chicago State University, the USAFA, and the University of Cincinnati primarily used the UW Tutorials activities without significant modifications (at least based on their descriptions in the literature). However, the University of Maryland PER group modified the UW Tutorials to utilize microcomputer-based laboratory equipment. The resultant curriculum is typically referred to as Activity-Based Physics Tutorials [21,22]. Average normalized student learning gains of 0.35 were reported through the use of Activity-Based Tutorials at the University of Maryland between Fall 1993 and Spring 1997 (N=546 over 9 courses) [21,23]. The results of this study provided additional evidence of the positive effects of a Tutorial approach for students’ conceptual learning across a broad range of topics.

Since the publication of the UW Tutorials, other researchers and curriculum developers have used this model to develop curricula for additional domains of physics. A few examples include: Bao & Redish’s Tutorials on classical probabilistic interpretations of physical systems [24], Ambrose’s Tutorials for intermediate mechanics [25], and Singh’s Tutorials for undergraduate quantum mechanics [26]. Approaches similar to the Tutorial approach have been appearing simultaneously in other disciplines such as Peer-led team learning (PLTL) [27] and Process-oriented guided-inquiry learning (POGIL) [28] in chemistry education. Recently, researchers at the University of Maryland have significantly modified the UW model for tutorial development [29] based on a different theoretical approach (a resource-based model of the mind [30]) and dramatically different learning goals (student development of a scientific epistemology [31,32]). These efforts have resulted in the University of Maryland Open-Source Physics Tutorials [33,34].

B. Research on Tutorial Implementation and Student Engagement Practices

More recently PER researchers have begun to study differences in implementation and student engagement in Tutorial environments. Here we briefly summarize two recent lines of investigation. One aims to better understand the importance of the actions and beliefs of the Tutorial educators (typically teaching assistants) in the Tutorials recitations [35,36,37], and the second aims to better understand exactly what students are doing in Tutorial learning environments [38,39]. These lines of research shed light on how teaching assistants and students are likely to engage with Tutorial curricular materials.

1. The Role of the Teaching Assistants in Tutorial Implementation

One of the first PER research studies on the role of teaching assistants in Tutorial learning environments sought to understand the necessary training required for Tutorial teaching assistants. Researchers investigated the impact of varying the instructor’s role in running
Tutorials on students’ learning [35]. An optional extra-credit opportunity was offered to students currently enrolled in an introductory physics course to participate in a single UW Tutorial activity on “Changes in Energy and Momentum” not currently being used in the course. Students were assigned to one of four recitation classes taught in different styles: 1) a traditional lecture taught by a physics faculty member following the structure of the Tutorial, but with the answers being told to the student explicitly, 2) students working individually through the tutorial with the instructor providing only an answer key given to students to allow them to check their work, 3) students working in cooperative learning groups of three or four through a tutorial with the instructor providing only an answer key given to students to allow them to check their work, and 4) students working in cooperative learning groups of three or four through a tutorial with the instructor performing checkpoints using Socratic dialogue. Students’ understanding of energy and momentum was assessed pre/post using questions developed at UW. The percentage of students getting all four parts of the posttest questions correct were almost twice as large for students in style 4 than for students in any of the other three recitation styles. This study showed that having Tutorial educators engage students in Socratic dialogue around their collaborative completion of a particular Tutorial led to higher student posttest performance on that topic as compared to either lecture format, individual or collaborative student completion of the Tutorial with answer key provided for self-correction. This research suggests that particular actions of the Tutorial educators may have a significant influence on student learning outcomes.

Researchers at the University of Maryland have systematically examined videotapes of students working on Tutorials and coded for whom or what initiated an interaction between a group of students and a TA [36]. Each interaction between the students and the TA was coded as beginning due to: S) Student initiated, T) TA initiated, and W) Worksheet initiated (checkpoint). It was found that TAs initiated 64% of interactions, students initiated 24% of interactions, and worksheets initiated 12% of interactions through checkpoints. The researchers then compared coded interactions across student groups, across TAs, and across Tutorial activities. They found that for all student groups there were more TA-initiated interactions than any other kind. Six of the eight TAs had, on average, more TA-initiated interactions than any other kind. TA-initiated interactions were the most common type for every Tutorial, although it also appears that student-initiated interactions dropped off midway through the semester. The researchers’ tentatively claim that the interaction pattern is mostly determined by the TA. The researchers worry that the Tutorial environments may not be as student-centered as they might have hoped due to the limited amount of student-initiated interactions. This research study also suggested that TAs may have significant influence over initiating interactions in their recitations sections.

Comparing TAs at the University of Maryland College Park and the University of Colorado at Boulder, Goertzen, Scherr, and Elby describe the degree of TA “buy-in” to reformed instruction which is defined to be the alignment of a TA’s stated set of beliefs about how physics should be taught compared to the beliefs of the curriculum developers [37]. TAs’ degree of buy-in is coded based on statements made in interviews and the researchers caution that broad characterizations of TAs as “not buying in” risk obscuring valuable information about specific attitudes and skills that TAs already have. Through an illustrative case study, ‘lack of buy-in’ is shown to have clear instructional impacts in the classroom suggesting that buy-in is a necessary (but insufficient) component of effective curriculum implementation. The researchers state that the TA preparatory meetings are very similar at both institutions and claim that differences in TA buy-in across CU and the University of Maryland are due to differences in the social and
environmental contexts. This article guides us to focus more broadly than just individual TAs and examine the social and environmental contexts within which TAs learn and work.

2. Observational Research into Student Behaviors and Associated Student Framing in Tutorials

By analyzing video-recordings of student-group activity in Tutorials sessions, Scherr and Hammer identify four frames related to stable modes of student behavior based on cues such as posture, gestures, vocal register, and visual focus [38]. Scherr and Hammer related these stable modes of student behavior to four frames common to students’ activity in Tutorials learning environments: 1) Worksheet Frame, 2) Discussion Frame, 3) TA Frame, and 4) Joking Frame. A frame describes how an individual or group forms a sense of “What is it that’s going on here?” Framing presents a communicative task in which participants collaboratively establish the nature of their shared activity and are constituted by verbal and nonverbal interaction. Each identified frame is associated with a set of behavioral cues and an associated set of expectations. For example, the Worksheet Frame consists of students with hands quiet, face neutral, body leans forward, eyes on paper, with only brief glances at peers, and muttering with expectations of minimal interaction, individual activity, attention belonging to the worksheet, and peers not attending to details of speech. The Discussion Frame, in contrast, consists of students gesturing prolifically, using an animated tone and facial expressions, sitting up straight, using eye contact, and making clear utterances with expectations that peers are watching each other and wanting to understand, intellectually and emotionally engaged, attention belonging to peers, and peer interest in details of speech. Scherr and Hammer’s research study describes a methodology for identifying frames and shifts in frames in immediate activity providing evidence of both variability and local coherences.

Building on the work of Scherr and Hammer, researchers from the University of Maryland began to investigate the relationship between students’ frame (Discussion, Worksheet, Socializing, and Receptive to TA) and the substance of student reasoning in UM Tutorials [39]. Russ’s coding scheme for identifying mechanistic reasoning in students’ physical explanations is used to single out “chaining” as a particularly sophisticated form of mechanistic reasoning [40]. Results of this investigation suggest that advanced mechanistic reasoning (chaining) occurs primarily during the animated discussion frame. Additionally, multiple reasons for students shifting frames are identified including: explicit frame negotiation, contrasting lines of reasoning, and contrasting student epistemological stances taken. This investigation reminds us that students actively participate and shape the Tutorial environments.

C. Replicating UW’s Student Learning Outcomes at CU through the use of Tutorials

Since the initial implementation of the Tutorials at CU in Phys1, the Physics Education Research Group at Colorado has sought to understand the impacts of this curriculum on students. In describing Tutorial implementation, pseudonyms have been assigned alphabetically to all faculty involved in teaching the introductory courses. During the first two semesters of Tutorial use at CU by Professor Arthur, a variety of student outcomes were assessed [41], including students’ topic-specific understanding based on weekly UW pretest and posttest performance, student conceptual learning gains measured by both the FCI [18] and the Force and Motion Conceptual Evaluation (FMCE) [42], students’ attitudes and beliefs about learning using the Colorado Learning Attitudes about Science Survey (CLASS) [43], and students’ affective response to the Tutorials and course through end-of-term surveys. Based on weekly UW pretest and posttest assessments of students’ topic-specific understanding, Finkelstein and Pollock were
able to demonstrate remarkable replication of the student achievement documented at the UW on specific physics topics such as combining Newton’s Laws, applying Newton’s third law, and constructing force diagrams [41].

In addition to these replication measures, Finkelstein and Pollock also documented students’ improved conceptual mastery—an average matched normalized student learning gain of 0.67 on the FCI during the Fall 2003 semester and an average matched normalized student learning gain of 0.63 on the FMCE during the Spring 2004 semester [41]. These results placed both initial implementations of the Tutorials in Phys1 at CU in the “high-gain” category defined by Hake [44]. Results of the CLASS during these two initial implementations at CU demonstrate no significant change in students’ attitudes and beliefs about learning physics over the semester [41] which is arguably a positive result compared to the degradation of students’ expectations in introductory physics measured elsewhere [45,46]. Based on online survey responses towards the end of the term, students report an overall favorable response to the utility of the Tutorials and group work; however, a large fraction of students report that they did not enjoy the Tutorials [41]. Based on this extensive documentation, Finkelstein and Pollock proceed to describe the conditions that allowed for the successful appropriation of Tutorials in Phys1 from a frames-of-context perspective [47]. An analysis from this perspective was fruitful for understanding how students’ experiences with the Tutorials shape (and are shaped by) broader frames of context (from departmental norms to specific Tutorial tasks). Finkelstein and Pollock’s paper concludes with calls for additional research studies to investigate if other faculty members will adopt Tutorial practices and whether students develop a more positive response to these practices over time.

D. Longitudinal Documentation of Student Learning at CU using the Tutorials

After replicating the student conceptual learning gains achieved at the University of Washington [41] and documenting initial success of the Tutorials in Phys1 at CU, the department began an effort to expand and sustain the use of the Tutorials. During Fall 2004, a team of PER faculty—Professors Arthur and Felix—began using the Tutorials in Phys2. During this initial implementation of the Tutorials in Phys2, high normalized student learning gains ($g=0.44$; $N_{matched}=400$) were documented [48] using the Brief Electricity and Magnetism Assessment (BEMA) [49]. After this initial implementation by the original CU Tutorials adopter, the PER group sought to measure a secondary implementation by Professors Felix and Harold (a non-PER faculty member) in which Professor Arthur’s participation was minimal.

The initial hand-off from Professor Arthur to Professor Felix arguably occurred under the best possible circumstances. Professor Felix was mentored by Professor Arthur during the prior semester and had extensive knowledge of PER research outcomes and pedagogies. Using the BEMA results, Pollock and Finkelstein were able to show that the average posttest scores were the same (59%) and average normalized learning gains indistinguishable: $44\% \pm 1\%$ for the initial Phys2 implementation and $43\% \pm 1\%$ for the secondary implementation [48]. Histograms of students’ pre and posttest performance for the initial (Fall 2004) and secondary (Spring 2005) implementations can be found in Figure 20.
During Fall 2005, Professor Harold was the first non-PER faculty to be the lead instructor teaching with the Tutorials. Significant student learning gains were measured during this implementation (33% ± 1%) although gains were lower than was found with PER instructors previously (44% ± 1%) [48]. This curriculum was demonstrated to be effective at CU over multiple semesters, but history supplies many examples of effective innovations not being adopted or sustained [50]. Despite these promising initial results, the Colorado PER group continued to collect student conceptual learning gains data in both Phys1 (See Table L) and Phys2 (See Table M) as additional instructors moved through the courses. For each semester of data present in this thesis, a pseudonym has been assigned to each professor and since each pseudonym begins with a different letter the professors are referenced in tables and figures by the first letter of their pseudonym. Generally two professors work on a given course. The professor listed first is in charge of the lecture portion of the course. The second faculty member listed is the ‘back-up’ instructor who has little face time with students, and is primarily in charge of Tutorial educator training, homework, and exams.
Table L: Summary of FMCE normalized learning gains over twelve semesters of data collection taken from Ref. 48 with recent learning gains data added.

<table>
<thead>
<tr>
<th>Semester</th>
<th>Lead/Secondary Professors</th>
<th>Recitation Style</th>
<th>Avg. Posttest FMCE Scores&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Avg. Normalized FMCE Gain&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa03</td>
<td>A</td>
<td>Tutorials</td>
<td>81 (FCI)</td>
<td>0.63</td>
</tr>
<tr>
<td>Sp04</td>
<td>A</td>
<td>Tutorials</td>
<td>74</td>
<td>0.64</td>
</tr>
<tr>
<td>Fa04</td>
<td>D/E</td>
<td>Workbooks</td>
<td>69</td>
<td>0.54</td>
</tr>
<tr>
<td>Sp05</td>
<td>E/G</td>
<td>Traditional</td>
<td>58</td>
<td>0.42</td>
</tr>
<tr>
<td>Fa05</td>
<td>G/I</td>
<td>Traditional</td>
<td>58</td>
<td>0.39</td>
</tr>
<tr>
<td>Sp06</td>
<td>J/K</td>
<td>Tutorials</td>
<td>60</td>
<td>0.45</td>
</tr>
<tr>
<td>Fa06</td>
<td>N/D</td>
<td>Tutorials</td>
<td>67</td>
<td>0.51</td>
</tr>
<tr>
<td>Sp07</td>
<td>H/K</td>
<td>Tutorials</td>
<td>62</td>
<td>0.46</td>
</tr>
<tr>
<td>Fa07</td>
<td>J/P</td>
<td>Tutorials</td>
<td>69</td>
<td>0.54</td>
</tr>
<tr>
<td>Sp08</td>
<td>S/T</td>
<td>Tutorials</td>
<td>52</td>
<td>0.32</td>
</tr>
<tr>
<td>Fa08</td>
<td>S/V</td>
<td>Tutorials</td>
<td>72</td>
<td>0.58</td>
</tr>
<tr>
<td>Sp09</td>
<td>F/W</td>
<td>Tutorials</td>
<td>69</td>
<td>0.54</td>
</tr>
</tbody>
</table>

<sup>a</sup>The average posttest score is a percentage, scored using Thornton’s recommended FMCE rubric (Ref. 51), for those students with matched pre-post data. (The standard error on the mean is between 1% and 2% for all terms.) Only Fa03 used the FCI (Ref. 18) exam pre and post, all other terms used the FMCE.

<sup>b</sup>Normalized gain in the last column is computed as the gain of the average prescores and postscores for matched students. (The standard error of the average gains is roughly ±0.02 for all terms.)

Table M: Summary of BEMA normalized learning gains over ten semesters of Tutorial use in recitations taken from Ref. 30 with recent learning gains data added. The standard error on the mean for this table is 1-2% on posttest scores, and roughly ± 0.02 for normalized gains.

<table>
<thead>
<tr>
<th>Semester</th>
<th>Lead/Secondary Professors</th>
<th>Avg. Posttest BEMA Scores (%)</th>
<th>Avg. Normalized BEMA Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa04</td>
<td>A/F</td>
<td>59</td>
<td>0.44</td>
</tr>
<tr>
<td>Sp05</td>
<td>F/H</td>
<td>59</td>
<td>0.43</td>
</tr>
<tr>
<td>Fa05</td>
<td>H</td>
<td>50</td>
<td>0.33</td>
</tr>
<tr>
<td>Sp06</td>
<td>L/M</td>
<td>53</td>
<td>0.37</td>
</tr>
<tr>
<td>Fa06</td>
<td>H/A</td>
<td>56</td>
<td>0.40</td>
</tr>
<tr>
<td>Sp07</td>
<td>O/N</td>
<td>53</td>
<td>0.37&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fa07</td>
<td>A/R</td>
<td>61</td>
<td>0.47</td>
</tr>
<tr>
<td>Sp08</td>
<td>P/J</td>
<td>58</td>
<td>0.41</td>
</tr>
<tr>
<td>Fa08</td>
<td>R/M</td>
<td>54</td>
<td>0.36</td>
</tr>
<tr>
<td>Sp09</td>
<td>H/N</td>
<td>56</td>
<td>0.40&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>No pretest was given in Sp07 and Sp09, but the average pretest has been 26% ± 1% every other term, so this was used to estimate the normalized gain for that term.

Using the Tutorials curriculum at CU in Phys1 over nine semesters, FMCE posttest scores have ranged from 52% to 74% and FMCE normalized learning gains have ranged from 0.32 to 0.64. Using the Tutorials curriculum at CU in Phys2 over ten semesters, BEMA posttest scores have ranged from 50% to 61% and BEMA normalized learning gains have ranged from 0.33 to 0.47. These results show dramatic variation in student conceptual learning outcomes despite these courses, on the surface, using the same curricular tools. Commenting on a subset of these data, Finkelstein and Pollock claim, “…there is sizable variation of success among these implementations, suggesting a potential effect of faculty in these environments” (Ref. 48, pg.4).
Finkelstein and Pollock call for subsequent work to examine subtle variations in implementation including what messages are sent and how they are sent at a variety of levels of the course (Ref. 48, pg. 6). Tutorial implementation practices at CU are described later in this chapter and students’ interpretations of Tutorial use in these courses will be presented in the following chapter.

IV. Study of Tutorial Adoption and Institutionalization at CU

There is significant interest in understanding the nature of educational change [52] and this analysis of Tutorial use at CU provides a case study in such institutional change. This case study addresses the question: how did the Tutorials get established at CU? By drawing on extensive analysis of historical documents and artifacts such as grant proposals, grant reports, meeting minutes, and email records of correspondence, as well as targeted interviews of key personnel, this work contributes to the development of a model for change processes and institutionalization of educational transformations (described in the following section). The change process associated with the adoption and institutionalization of the Tutorials is described at the scale of the instructor, the department and the institution as well as the role of support from external funding sources.

A. Setting the Stage: Departmental History

The CU physics department’s dedication to high quality education and educational innovations has a long history. The department has a legacy of dedicated and innovative educators such as George Gamow, Frank Oppenheimer, Al Bartlett, and John Taylor [16]. More recently, there were many factors that contributed to the departmental support for education prior to efforts to implement the Tutorials. The support and involvement of Nobel Laureate, Carl Wieman in Physics Education Research (PER) and course transformation added greatly to departmental awareness and respect for the field. In 2001, the initial use of personal response systems (such as HiTT [53] or iClicker [54]) and Peer Instruction [13] began in CU physics. Within two years, Peer Instruction became commonly used by all faculty members in the large introductory physics courses. In 2002, the physics department faculty began to gather for brown bag discussions in education. These meetings, held over lunch, were a forum for faculty in the department to discuss relevant issues in education in an informal manner. A sample of discussion topics during the first year included: educational technologies and curricula being developed at other institutions, sharing of locally developed course materials, and strategies for teaching to a broad spectrum of student backgrounds [55]. In 2002, the physics department also granted its first PhD in the field of PER. In 2003, Carl Wieman acquired an additional instructor line from university administrators to support his work in PER which was subsequently expanded into a tenure line for PER within the physics department. The physics department however was left to decide if it would support a PER program and accept PER research within the physics department. The department voted to accept the tenure line in PER and formally establish a PER research group in 2003. All of these efforts set the stage for the department to adopt and sustain the use of the Tutorials.

B. External Funding Sources

The initial implementation of the Tutorials was made possible by the interest and energy of Professor Steven Pollock, the initial adopter of the curriculum at CU, but could not have succeeded without the support of the STEM-TP Grant (NSF#03022134). The STEM-TP grant
began the CU LA program. This program recruited science majors for early educational experiences in undergraduate courses and supported the transformation of the way science departments teach at the undergraduate level. These undergraduate LAs provided the critical infrastructure to increase the teacher to student ratio required for small group work as recommended by the Tutorial designers. The evaluation, sustained use, and expansion of the Tutorials was accomplished with the support of Colorado PhysTEC Grant from the American Physical Society (APS) and the American Association of Physics Teachers (AAPT) as well as Course, Curriculum, and Laboratory Improvement (CCLI) Grant (NSF #0410744). These two grants helped to expand and to bridge funding from the STEM-TP grant and to sustain the LA program. Currently (as of fall 2009), the LA program is supported from: the CU physics department, the University, private donors, and an NSF Teacher Professional Continuum (TPC) grant #0554616 (which researches and partially supports the program).

C. Instructor Involvement

Over the course of a six year period that CU has been using the Tutorials in its introductory course, (Fall 2003-Spring 2009), 16 different physics instructors have been involved in implementing the Tutorials. Pseudonyms have been assigned to all faculty members that taught the introductory sequence during this six year period which can be found below in Table N. All pseudonyms start with a different letter and will occasionally be abbreviated by just the first letter of their pseudonym. Of the 16 different physics instructors involved in using the Tutorials curriculum, only three (Professors Arthur, Felix and Violet) are members of the PER Group at CU. During the first two semesters of Tutorial use at CU (Fall 2003-Spring 2004), a single PER faculty member, Professor Arthur, taught Phys1 using the Tutorials. For extensive descriptions of these primary implementations, please see Ref. 41. The initial implementations of Tutorials in both Phys1 and Phys2 were led by PER faculty.

Beginning with the Spring 2004 semester, the department decided to assign two faculty members to jointly teach a single course—a significant transition for the department. However, the two faculty members assigned to teach the Phys1 course, Professors Donald and Edgar, decided not to use the UW Tutorials (although limited funds for LAs was likely a factor also). This marked the beginning of a hiatus of Tutorials use in Phys1 which lasted for a year and a half. Meanwhile, Professor Arthur, the original adopter of the Tutorials, transitioned to teach the Phys2 course with another PER faculty member, Professor Felix. Professors Arthur and Felix were the first professors to team-teach with the Tutorials during Fall 2004. Since Fall 2004, it has been the norm for the physics department to assign two faculty members to team-teach the Phys1 and Phys2 courses. (The one exception during the Fa05 semester of Phys2 was due to a last minute family emergency of a faculty member.) The department, however, did not explicitly specify how the teaching responsibilities should be shared. The general model that emerged consisted on one faculty member, usually considered the lead instructor, being responsible for preparing and delivering the lectures while the second professor, usually considered the secondary (or “backup”) professor, would be responsible for out-of-class assignments and course management such as the Tutorials, the pretests, the TA/LA training meetings, and the homework assignments. The exams tended to be jointly constructed by both professors. This means that the lead instructor is dominant in establishing the class culture in lecture and the secondary instructor primarily influences the Tutorial situations. In practice, however, there is significant variation in the distribution of teaching responsibilities from semester-to-semester as will be discussed in more detail later in this chapter.
Table N: (Color) Pseudonyms for instructors involved in Phys1 and Phys2 between Fall 2003 and Spring 2009 with associated recitation curriculum. Semesters where Tutorials were used are shaded green.

<table>
<thead>
<tr>
<th>Semester</th>
<th>Phys1</th>
<th>Phys2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2003</td>
<td>Professor Arthur (A)</td>
<td>Professor Bartholomew (B)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>No Tutorials (Traditional)</td>
</tr>
<tr>
<td>Sp 2004</td>
<td>Professor Arthur (A)</td>
<td>Professor Calvin (C)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>No Tutorials (Traditional)</td>
</tr>
<tr>
<td>Fall 2004</td>
<td>Professors Donald (D) &amp; Edgar (E)</td>
<td>Professors Arthur (A) &amp; Felix (F)</td>
</tr>
<tr>
<td></td>
<td>No Tutorials (Knight Workbook)</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Sp 2005</td>
<td>Professors Edgar (E) &amp; Gus (G)</td>
<td>Professors Felix (F) &amp; Harold (H)</td>
</tr>
<tr>
<td></td>
<td>No Tutorials (Traditional)</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Fall 2005</td>
<td>Professors Gus (G) &amp; Ira (I)</td>
<td>Professor Harold (H)</td>
</tr>
<tr>
<td></td>
<td>No Tutorials (Traditional)</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Sp 2006</td>
<td>Professors Jasper (J) &amp; Kent (K)</td>
<td>Professors Linus (L) &amp; Milo (M)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Fall 2006</td>
<td>Professors Nemo (N) &amp; Donald (D)</td>
<td>Professors Harold (H) &amp; Arthur (A)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Sp 2007</td>
<td>Professors Harold (H) &amp; Kent (K)</td>
<td>Professors Opus (O) &amp; Nemo (N)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Fall 2007</td>
<td>Professors Jasper (J) &amp; Petunia (P)</td>
<td>Professor Arthur (A) &amp; Ron (R)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Sp 2008</td>
<td>Professors Stuart (S) &amp; Tabitha (T)</td>
<td>Professors Petunia (P) &amp; Jasper (J)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Fall 2008</td>
<td>Professors Stuart (S) &amp; Violet (V)</td>
<td>Professors Ron (R) &amp; Milo (M)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
<tr>
<td>Sp 2009</td>
<td>Professors Felix (F) &amp; Wilbur (W)</td>
<td>Professors Harold (H) &amp; Nemo (N)</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>Tutorials</td>
</tr>
</tbody>
</table>

D. Recruitment and Training of Educators to Lead the Tutorial Recitation Sections

The CU physics department supplies one graduate teaching assistant (TA) to teach each of the recitation sections. The departmentally supplied TAs are typically first year graduate students in the physics department and are only occasionally more veteran graduate students. These graduate students alone were insufficient to meet the recommended 1:15 teacher to student ratio for implementing the Tutorials [45].

By partnering with the school of education and other science departments, CU was awarded a science, technology, engineering and mathematics teacher preparation (STEM-TP) grant which supported the development of the LA Program at CU [12]. This program has three goals: 1) To increase the quantity and quality of STEM K-12 teachers by recruiting science majors for early educational experiences in undergraduate courses, 2) To transform the way science departments teach science at the undergraduate level, and 3) To research these systems to improve the education of undergraduates and teachers alike. The LA program supported the recruitment, hiring, and training of high-achieving undergraduates with an expressed interest in teaching who had recently taken Phys1 and Phys2. One to two LAs joined the departmentally supported TA to teach each recitation section in the Phys1 and Phys2 courses. The training of these LAs was supported both by the Physics Department and by the School of Education. All
undergraduate LAs during their first semester are required to enroll in a one credit LA-specific pedagogy course in the School of Education team-taught by a science education professor and a high school master teacher. With the goal of supporting students’ integration of theory, pedagogy, and practice, this course touches on theoretical issues such as conceptual development, conceptual change, collaborative learning, and students’ conceptions of various mathematics and science topics, as well as practical issues encountered in facilitating learning, managing the classroom, formative and summative assessment, and differentiating instruction in a collaborative learning environment.

The physics department also developed weekly training meetings for TAs and LAs that typically last 1-2 hrs in accord with the UW’s approach to recitation training. In both Phys1 and Phys2, TAs and LAs meet with the secondary professor to prepare to teach the next week’s Tutorial activity. At this meeting, TAs and LAs typically examine incoming student thinking on the topic to be taught that week based on student pretest responses. Next, the instructor leads a discussion with the TAs and LAs about common ideas they expect students to come in with. The TAs and LAs then work through the Tutorial activity familiarizing themselves with the structure of the activity itself as well as refreshing themselves with or relearning material. The secondary instructor usually models appropriate Socratic questioning with the TAs and LAs as they work through the activity, drawing attention to how to guide student learning. Although there are many common features to these departmental training meetings, the nature of these meetings varies greatly from semester-to-semester, strongly influenced by the participation and inclinations of the secondary instructor which will be further discussed later in this chapter.

E. Changes in Departmental Support

After the initial implementation, the CU physics department gradually committed additional resources to support the Tutorials. The dedication of these resources began after the Tutorials were first used in a more traditional classroom setting with individual chairs and desks. After some demonstration of the effectiveness of this curriculum, in 2004 the department committed dedicated classroom space, which was created from the larger student laboratory area. The department purchased larger tables with moveable chairs and equipment for laboratory activities using money from the course fees fund. These resources allowed the Tutorials to operate in a dedicated space that more easily facilitated group interactions. Shortly after the commitment of these resources, the department decided to systematically institute team teaching assignments as mentioned before. The intention behind introducing team teaching was to lower the load on individuals so that they will have more time to: reflect on their teaching, prepare for teaching, and do the work necessary for implementing innovative teaching strategies [55]. While the division of labor varied across teaching pairs, often one professor would take responsibility for the lecture periods (lead), while the other would manage the TAs, recitation sections, course assignments, and course logistics (secondary).

Although the physics department has a long history of partnering with the Graduate Teacher Program (GTP), the physics department began to depend on the GTP’s Lead TA to support the implementation of the Tutorials. The GTP supports a small raise in salary for a Lead TA in order for them to spend time mentoring other graduate students in the department about teaching. The Lead TA began to take some responsibility for the set-up and maintenance of equipment for the Tutorial recitations although this was initially in addition to a half-time TA teaching assignment. Beginning in fall 2005, this Lead TA shifted such that the Lead TA role is partially supported by the department—they are given only a quarter time TA assignment with
students and the other quarter time appointment is spent supporting Tutorial set-up and TA training. This created a Lead TA position that was not an added burden to the chosen Lead TA, but rather a departmentally supported position to support the development of other TAs and LAs. Over the course of a six year period that CU has been using the Tutorials in its introductory course, (Fall 2003-Spring 2009), 5 different Lead TAs have been involved in implementing the Tutorials—only one of which was involved prior to expanded departmentally supported position.

Additional staff support was available for a few semesters during the early adoption of the Tutorials through the PhysTEC grant. The PhysTEC grant supported a high school physics teacher to take a leave of absence from their school in order to take a sabbatical with the CU physics department and PER group. During some semesters (Fall 2006-Spring 2007), this master teacher was heavily involved in running or supporting faculty in the Tutorial Training meetings. When a master teacher was involved, they tended to play a similar role as the Lead TA.

The physics department continued its brown bag discussions and its partnership with the Learning Assistant program. To support the development of the PER group and efforts to recruit physics majors to teaching, the physics department added an elective course in Fall 2004—Teaching and Learning Physics [47,56,57] which routinely places undergraduate and graduate students as volunteer LAs in the Tutorial courses for teaching and research experiences. Recently (2008), the department also awarded tenure to its first professor in the field of Physics Education Research.

In Spring 2005, the Teaching Evaluation Committee in the department decided to further modify the peer teaching evaluation criteria. These modifications placed increasing emphasis on student engagement and coordination of course learning goals and course activities. These evaluations were designed to recognize innovative efforts using Peer Instruction, Tutorials, and other student-centered instructional strategies. Prior to these changes, evaluation criteria were more heavily based on characteristic of good lecturing such clarity of speech, effective use of the chalkboard, enthusiasm, clarity of explanation, preparation, and effective class engagement.
Figure 21: (Color) Timeline showing professor involvement, changes in departmental and institutional support, and external funding sources. Each professor in this timeline is labeled by the first letter of their pseudonym. The initial implementation for each course is shown in Bold & Red.
F. Changes in Institutional Support

As the physics department showed increasing commitment to sustaining the Tutorials, the university began to support these efforts as well, see Figure 21. In the sixth semester of using the Tutorials (Fall 2006), the Dean of Arts and Sciences began to commit intermittent funds to support the use of the Tutorials (supporting roughly 20% of the LAs). The physics department also partnered with the Faculty Teaching Excellence Program (FTEP) and the GTP. The partnership with FTEP provided summer support for faculty research on course transformations, development of new course materials, and the creation of an annual two-day teaching assistant workshop for new incoming graduate students (beginning in 2003). The GTP helped to support a lead graduate teacher to mentor other teaching assistants and typically assist with Tutorial preparation meetings and equipment set-up as discussed in detail in the prior section. In Fall 2007, the Provost along with the Dean of Arts and Sciences and the School of Education committed funding for these efforts and it appears that this support will continue into the foreseeable future. This commitment was spurred in part by a significant ($120K) private donation to the LA program, along with concerted lobbying efforts of LA program PIs.

G. Limitations of Our Methodological Approach to Studying Change

There are important aspects of departmental and institutional cultural change that were not systematically captured by our methodological approach (focusing on historical artifact analysis supplemented by some interviews and observations). We are aware of many informal discussions between physics faculty which occur in the halls or elevators spontaneously. When professors are assigned a new course to teach in the CU physics department, they will commonly talk to faculty members who have taught the course recently. In addition to these informal discussions about education that occur between faculty members at the university, we are also aware of casual discussions about education occurring at informal social gatherings. Specifically, one faculty member (Professor Ron) was first informed of Peer Instruction and clickers at a BBQ at another faculty member’s house. The discussion that ensued at the BBQ resulted in the faculty member leaving the BBQ with the intention to try clickers and Peer Instruction in his course the following semester. We believe that the prevalence of and the informal/colloquial nature of these discussions may be an important element of the departmental change that occurred at the University of Colorado; however these discussions are not inscribed in documents (i.e. emails, meeting minutes…) and were therefore not included in our systematic analysis. Some faculty members believe that the emphasis and evaluation of teaching in the department has also changed; however these changes are largely informal and are not always captured in the formal hiring and promotion documents (beyond the peer teaching evaluation system already described). The analysis based on these peer evaluation rubrics may not capture the full degree to which teaching is valued in hiring and promotion in the physics department.

Professor Arthur also plays important behind-the-scenes roles in allowing the Tutorial system to persist. These gentle nudges at critical times during the semester are difficult to capture, but are likely to be integral to Tutorial success. These behind-the-scenes nudges include reminders about when to order additional Tutorial activity books for the student bookstore, reminders about application deadlines to request learning assistants, reminders about hiring deadlines for prospective learning assistants, etc. Professor Arthur also runs workshops with all of the incoming graduate students at the beginning of the academic year, so that all graduate students are aware of the reforms that are being implemented in the introductory physics courses. This early exposure of graduate students to how things are done in the CU physics department
may play an important role in gaining TA buy-in. Professor Arthur is also often present at the first couple of TA/LA training meetings every semester. At these early meetings, Professor Arthur finds another opportunity to make explicit the purpose and worth of the Tutorials to both graduate students and the faculty members teaching the course. The many behind-the-scenes influences of Professor Arthur may not be systematically captured by our methodological approach.

H. Summary of Tutorial Adoption and Implementation

Through this case study analysis a sufficient, but not a necessary criterion for institutional change is described. The account above suggests the following elements may be instrumental for sustaining change: 1) Address and coordinate multiple levels of the educational system (individuals, department, interdepartmental collaborations, partnerships with programs across campus and institutional administrators), 2) Develop local structures that support reflective practice of faculty—informal discussion about instruction and learning, the sharing of course materials and assessment results, 3) Cultivate faculty members as advocates for educational transformation and 4) Dedicate infrastructure and resources to support the change process. Some of these elements by be particularly supported by the presence of a PER group. Although these factors emerge as important, additional case studies across institutions are needed to fully characterize the critical features of institutional change.

V. Faculty Choice of Adoption

Here, the fine-grained features of faculty adoption are explored: the process faculty went through when deciding to adopt the curriculum and the reasons given for deciding to use the Tutorials. While the commitment of funding and resources were necessary, the Tutorials curriculum could not be sustained without the voluntary involvement of many physics professors. At CU, curricular decisions are not designated by the department, nor decided through a committee process. Historically, curricular decisions have fluctuated based on the availability of resources and the inclinations of the professors assigned to the course. To sustain the use of the Tutorials, CU adopted a different model from the Tutorial authors. In the CU case, the responsibility for the Tutorials is distributed throughout the department rather than left to a specific group. At UW, the Physics Education Group (PEG) and Tutorial authors are the instructors assigned to run the Tutorials at UW, which ensures sustainability so long as the PEG group exists and is willing to run the courses. The alternative model at CU requires the physics faculty to agree to implement the Tutorials as part of their teaching assignments. Over the last twelve semesters implementing the Tutorials, 16 different faculty members have implemented this curriculum, only three of whom are PER faculty. The three-semester break from using the Tutorials in the Phys1 is due both to faculty choice and to the lack of institutional support: i.e. insufficient funds to pay LAs [48]. The next investigation describes how non-PER faculty members, new to teaching with the Tutorials, decided to adopt the Tutorials and the reasons they give for adoption.

A. The Decision Process

Between 2006 and 2007, six faculty members were interviewed prior to their first involvement in implementing the Tutorials to identify why they opted to implement Tutorials. Of these six professors, four felt personally involved in the decision to use the Tutorials curriculum (J, L, M, and P), see lower panel of Figure 21. The other two faculty members (K
and O) expressed that they were not involved in deciding to use the curriculum, but had defaulted to the decision of the other instructor in the team-teaching pair. In one case (Professor Kent), the decision was passed over to the lecture professor because the recitation professor perceived the course as primarily the responsibility of the lecture professor. In the second case (Professor Opus), the lecture professor was not a permanent faculty member and did not express awareness about how decisions like these were made. It appears that in this instance the recitation professor (Professor Nemo) alone chose to adopt the Tutorials (based on conversations with his team-teaching partner Professor Opus). Of the four professors interviewed that felt personally involved in the decision to adopt the Tutorials, three (Professors Linus, Milo, and Petunia) explicitly expressed that they had decided jointly with their team-teaching partner. These interviews also revealed that few of the instructors involved in using the Tutorials for the first time had looked at the Tutorial activity book. For example in the first interview with Professor Opus (01/16/07) a few days before the first day of classes teaching with the Tutorials he said, “I don’t know very much about them [The Tutorials]. I have a copy in the office, but to be frank, I haven’t had a chance to look through that.” Through these interview discussions, it became clear that four out of six of the professors preparing to use the Tutorials had not looked at the actual curricular activities prior to deciding to use the curriculum.

B. Reasons Given for Initial Adoption

The interview data from the four professors engaged in the decision-making process were transcribed and coded for instances where each professor mentioned the Tutorials or the decision to adopt the Tutorials. Key reasons that faculty members give for adopting the Tutorials include:

- **the old model of lecturing during recitations is ineffective** (n=3), For example, “it seemed like a no-brainer that we’d do the tutorials. The classic recitation. I mean, what are the other choices? The TA stand up and do problems? You might as well just not have them go to recitation. I know that’s pretty useless” (Professor Petunia, 09/05/07).

- **the argument made by Professor Arthur about the success of the Tutorials based on local student learning gains data was compelling** (n=3), For example, “What I don’t know is the extent to which the tutorials have been evaluated, say at the University of Washington. I know about Pollock’s efforts to understand learning gains. It looks like the tutorials do lead to learning gains. That’s good enough for me right now” (Professor Jasper, 02/01/06).

- **the structure of the Tutorials relieves teaching pressures from under-prepared TAs** (n=2), For example, “Most TAs are just thrown into the process, they are just new enough to the whole thing to feel uncomfortable in the front… So the tutorial structure relieves the TAs of having to stand there and deliver. And I think that alone is a good thing” (Professor Jasper, 02/01/06).

- **specific pedagogical aspects of the curriculum, e.g. Use of peer discussion, and attention to specific student misconceptions** (n=2), For example, “I do like the repetition, because you go through one thing and then you change it… They clearly are hitting upon the classic misconceptions. The zero acceleration when the object is stopped, I mean, you’ve seen that for years. Everybody always has that. So they’ve clearly identified some of those things. That’s nice” (Professor Petunia, 09/05/07).

- **Help and resources (LAs and additional faculty) were readily available for implementing the Tutorials** (n=1), For example, “I think that the biggest thing was that it was clear that there was going to be help. It wasn’t clear how much help, but the idea that there might
be learning assistants and that that might work out pretty well” (Professor Linus, 02/01/06).

There are a variety of potentially significant implications of these reasons faculty give for curriculum adoption [58,59]. First, there appears to be an initial dissatisfaction with traditional lecture methods. Second, in the interview data there is no mention of data from other institutions or nationally. One professor explicitly stated that he was not aware of data from UW, but was aware of data on the Tutorials collected locally. The importance of local data may be particularly important in thinking about the adoption process at other institutions. Faculty members may not be convinced by data that are non-local. Third, the additional resources made available for implementing the Tutorials may influence professors’ decision-making process and willingness to participate. Because of the LA hiring process, faculty members are asked to decide about whether they would like to use LAs in their upcoming course before they are explicitly faced with designing their course for the semester. In this sense the question comes early in the professor’s thinking about the course and is phrased as “would you like us to find LAs for you in order to teach with the Tutorial curriculum?” The order of these considerations is less about convincing faculty to use the Tutorials and has more to do with the practicality of recruiting and hiring LAs. This timing may have important implications for the adoption process—faculty members are expected to make decisions about their courses far in advance in order to get particular support or resources.

C. Summary of Faculty Adoption Choices

Interviews with faculty prior to using the Tutorials reveal that the actual process of deciding to use the Tutorials curriculum is a complex process when team-teaching assignments are involved. At times, one instructor in the team may decide to adopt Tutorials on their own while at other times the team-teaching pairs may decide together. Interviews also showed that few instructors had spent time looking at the Tutorials activities themselves prior to deciding to adopt them. Key reasons for adopting the Tutorials were identified—lecture seen as ineffective, local data on student learning were compelling, Tutorials reduce pressure on under-prepared TAs, specific pedagogical aspects of the curriculum, and the additional help and resources available. This analysis of interviews with faculty corroborates results from the prior case study analysis which suggests the importance of developing local structures that support reflective practice of faculty—specifically the sharing of assessment results and dedicating resources to support the change process.

As a result of these change processes, there is some evidence that the Tutorials have been accepted as the norm and are perceived by some faculty as an integral part of supporting student learning. As one professor states, “I think that without the Tutorials there’d be a huge vacuum that would be impossible to fill from the lectures. They’ve become in my mind pretty much part of physics 1, and physics 2... So it would be hard for me to imagine not having them” (Professor Harold, 09/05/07, following four semesters of teaching with the Tutorials). At least to this point it appears that the Tutorials have been sustained at CU.

VI. Modeling the Complexities of Tutorial Implementation

The case study analysis provided a historical account of many of the structural changes that occurred throughout the adoption and continued use of the Tutorials, but focused on the educational system within which the Tutorials were embedded. Based on prior work showing
semester-to-semester variations in students’ conceptual understanding, here particular attention is paid to variations in implementation that may influence students’ experiences over the time scale of a semester. In order to identify key variations in Tutorials use at CU, a frames-of-context model is applied to this system [47,60]. A frames-of-context model shows how the circumstances of a student working on a Tutorial task shape (and are shaped by) broader aspects of this educational system. This framework is a useful next step because it organizes the complexities of the Tutorial system and helps to define components of the system to investigate further in future research studies.

A. Frames of Context

The broad theoretical framing of this analysis draws from the socio-cultural perspective of student learning which focuses on the social, cultural, and contextually-bounded nature of learning. The early roots of this perspective can be traced to the Russian school of psychologists founded mostly on the works of Lev Vygotsky—one of the first psychologists to recognize that higher psychological processes are in their essence socially situated [61]. In Michael Cole’s book *Cultural Psychology*, he describes their work as follows: “The central thesis of the Russian cultural-historical school is that the structure and development of human psychological processes emerge through culturally mediated, historically developing, practical activity” (Ref. 60, pg. 108). Cole proceeds to build on Vygotsky’s work and develop a rich description of culture and the role of context in activity which will be revisited later in this section. For now, let us discuss how viewing student learning as socially situated in activities applies to understanding the use of the Tutorials.

For the purposes of this chapter, context can be thought of as “the interconnected collection of factors that frame the structure and meaning of human actions” [Ref. 62, pg. 198]. This definition of context moves beyond a definition of context as the external environment or that which surrounds. In the definition provided, context is multi-faceted consisting of the task being worked on, artifacts and tools used during this work, the personal histories of the participants, the social and physical setting. The construct of context can clearly become unwieldy for research purposes without further differentiation—what factors exactly frame the structure and meaning of human actions?

At this point it is useful to return to the work of Cole for a richer understanding and model of context. Cole argues against the common use of context as a backdrop or an empty container waiting to be filled by individuals and their activities. He claims that particularly for studying the richness of learning, context must be taken as more complex and substantial element of the learning process. To develop this model of context, Cole draws on “the Latin root of the term, *contextere*, which means to weave together” (Ref. 60, pg. 135). He develops this model further and describes context and task as mutually constitutive, meaning that context and task interact and define each other. Cole proceeds to discuss how multiple layers of context surround a given task and how interactions across these contextual levels define the activities that take place. Finkelstein applies Cole’s conception of multiple contextual layers and describes a particular physics classroom context in terms of nested contexts [47].

Using a model of nested contexts, the most relevant contexts related to this study are shown in Figure 22 below. Individuals involved in these course level tasks are participating in many different levels of contexts simultaneously (even more than are shown directly) and these different contextual levels are interacting with and defining each other and the activities that are taking place. Replication studies have shown that subtle contextual features of communities and
educational environments can have dramatic impacts on the effectiveness of educational innovations [41]. Additionally, prior research has shown that some faculty members attribute differences in their teaching practices and stated beliefs due to external systemic constraints [63]. It was based on these results that the most relevant surrounding contexts were selected to appear in this model. A Frame-of-Context model helps us to broaden our perspective on what might influence Tutorial implementation.

Levels of context depicted in Figure 22 mutually constitute each other and the boundaries shown as separating these levels are purely for analytical purposes. For our analytical purposes, levels were chosen such that they represented different spatial and temporal scales. For example, a given Tutorial task occurs in a relatively small physical space such as a given table in the physics basement and tasks are completed on relative short time scales for example a given task may take up only a small fraction of a 50 minute Tutorial class. Many key artifacts, actors and practices vary across levels and describing these components at each level helps us to identify critical features that influence Tutorial implementation. Some artifact, actors, and practices may appear in the descriptions of multiple levels of context. This is an indication that levels of context bleed into each other.

Of course, all of the actors involved in this educational system participate in other communities which constitute additional levels of context. Only academic levels of context that engage many of the participants associated with the Tutorial system are featured. In this representation, time or history becomes difficult to depict. The histories of particular actors are described during certain levels of context where their prior experiences seem particularly relevant.
The bulk of the Tutorial system can be described by examining five frames of context, listed from broader to most specific: 1) The institutional level—the broader University of Colorado system, interdepartmental collaborations, and partnerships between programs, 2) The departmental level—the CU physics department, 3) The course level—either Phys1 or Phys2, 4) The situation level—activities that make up the course such as lecture, recitations, or TA/LA training meetings etc., 5) The task level—activities that make up situations such as students completing a tutorial activity in recitations or TAs and LAs discussing student thinking in a Tutorial training meeting. Notably, the level usually described is the course level. However, the course level consists of many components (situations) which can combine in various ways to make up the course: the TA/LA training meetings, the lecture, the recitation sections, and various out-of-class activities such as grading, homework, exams, etc. The dotted arrows in Figure 22 depict possible efforts to coordinate these various course situations. In order to understand the use of Tutorials, two key situations (and key constituent tasks) highlighted in green (or shaded)
in Figure 22 are examined as well as aspects of the course, department, and institution that impact and are impacted by the Tutorials.

For each of these levels of context, I focus on describing 1) Key practices, 2) Key actors, and 3) Key artifacts. These three components define what is going on at each level [47,60,64] and constitute key levers for change in this educational system [65]. For example, one may attempt to modify this educational system by inserting new materials such as the Tutorials workbook, or by convincing university administrators of the importance of the LA program. After describing each level, aspects of the level that tend to vary, impacting Tutorial implementation, are summarized.

B. Tutorial Task and Situation Levels

At the level of Tutorial task, one would find a small group of students (2-5 students) working on answering specific questions in the Tutorials workbook possibly using simple equipment and interacting with their TA or LA(s). The TAs and LAs would be found to be walking around the room checking in with student groups and possibly asking questions about their reasoning for various answers. The key actors at the Tutorial task level include the students, the TAs, the LAs, and possibly the secondary professor, if the professor tends to visit the recitation section when the Tutorials were actually running. The presence of the secondary professor has been observed to vary greatly—some semesters the secondary professor may never visit the Tutorials being run with students while other semesters the secondary professor visits multiple Tutorial sessions every week. The key artifacts that shape the experience of participants at this level include the Tutorial activity workbooks, the simple topic-specific physics equipment, large white paper in the center of the groups’ table providing a collective writing space, and possible Tutorial handouts. The Tutorial tasks associated with Phys1 may be considered significantly different than the activities associated with Phys2 since the Phys1 activities rarely involve hands-on equipment. On the other hand, the vast majority of Phys2 Tutorials include simple equipment to be used in each activity.

Examining one level broader, the Tutorial situation level draws attention to patterns of interaction and activity across many Tutorial tasks and across multiple Tutorial sessions. During Tutorial sessions, TAs are typically found to spend the first few minutes of a recitation section making announcements—reminding students about upcoming exams and homework due dates, stating the Tutorial workbook pages to be completed during this session, and assigning the next week’s Tutorial homework. The TA may also be observed to be handing out graded Tutorial homework from the prior week, or even midterm exams. Significant variation in how TAs conduct these introductory remarks has been observed. Some TAs spend time reviewing or lecturing on common mistakes from the homework and others keep their comments very brief allowing students to immediately get started on the Tutorial activity. It would be at the situation level that we might expect the “elicit, confront, resolve” approach of the Tutorial activities to frame students’ experiences with the Tutorial. Students over multiple Tutorial situations will come to develop expectations about what their role is and what the role of their TA is based on collections of past interactions while completing the Tutorial. Student-TA interactions have been observed to vary greatly from section to section, as has been noted elsewhere [35]. The Tutorial recitations themselves occur in part of the basement-level laboratory space with temporary walls separating different recitation sections. With multiple ongoing sections, this space ends up being quite busy and fairly loud. At this level, key actors include the students, the TAs, the LAs, and possibly the secondary professor if he or she is present. Key artifacts at this
level include students’ completed Tutorial homework, TA-graded Tutorial homework, the TA’s handout with announcements on it constructed by the secondary professor, large tables for group work, moveable chairs, temporary walls, TAs’ & LAs’ Tutorial book with notes and their answers to particular Tutorial questions.

Although the personal histories of individual actors in these environments do not fit neatly into one level of context, the personal histories of the LAs and TAs are described here since it is at this level where TAs and LAs have the most control over establishing the norms within their Tutorial sections. Since the majority of TAs are first year graduate students, many of them have limited teaching experiences beyond tutoring. However, most first year graduate students are awarded a full-year long teaching assistantship meaning that TAs in the Spring semester may have previous experience teaching at CU. An additional complexity is whether the TA taught with the Tutorials during the Fall semester or with some other course. If the TA has prior experiences teaching with the Tutorials, then he or she may have developed ideas about how to teach with the Tutorials based both on their prior experiences in the classroom and through the training meetings with other secondary professors. The few TAs who choose to teach in their later years as a graduate student are more anomalous and their reasons for teaching may vary greatly from an interest in developing teaching expertise to a lack of research funding. The prior experiences of TAs are rarely documented systematically or shared with faculty within the course or the department. Similarly, the majority of LAs are first-time LAs although there are typically a few returning LAs. The LA program attempts to place returning LAs into new teaching environments for variety, but an LA returning from teaching in Phys1 may end up teaching as an LA in Phys2 during another term. Another complicating factor contributing to semester-to-semester variations is the addition of other educators to the Tutorial environments from the Teaching and Learning Physics course. This course is not offered every semester and these additional educators may have different backgrounds and a specific interest in teaching and learning based on their selection of this particular physics elective.

Students’ personal histories will influence their engagement in the Tutorials tasks and situations as well. Students arrive to the university and this particular course with specific expectations about what they think they should do and what the educators should do. Since the typical classroom is not organized as the Tutorials are, many students may experience a clash of their expectations and the expectations enacted in the Tutorial environment. Due to the discontinuous use of the Tutorials in Phys1, not all students taking Phys2 with the Tutorials have had prior experience learning physics in a Tutorial structure. Semesters where the majority of students have no experience in Tutorial-like learning environments are referred to as Tutorial newcomer semesters. Students’ prior experience with Tutorials or lack of prior experience with Tutorials may influence their interpretations and engagement in the Tutorial environments.

Key variations in implementation identified across the Tutorial task and situation levels include: variations in the reliance on equipment during the Tutorials activities between Phys1 and Phys2, participation of the secondary professor in the Tutorials while students are completing the Tutorials, prior teaching experiences of TAs & LAs, prior experiences of the enrolled students with Tutorial-like physics activities, and patterns of interaction between TAs/LAs and students.

C. Tutorial Training Meeting Task and Situation Levels

At the Tutorial training meeting task level, one would find small groups of TAs and LAs working on answering specific questions in the Tutorial activity, using simple equipment,
interacting with the secondary professor and possibly the lead TA, completing the Tutorial pretest, sharing and reflecting on the prior week’s Tutorial sessions, reading student responses to the pretest, and maybe discussing incoming students’ thinking based on old pretest responses with each other and the secondary professor. The secondary professor and lead TA might be found to be walking around the room checking in with groups of LAs and TAs, possibly asking questions about their reasoning for various answers, and drawing their attention to how they might support students working through these activities. The key actors at the task level of the Tutorial training meeting include the TAs, the LAs, the secondary professor, and the lead TA. Key artifacts at this level include Tutorial activity workbooks, simple topic-specific physics equipment, large white paper used as a collective writing space, and past student responses to Tutorial pretests.

Examining one level broader, the Tutorial training meeting situation level draws attention to patterns of interaction and activity across many Tutorial training meeting tasks and across multiple Tutorial training meetings. At the Tutorial training meeting situation level, one would find the secondary professor leading an initial discussion with TAs and LAs about the prior week’s Tutorial session, addressing concerns that TAs and LAs bring up from the past Tutorial session, describing a handout with this next week’s announcements on it, handing out copies of student responses to this week’s pretest, leading a discussion about common incoming student ideas for the next Tutorial, walking around the room checking in with groups of LAs and TAs as they complete the Tutorial, asking questions about TAs/LAs’ reasoning on the Tutorial, and possibly drawing TAs/LAs’ attention to how they might support students working through these activities. The Lead TA maybe involved in equipment maintenance and set-up or discussing the Tutorial activity with other TAs and LAs. The extent to which the Lead TA models Socratic questioning with the TAs and LAs varies depending on the preferences of the secondary professor and the Lead TA. Sometimes, the lecture professor is also present for some fraction of this training meeting. When this occurs, the lead and secondary instructors tend to spend some time synchronizing their efforts, sharing what material will be covered in lecture prior to the Tutorial activity that week. Many semesters the lecture professor will never be seen in the Tutorial training meetings. The TAs and LAs engage in many of the actions described in the Tutorial training meeting task level in addition to eating candy and snacks provided by the secondary professor. The instructional practices associated with the TA/LA training meetings will be discussed further though case studies presented in the next chapter.

Other factors contributing to the Tutorial training meeting situations is the presence of the initial adopter and PER researcher, Professor Arthur. Typically Professor Arthur attends the first two to three training meetings and gives a brief description to the TAs and LAs about the philosophy behind the Tutorials, encouraging the educators to not tell students the answer, but to ask students questions to help guide them to a more complete understanding. Although Professor Arthur is not present for many Tutorial training situations, his comments may be influential due to their placement at the beginning of the semester. The key actors at the situation level of the Tutorial training meeting include the TAs, the LAs, the secondary professor, the lead TA, Professor Arthur, and possibly the lead instructor. Key artifacts at this level include Tutorial activity workbooks, simple topic-specific physics equipment, large white paper used as a collective writing space, old student pretest responses, locally developed binder full of past materials used during the training meetings, the Tutorials instructor’s manual [2], and candy or snacks.

117
Key variations in implementation identified across the Tutorial Training Meeting task and situation levels include: presence of the lead instructor at the training meetings, participation of the lead TA in the training meetings, participation of the secondary professor in the training meetings, role of having the TAs/LAs take the Tutorial pretest, degree of emphasis placed on reflecting on the prior week’s Tutorial session, having the TAs and LAs learn the physics content, developing specific instructional moves or questions for interacting with students, and correcting wrong incoming ideas students have versus developing strategies for building on the useful intuitions that students bring to the class.

D. Course Level

At the course level, the professors for the course are engaged in designing the course syllabus and course website, drawing on materials from past courses, setting up grading policies, recording and assigning students’ grades, deciding on the appropriate division of labor between the team-teaching professors, planning and delivering lectures with corresponding clicker questions, planning the Tutorial activity schedule for the semester, setting office hours, writing exams, and constructing the computer-based homework assignments. Many of these activities help to establish the norms of the course and what students will perceive as valued in the course. There are semester-by-semester variations in the grading policies, clicker questions used during lecture, the division of labor between professors, the norms associated with lecture environment and the norms associated with the TA/LA training meetings. Some semesters the secondary professor may have no contact with students (not offering office hours, or attending Tutorial recitation sections) while other semesters the secondary professor may have extensive office hours and consistently visit Tutorial recitation sections.

Although the personal histories of individual actors in these environments do not fit neatly into one level of context, the personal histories of the physics professors are described here since it is at this level where physics professors have the most control over establishing the norms within their course. As mentioned before, physics professors have a long history of engagement in academia and as students they were unlikely to experience a Tutorial-like learning activity. These prior experiences both as students and professors may influence how they engage in teaching with the Tutorials. More recent experiences with teaching large enrollment physics courses may also influence physics professors’ initial reactions to using the Tutorials. A professor new to teaching large enrollment courses may use the Tutorials differently as they simultaneously learn to manage a large course. Similarly, professors with prior experiences teaching with the Tutorials may draw on these past experiences when designing their current course. Professors’ prior experiences with Tutorials may be particularly influential if significant faculty peer mentoring occurred during prior Tutorial use.

Another key actor at this level is the lead graduate teaching assistant which is supported both through the Graduate Teacher Program (GTP) and the Physics Department to play a mentoring and leadership role for other TAs. At the course level, the lead graduate teaching assistant is involved in maintaining the supply of equipment for the Tutorials, setting up the Tutorial equipment for each TA/LA training meeting as well as for the students’ recitation sections, and occasionally participating in the TA/LA training meetings as TA or as an additional educator modeling Socratic questioning as the TAs/LAs work through the Tutorial activities. Involvement of the lead TA during the actual TA/LA training meeting varies from semester to semester depending both on the lead TA and the preferences of the secondary professor. The
Lead TA is described at the course level, since their involvement in the course spans multiple course activities.

At the course level, students are engaged in attending lecture, attending Tutorial recitation sessions, and other out-of-class assignments associated with the course such as preparing for exams, taking exams, completing Tutorial pretests and homework, finishing computer-based homework, possibly reading the textbook for the course, and possibly reviewing the course lecture notes and clicker questions. At this level, the key actors include the lead and secondary professors, the Lead TA, and the students. Key artifacts at the course level include the course syllabus, the Tutorial activity workbook, the Tutorial homework book, the students’ personal clicker, the textbook, online course materials such as lecture notes, clicker questions, homework solutions, and/or exam solutions. Key variations at the course level include grading policies outlined in the syllabus, clicker questions used during lecture, the division of labor between professors, role of the Lead TA in the TA/LA meetings, the norms associated with lecture environment and the norms associated with the TA/LA training meetings.

E. Departmental Level

At the departmental level, the PER group is involved in initially bringing in and implementing the Tutorial curriculum, mentoring other instructors into teaching with the Tutorials, writing a local user’s guide for running the Tutorials, recruiting potential LAs, interviewing prospective LAs, hiring LAs for the upcoming semester, writing grants to external agencies to research and fund the LA program, advocating for LA Program funding to university administrators and private donors, assessing the impacts of course transformations on students, teaching a course on teaching and learning physics which places volunteer LAs into the intro courses, and presenting research results nationally and locally. Other physics faculty members are engaged in attending brown bag discussions about issues in education, partnering with the Faculty Teaching Excellence Program (FTEP) to support curriculum development and assessment of course transformations, mentoring other instructors into teaching with the Tutorials, sharing past course materials, supporting team-teaching assignments, and voting to award tenure (and promotion) to a PER professor to sustain the PER group locally. The department chair is engaged in supporting the Tutorials through the dedication of classroom space, hiring a PER tenure-track professor, contributing funding for LAs, providing financial support for equipment, partnering with the GTP, supporting the existence of the Lead TA, extending the position of the Lead TA, demonstrating enthusiasm for the Tutorial transformations by requesting teaching assignments in Tutorial courses, and supporting the creation of a course on Teaching and Learning Physics. Key actors at this level include the PER group faculty, past and current physics department chairs, and the physics faculty involved in teaching the introductory courses. Key artifacts at the departmental level include the course catalogue with physics course descriptions, meeting minutes from Brown Bag discussions on education [55], grant proposals, annual grant reports, and course artifacts such as exams, clicker questions, syllabi, past Tutorial activity schedules, and past online homework assignments. When examining course artifacts to understand course norms one must take caution since authorship of the document usually becomes ambiguous and difficult to pin down due to the departmental norm of sharing materials from prior implementations of the course. Changes at this scale tend to occur on longer time scales. Some of the shifts that occur at the departmental level can be seen in the timeline and associated historical analysis, see Figure 21.
F. Institutional Level

By examining the institutional frame of context, attention can be drawn to the partnerships between the physics department and other university programs and departments that were coincident with the implementation of the Tutorials. These partnerships include FTEP, GTP, the School of Education, and other STEM departments. One particularly important partnership is between the physics department and the School of Education (SoE). The SoE added a pedagogy course for first-time LAs when the LA program began. One SoE instructor and one high school master teacher have been teaching this SoE course since its inception (although during the Fall 2008 semester the SoE instructor changed). This course also serves LAs from other STEM departments. Interdisciplinary collaborations across the traditional departmental boundaries have contributed to expanding and sustaining the CU LA program which supports student-centered course transformations such as the Tutorials. Partnerships with the SoE and other STEM departments have lead to many successful grant applications which fund the LA program and support research on course transformation efforts associated with the LA program. Financial support from the Dean of Arts and Sciences, the Dean of the SoE, and the University Provost has been essential to sustaining the LA program thus to the implementation of the Tutorials. Key actors at the institutional level include the Dean of Arts and Sciences, the Dean of the SoE, the University Provost, PER advocates, other SoE collaborators, and Discipline-based Education Research (DBER) advocates in STEM departments. Key artifacts at the institutional level include interdisciplinary grant proposals, annual grant reports, budget records, LA course syllabi, LA recruitment materials, LA hiring applications, LA reflections associated with the SoE course, end-of-term posters from LAs, interdisciplinary education research meeting schedules, attendance sheets from interdisciplinary meetings, internal proposals to university administrators to sustain the LA program, published research articles and presentations from key actors, and email correspondence between key actors. As with the departmental level, changes at this scale tend to occur more slowly. Some of the shifts that occur at the institutional level can be seen in the timeline and associated historical analysis.

G. Summary of Frames of Context Analysis

A frames-of-context analysis allows us to identify key factors shaping the implementation of Tutorials at CU. This analysis highlights the broader institutional and departmental changes that have accompanied the adoption of the Tutorials at CU. Although these changes may allow (or inhibit) successful adoption, this analysis points out that these institutional and departmental changes tend to occur at time scales much longer than a given semester. For this reason, shifts that occur at these levels are unlikely to help us explain the variations in students’ learning from term to term.

The few research-based descriptions of educational change that exist tend to focus on the choices that individual faculty members make largely at the course level without sufficient attention to the situation and task levels which faculty members participate. We find that course level descriptions are insufficient to capture the rich nature of these changing environments. Particularly, in the case of the Tutorials there are at least two specific task and situation levels that constitute the course level: the TA/LA training meetings and the Tutorial recitations. The task and situation levels of both the TA/LA training meetings and the Tutorial recitations seem to vary greatly.

In the TA/LA training meeting, the focus of the activity can vary greatly ranging from a focus on the educators learning the correct physics and moving on, to an explicit focus on
specific questions to ask students at particular points in the activity with an emphasis on how to scaffold students’ conceptual understanding based on their incoming ideas. These different emphases can be seen in the relative amount of time spent discussing the educator’s own understanding of the materials, discussing incoming student ideas, and devising explicit series of questions to guide students.

In the Tutorial recitations themselves, TAs and LAs are largely responsible for establishing the expectations in the classroom. TAs and LAs may provide students with answers and/or extensive explanations to varying degrees. The interactions between the students and the educators in these environments may differently support a view that students and their peers can develop these answers and explanations for themselves. Similarly, some TAs and LAs may encourage students to blow through the Tutorials by not checking in with students about their reasoning and letting students leave early. Other TAs may consistently check-in with their students expecting them to articulate reasons for the answers that they came to.

At the course level, faculty members make key choices about if and how to connect the Tutorials activities and the lecture, how to assess Tutorial-related activities, how to design exams, and the timing of material introduced in lecture and Tutorials. At the departmental level, endorsement of these course practices through changing evaluation criteria and participation of the broader faculty as well as the dedication of resources stand out as salient. The PER group seems to have played an important role in the development and continued support for Tutorial use. At the institutional level, partnerships across STEM departments and the School of Education as well as efforts of interdisciplinary teams to secure external and internal university funding has also sustained these programs.

VII. Conclusions

This chapter has presented three coupled research studies: 1) Tutorial adoption and institutionalization at CU, 2) Individual faculty members’ tutorial adoption decisions, and 3) Tutorial implementation overview from a frames-of-context perspective. Drawing on an in-depth case study approach, these coupled studies contribute to a more nuanced model of educational change processes.

The first research study established that the CU physics department began the Tutorial adoption process building from: 1) a long legacy of dedicated educators at CU, 2) An existing informal structure to support faculty in discussing and reflecting on instructional approaches and student learning, 3) some preliminary faculty dissatisfaction with a traditional lecture approach, and 4) some preliminary success in adopting Peer Instruction in CU’s introductory physics courses. This analysis draws attention to the fact that purely describing the experiences of individual faculty members would not fully capture the complexity of this change process. Individual faculty members—the initial PER adopter, additional non-PER faculty members participating in implementing the Tutorials, and faculty advocates—played important roles in the adoption and institutionalization process, but other simultaneous changes throughout the educational system cannot be understated. These simultaneous changes throughout the education system included shifts in external funding, shifts in the dedication of departmental resources, shifts in support from university administrators, expanding partnerships across STEM departments, expanding partnerships between the School of Education and School of Arts and Sciences, and expanding partnerships between the physics department and other university programs across campus. By broadening our lens and examining faculty within the situations that they work, we see that structural changes in how things are done are coupled with changes in
how individual faculty members engage in teaching their courses. These findings call into question the common assumption of disseminations approaches that typically focus solely on individual faculty members’ adoption and individual use of curricular materials. These findings suggest that individual faculty members involved in these change processes may also need to learn to build partnerships and act as advocates within their institutional contexts to achieve sustainable change and that approaches to educational change might be more successful by coordinating and addressing multiple levels of the educational system simultaneously.

The second research study established that faculty members’ early decision to adopt the Tutorials was more complex due to team-teaching assignments. At times, a single faculty member decided about adopting the Tutorials independently from his/her co-teacher, while at other times team-teachers collaborated in deciding to adopt the Tutorials. Many faculty members openly admitted not having looked at the actual Tutorial activities prior to deciding to use them. Faculty members who felt involved in the decision to adopt the Tutorials expressed the following reasons for their decision: 1) Lecturing in recitation is largely ineffective, 2) Local data on student learning was compelling, 3) Tutorials relieve pressure from under-prepared TAs, 4) Specific pedagogical aspects of the curriculum such as peer discussion and attention to student misconceptions, and 5) The availability of help and resources in using the Tutorials. These results suggest: some initial individual dissatisfaction with current instructional methods may be critical in decisions to try something new, locally collected data maybe particularly compelling for other faculty members, and support and resources offered for the new methods may be persuasive (although the timing of these offers may be critical). Corroborating the findings from the first study, the results of this interview study suggests the importance of developing local structures that support reflective practice of faculty—specifically the sharing of assessment results. Encouraging reflective practices and the sharing of assessment results may be particularly supported by the presence of a physics education research group, although participation of and communication with the broader faculty also seems important.

The third analysis from a frames-of-context analysis demonstrated that adoption of the Tutorials required changes throughout a complex, nested educational system. This descriptive analysis suggests that a model of change (or claims about change) requires simultaneous attention to multiple levels of the educational system. At the same time, in studying and researching change, we must attend to the multiple levels and how these levels interact. Because this is a messy and complex system, we create a model for talking about components of the system. We create artificial boundaries for analytical purposes within this dynamic system of interacting variables. A frames-of-context framework allows us to identify key actors, practices and artifacts. While this system is complex, we expect semester-to-semester variations to be primarily influenced by three frames of context (task, situation, and course), as Cole argues the most significant influence on a given frame (situations where students complete the Tutorials activities) are the frames immediately above and below (course and task levels).

Based on frames-of-context analysis and our preliminary observations in Tutorial recitations, TA/LA Tutorial Training Meetings and associated course lectures, we expect that there is significant variation in how the Tutorials are implemented and integrated into the course. This analysis helps us to define Tutorial implementation—what does it involve and identify particular ways in which we might expect implementation to vary. For example there are choices that faculty make at the course level such as syllabus descriptions and the design of assessments that affect Tutorial implementation. There are choices at the level of the TA/LA training meetings where the secondary instructor’s practices frame what the purpose of the
Tutorial activities for TAs and LAs. There are also activities that exist at the same scale as the Tutorial recitations such as the lecture activity and out-of-class homework activities that may support the goals of the Tutorials or sit in tension with the Tutorials. All of these decisions constitute influence of physics faculty on Tutorials implementation.

We hypothesize that differences in Tutorial implementation will be visible in students’ perceptions of Tutorials reflected in end-of-term surveys. In the next chapter, we proceed to investigate whether students’ survey responses on end-of-term surveys about the utility and enjoyment associated with the Tutorials varies from semester-to-semester. Once semester-to-semester variations are identified in students’ multiple-choice survey responses, students’ long answer responses are examined to identify aspects of the Tutorials system that are particularly salient for students. We anticipate that these course level measures of students’ perceptions of Tutorials are associated with finer grained variation that may be observed at the Tutorial situation and task levels.

**Peer-reviewed Publications based on this work:**


References (Chapter 4)


[28] D.M. Hanson, Instructor’s guide to process-oriented guided-inquiry learning (Pacific Crest, Lisle, IL, 2006).


[55] Brown Bag Discussion minutes can be found at http://www.colorado.edu/physics/EducationIssues/resources/brownbags.htm


Chapter 5: Understanding the Students’ Perspectives on the Tutorials

I. Introduction

In this chapter, students’ end-of-term survey responses concerning the utility and enjoyment associated with the Tutorials [1] over five years of implementation at the University of Colorado (CU) are analyzed in order to 1) better understand the meaning students are making of this reform and 2) to understand the nature of variations in students’ perceptions from semester to semester. For example, students reported on the degree of sharing their reasoning with their peers in Tutorials and students reported on the utility of the Tutorials for their learning. Students’ perceptions of the Tutorials are considered important to understand because they influence how students choose to engage (or not engage) in Tutorial related activities. Additionally, early in the process of Tutorial adoption at CU, faculty members worried about (what was perceived to be) a large fraction of students complaining about the Tutorials. Faculty expressed concern over students’ negative response to the Tutorials and thought that if this negative experience were representative, the use of the Tutorials in introductory courses could be reducing students’ interest and motivation towards studying physics. Students’ survey responses are examined to investigate whether these vocal student complaints well-represent the perspectives of the larger student population. Students’ long answer responses on the end-of-term survey responses are also a rich starting point for understanding what about the Tutorials is particularly salient for the students.

In addition to investigating students’ overall positive or negative response to Tutorials, students’ survey responses are investigated to identify potential causes of semester-to-semester variations in students’ perceptions of Tutorials. This investigation is motivated by 1) demonstrated semester-to-semester variations in other outcome measures such as student conceptual learning [2] and 2) field notes from systematic observations of Tutorials implementation which suggests dramatic variations in Tutorial implementation (particularly in the TA & LA training meetings associated with the Tutorial courses) as discussed in the previous chapter. Even though surface features of Tutorial implementations are similar (i.e. the same worksheet materials, equipment, classrooms, training sessions, etc.), observations suggest that both the primary and secondary instructors’ decisions while implementing the Tutorials may influence students’ experiences with the Tutorials both directly and indirectly. For survey questions in which semester-to-semester variations are identified, we investigate whether different characteristic features of this system might account for the variations in students’ perceptions of Tutorials. Variations in student demographic characteristics are not found to largely account for variations in students’ perceptions of Tutorials. Although this is a complex system, it appears that instructors’ implementation decisions may account for variations in students’ perceptions.

Based on students’ long answer responses, aspects of Tutorials use which are salient for students are further investigated through the design of additional survey items. A second more detailed survey begins to investigate students’ perceptions of classroom norms, such as student-student interactions, student-TA interactions, and the coordination of course components, surrounding Tutorial use in an approach parallel to Peer Instruction investigations presented in Chapter 3. Through these investigations, we intend to uncover elements of classroom norms associated with Tutorials which could lead to fleshed out definitions of norms as part of further research.
II. Prior Research into Students’ Experiences in Tutorial Learning Environments

In trying to understand student learning, researchers have found that understanding student engagement is critical. Researchers have discovered that student engagement can be influenced by many aspects of the classroom context such as the tools available, the social relationships among participants, and the prior histories of individuals engaging in these tasks. Additionally, researchers have learned that understanding student engagement necessitates understanding how students experience learning tasks and environments. Researchers have devised a variety of related constructs and associated assessment instruments to better understand students’ own perspectives on their experiences learning physics. These instruments include: the Maryland Physics Expectations (MPEX) Survey [3], the Colorado Learning Attitudes about Science Survey (CLASS) [4], the in-class Conflict and Anxiety Recognition Evaluation (iCARE) [5], Motivated Strategies for Learning Questionnaire (MSLQ) [6], and typical midterm and end-of-term course evaluations. Each of these research areas are summarized since they all attempt to build an understanding of the meaning that students are making of physics and physics instruction. We focus particularly on work done to understand students’ experiences learning physics with curricular materials designed based on a cognitive conflict model of conceptual change, since this is the theoretical perspective adopted by the designers of the Tutorials curriculum [7].

A. Research into Students’ Attitudes and Beliefs about Physics and Learning Physics

Using a case study research design based on in-depth longitudinal student interviews, David Hammer established that students not only arrive to our introductory physics course with prior ideas about the physical world, but also prior ideas about school and the discipline physics that influence how students engage in physics learning activities [8]. Hammer suggests that our instruction may challenge or reinforce students’ prior ideas about the nature of learning and the nature of physics which may not necessarily be expert-like.

Building on Hammer’s prior work, Redish, Saul, and Steinberg developed the Maryland Physics Expectations (MPEX) survey to systematically probe student attitudes, beliefs, and assumptions about physics in large enrollment introductory physics courses [3]. This survey was used to explore students’ expectations about learning physics across six different institutions (each using a variety of different physics curricula—UW Tutorials [1], Workshop Physics [9], and University of Minnesota’s Cooperative Group Problem Solving [10]). Research conducted using MPEX established that students begin introductory physics courses with expectations that significantly differ from that of expert scientists (overall student responses range from 50%-60% favorable or expert-like) [3]. Additionally, it was found that students’ expectations tend to deteriorate (become less favorable) as a result of the first term of introductory physics at all institutions surveyed despite using different research-based curricula [3]. Although many of the curricula surveyed produce substantial gains in students’ conceptual understanding, these research-based curricula did not demonstrate similar gains in students’ expert-like attitudes and beliefs about physics (in fact they were often found to regress).

Building on the instrument design of the MPEX, researchers at the University of Colorado sought to design a new instrument—the Colorado Learning Attitudes about Science Survey (CLASS)—to assess similar aspects of students’ beliefs about physics and learning physics. This instrument sought to carefully validate survey statements, such that they correspond to a single interpretation by students, and to assess broader aspects of physics learning including personal interest, aspects of problem solving, and the coupled beliefs of sense-
making and effort [4]. Use of this instrument has corroborated the results found with the MPEX—most university teaching practices have a detrimental impact on students’ attitudes and beliefs about science [4]. Additionally, they showed that teaching practices which explicitly address students’ beliefs about physics can have clearly measurable effects such that students’ attitudes and beliefs do not deteriorate [4]. While using the Tutorials in CU’s intro physics courses, students’ shift in CLASS scores in Phys1 and Phys2 have been documented to range from slight improvements to no significant change to substantial declines [11,12].

Since the development of the MPEX and the CLASS, research has demonstrated that with curricular materials and instructional practices that give extensive attention to students’ beliefs about learning science and the nature of science, students’ attitudes and beliefs about physics can improve significantly over a single course [13,14]. These significant and substantial improvements in students’ attitudes and beliefs about physics have not been demonstrated while using the UW Tutorials in Introductory Physics.

To summarize, students’ attitudes and beliefs about science typically become less expert-like after the completion of a single semester of introductory physics. This decline has been documented in courses utilizing the UW Tutorials curriculum. With some attention to students’ attitudes and beliefs, educators and researchers have shown that it is possible to prevent the decline in students’ attitudes and beliefs about physics while still using the UW Tutorials curriculum. However, substantial improvements in CLASS or MPEX scores have not been documented while using the UW Tutorials. The relationship between students’ shift in attitudes and beliefs about science and particular curricula or instructional practices is not clearly understood. It appears that unless students’ attitudes and beliefs about physics are explicitly considered as learning goals in both instructional practices and the development of curricular materials, students’ attitudes will not improve and are likely to decline.

B. Student Anxiety & Motivation within Cognitive Conflict Driven Learning Environments

Research has shown that a cognitive conflict model of conceptual change may be significantly flawed or at least incomplete [15,16]. For example, the presence of cognitive conflict is not always sufficient for conceptual change to occur [17]. Additionally, researchers have called for a model of conceptual change that takes into account interest and motivation not only as they pertain to individual characteristics of students, but as they pertain to contextual features of the classroom [15]. Some PER researchers have sought to develop alternative models of conceptual change to guide their instruction [18,19], while others have sought to better understand the students’ affective experiences associated with learning in an environment designed based on a cognitive conflict model of conceptual change [5,20]. Here we focus on summarizing the results of research studies to better understand students’ affective experiences associated with learning in environments designed based on a cognitive conflict model of conceptual change, since the Tutorials curriculum was largely informed by a cognitive conflict model of conceptual change [7].

Researchers at Ohio State University developed a survey instrument, the in-class Conflict and Anxiety Recognition Evaluation (iCARE), to monitor students’ anxiety associated with cognitive conflict during in-class learning activities [5]. At the end of each weekly instructional meeting, this survey was administered to students enrolled in a Physics-by-Inquiry (PBI) course (a curriculum which was also designed by UW and explicitly utilizes an elicit-confront-resolve model of learning [21]). A large fraction of students (93% on average) reported experiencing a cognitive conflict situation in a PBI activity although the type of situation (individual or social)
that promoted the cognitive conflict varied by PBI activity. Students’ reactions to conflict shifted over the course of the semester: the student reaction of “interest” decreased and the student reaction of “trying to pay attention” increased over the term. It was found that the prevalence of high anxiety amongst the students varied by activity (ranging from 0%-21% of students) and those students who experienced high anxiety during the activity immediately prior to the mid-term exam performed less well on the midterm as compared to low anxiety students [5].

In a related research study at Ohio State University, associations between student motivation, student anxiety associated with cognitive conflict situations, and student conceptual understanding were investigated [20]. Three assessments were given: 1) Motivation assessment, a subset of the MSLQ targeting intrinsic goal orientation, task value, and self-efficacy [6], 2) Cognitive conflict and associated anxiety assessment, iCARE [5], and 3) Pre/post assessments around particular PBI activities. This research study showed that the level of students’ prior knowledge plays an important role in students’ development of scientifically correct conceptions (higher correctness on pretest was found to be strongly correlated with higher correctness on posttest). The results also indicate that students who had a high level of motivation and reacted to conflict by “attempting to revise their current theory” tended to reach the scientifically accepted understanding through cognitive conflict situations in a PBI course. On the other hand, despite having a high (or middle) level of motivation, those who reacted to conflict by “making multiple predictions,” “lacking of self confidence,” and “using past personal experience” tended to be less successful in reaching the scientifically accepted understandings [20] within a PBI course.

To summarize, students are likely to experience conflict situations in learning environments designed from an elicit-confront-resolve perspective. Some students are likely to experience anxiety associated with these conflict situations. Many students will not successfully resolve these conflict situations. Students’ experiences with conflict, anxiety, and a lack of resolution may all contribute to students’ positive or negative interpretations of the utility and enjoyment associated with the Tutorial learning environment.

C. Students’ Responses to Midterm and End-of-term Evaluations in Tutorial Learning Environments

Based on mid-semester evaluations from students using the Tutorials at USAFA, students primarily wanted one change—an answer key [22]. Students expressed that they were frustrated with “teaching themselves” [Ref. 22, pg. 1170]. Students wanted more explanations given and reassurance of knowing that an answer was correct before moving on. However, frustration with not having an answer key was not evident in the end-of-term evaluations. These researchers claim that by the end of the course students had “adapted to the new style of teaching/learning and had become less dependent on the teacher” [Ref. 22, pg. 1171]. This research suggests that students’ response to Tutorials may vary throughout the semester or term and better understanding students’ experiences in Tutorial-like learning environments may require investigating students’ experiences at multiple points throughout the term.

Based on end-of-term student evaluations of students using the Tutorials at USAFA, students’ response to the Tutorials was mixed. Seventy-five percent of students stated that the Tutorials were good or very good at helping them understand physics concepts. Sixty-nine percent of students stated that the Tutorials were good or very good at helping them meet the course objectives. However, only 35% of students stated that the Tutorials were good or very
Based on end-of-term student evaluations of students during the first semester of Tutorials use at CU, students’ reported an overall favorable response to the utility of the Tutorials and the student group work in Tutorials [23]. However, a sizeable fraction of students’ reported not enjoying the Tutorials experience (60% of students reported not enjoying the Tutorials). Students did however respond that they enjoyed working in groups and talking with their peers in Tutorials (70% of students reported enjoyment) [23]. We look to better understand students’ experiences in these courses utilizing the Tutorials curriculum to identify possible influences on students’ positive or negative experiences.

III. Description of Setting and Methods

A. Description of Survey Methodology and Courses Surveyed

We investigate how students perceive the Tutorials and associated classroom cultures by collecting student responses to two different types of online surveys designed locally: Survey 1—data from an instrument designed to identify students’ perceived utility and enjoyment of Tutorials, and Survey 2—more detailed survey data that targets students’ reflections on particular interactions suspected to be tightly coupled to the Tutorials, such as peer interactions and student-TA interactions in the Tutorials, as well as the coordination of the Tutorials with other course components. Each of these surveys includes both Likert-scale items as well as open-ended short answer questions. Initial information about the coarse-grained features of these environments was provided by Survey 1. Student long-answer responses from Survey 1 told us why students chose particular answer options and guided the design of Survey 2 to examine classroom norms of Tutorial implementation in more detail.

Survey 1 contains seven questions that inquire about the students’ broad course level perceptions of the Tutorials. These include 1) the perceived degree of help the Tutorials provided, 2) the students’ enjoyment of the Tutorials, 3) the students’ overall feeling about the Tutorials, 4) the students’ reported attendance at the Tutorials, 5) the shift or change in the students’ perception of Tutorials over the course of the semester, 6) the relative usefulness of the Tutorials as compared to traditional recitations and 7) the relative enjoyment of the Tutorials as compared to traditional recitations. In Survey 1, there were two broad questions that targeted students’ experiences of each of the following: working with their peers in Tutorials, working with equipment in the Tutorials, and working with educators in the Tutorials. These survey questions can be found in Appendix A, Table I.

Three themes emerged from students’ long answer responses on Survey 1 concerning their reasons for rating the utility and enjoyment associated with Tutorials: interacting with peers, interacting with their TA, and how the Tutorials fit in with the rest of the course. For examples of student comments that were coded under these themes, see Appendix A, Table III. Using student language from their own long answer responses, 6 to 7 new questions were designed to probe each of these themes. New items on Survey 2 were designed as Likert-scale items with answer options ranging from strongly disagree to strongly agree. These survey items can be found in Appendix A, Table II. In semesters in which Survey 2 was administered, these additional questions were added to the end of the previous Survey 1 questions and administered as a single survey.

For each semester of data, a pseudonym has been assigned to each professor as discussed in the previous chapter. Since each pseudonym begins with a different letter, the professors are
referenced in tables and figures by the first letter of their pseudonym. Generally two professors work on a given course. The professor listed first is in charge of the lecture portion of the course. The second faculty member listed is the ‘back-up’ instructor who has little face time with students, and is primarily in charge of Tutorial educator training, homework, and exams. This division of teaching responsibilities is discussed in more detail in the prior chapter. Table O describes the professors involved in teaching Phys1 and Phys2 with the Tutorials during the semesters when Tutorial surveys were administered to students. In Table O, semesters where the Tutorials were used are shaded in green.

Table O: (Color) Summary of courses surveyed regarding students’ experiences in Tutorials.

<table>
<thead>
<tr>
<th>Semester</th>
<th>Phys1</th>
<th>Phys2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2004</td>
<td>No Survey Data Collected</td>
<td>Professors Arthur (A) &amp; Felix (F) Survey 1 Administered</td>
</tr>
<tr>
<td>Sp 2005</td>
<td>No Survey Data Collected</td>
<td>Professors Felix (F) &amp; Harold (H) Survey 1 Administered†</td>
</tr>
<tr>
<td>Fall 2005</td>
<td>No Survey Data Collected</td>
<td>Professor Harold (H) Survey 1 Administered†</td>
</tr>
<tr>
<td>Sp 2006</td>
<td>Professors Jasper (J) &amp; Kent (K) Survey 1 Administered†</td>
<td>Professors Linus (L) &amp; Milo (M) Survey 1 Administered†</td>
</tr>
<tr>
<td>Fall 2006</td>
<td>Professors Nemo (N) &amp; Donald (D) Survey 1 Administered†</td>
<td>Professors Harold (H) &amp; Arthur (A) Survey 1 Administered</td>
</tr>
<tr>
<td>Fall 2008</td>
<td>Professors Stuart (S) &amp; Violet (V) Survey 1 &amp; 2 Administered†</td>
<td>Professors Ron (R) &amp; Milo (M) Survey 1 &amp; 2 Administered</td>
</tr>
<tr>
<td>Sp 2009</td>
<td>Professors Felix (F) &amp; Wilbur (W) Survey 1 &amp; 2 Administered†</td>
<td>Professors Harold (H) &amp; Nemo (N) Survey 1 &amp; 2 Administered</td>
</tr>
</tbody>
</table>

†Tutorial newcomer semester: Majority of students have no experience with physics Tutorials.

Table O also shows that between fall 2004 and spring 2009 Survey 1 was administered periodically in both Phys1 and Phys2. Survey 2 was administered in both Phys1 and Phys2 during the 2008-2009 academic year. Not all students in Phys2 have had prior experience learning physics in a Tutorial environment in Phys1 due to the early discontinuous process of adopting the Tutorials [24] as shown in the timeline in the prior chapter. We will refer to Phys1 and Phys2 semesters in which the majority of students have no prior experience with physics Tutorials as Tutorial newcomer semesters. The F&H, H, and L&M semesters of Phys2 as well as all semesters of Phys1 are Tutorial newcomer semesters as indicated in Table O.

The surveys were administered during the 13th and 14th weeks of the 15-week semester. Surveys were announced in the lecture portion of the course and a link to the online survey was placed in a prominent location on the course web page. The surveys were optional, although students’ completion of the survey was typically rewarded by awarding some small amount of participation credit (affecting less than 1% of a student’s overall grade). Surveys were available online and no class time was devoted to allowing students to complete the surveys.

Survey response rates were calculated by beginning with the instructors’ grade sheets. From these grade sheets, a list of students in the course was generated. Demographic information and course grade information was gathered from the CU’s Planning, Budget and Analysis for students in the course. Students who dropped the course, withdrew from the course, or took an incomplete in the course were not included in the number of students enrolled in the course since they were not participating in the course by the end of the term. The percentage of students—receiving a withdrawal or incomplete in these course—ranges from 0.7% to 8.2%.
The number of students completing the survey and associated response rates for each course can be found below in Table P.

### Table P: Student survey response rates on Tutorial surveys administered.

<table>
<thead>
<tr>
<th>Term</th>
<th>Lead &amp; Secondary Professors</th>
<th>Course</th>
<th>Survey</th>
<th>N students enrolled</th>
<th>N survey respondents</th>
<th>Response Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa04</td>
<td>A&amp;F Phys2</td>
<td>1</td>
<td>461</td>
<td>361</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td>Sp05</td>
<td>F&amp;H Phys2</td>
<td>1</td>
<td>322</td>
<td>252</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td>Fa05</td>
<td>H Phys2</td>
<td>1</td>
<td>444</td>
<td>290</td>
<td>65%</td>
<td></td>
</tr>
<tr>
<td>Sp06</td>
<td>J&amp;K Phys1</td>
<td>1</td>
<td>491</td>
<td>103</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>Sp06</td>
<td>L&amp;M Phys2</td>
<td>1</td>
<td>351</td>
<td>216</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Fa06</td>
<td>N&amp;D Phys1</td>
<td>1</td>
<td>569</td>
<td>316</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>Fa06</td>
<td>H&amp;A Phys2</td>
<td>1</td>
<td>422</td>
<td>313</td>
<td>74%</td>
<td></td>
</tr>
<tr>
<td>Fa08</td>
<td>S&amp;V Phys1</td>
<td>1&amp;2</td>
<td>604</td>
<td>466</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>Fa08</td>
<td>R&amp;M Phys2</td>
<td>1&amp;2</td>
<td>451</td>
<td>322</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>Sp09</td>
<td>F&amp;W Phys1</td>
<td>1&amp;2</td>
<td>592</td>
<td>430</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>Sp09</td>
<td>H&amp;N Phys2</td>
<td>1&amp;2</td>
<td>433</td>
<td>280</td>
<td>65%</td>
<td></td>
</tr>
</tbody>
</table>

In the data presentation that follows, data from the spring 2006 semester of Phys1 have been excluded due to the anomalously low response rate this semester. For all semesters of data presented, the student survey response rates range from 56% to 78%. Since all students did not take the survey, this increases the possibility that survey respondents represent a non-random sample which adds additional uncertainty to the data. Past analyses of student surveys administered at the University of Colorado have shown that students who complete surveys tend to receive slightly higher grades in the course on average than students who do not complete the surveys [25]. The results described may not well represent lower performing students and care must be taken when generalizing to the entire student population. The data presented in this chapter spans ten implementations of Tutorials across two courses—three in Phys1 and seven in Phys2.

### B. Variations in Course Characteristics between Phys1 & Phys2

There are some important differences to be aware of between Phys1 and Phys2. The first most obvious difference is that the content being covered in the courses is different. Phys1110 primarily covers introductory mechanics including: “kinematics, dynamics, momentum of particles and rigid bodies, work and energy, gravitation, simple harmonic motion, and introduction to thermodynamics” [26]. Phys 1120 primarily covers electricity and magnetism including: “electricity, magnetism, wave motion, and optics” [26]. Differences in physics course content may be relevant in understanding students’ experiences in these courses.

Another important difference between Phys1 and Phys2 is the degree to which the UW Tutorial activities rely on hands-on equipment. In Phys1, there are very few Tutorial activities that utilize hands-on equipment. In contrast, most of the Phys2 Tutorial activities utilize hands-on equipment extensively. The use of equipment in the Phys2 Tutorials may influence both the nature of the tasks that the students are engaged in as well as influence how students perceive the Tutorial tasks.
The size of the course and the distribution of majors enrolled in the course varies between Phys1 and Phys2. Over the past five years approximately 575 students take Phys1 every semester, while only about 400 students take Phys2 every semester. Two lecture section of Phys1 are typically offered in both the fall and spring, while in Phys2 the fall semester offers two lecture sections and the spring only offers one lecture section. The size of an average lecture section of Phys1 (N~290) is roughly comparable to that of Phys2 (N~250). The distribution of majors shifts slightly between Phys1 and Phys2. Phys1 has a slightly higher fraction of non-science majors and undeclared students and a slightly lower fraction of engineering majors as compared to Phys2.

Overall trends across Phys1 and Phys2 may be mentioned briefly. However, semester-to-semester variations are examined separately within each course due to the significant differences between the two courses (i.e. course material, student familiarity with the material, use of equipment in the Tutorial activities, and major demographics).

C. Statistical Methods for Analysis of Student Survey Data

A variety of statistical methods are used to identify semester-by-semester differences in student responses on 13 Tutorial-specific survey questions. Student responses from the same course (i.e. Phys1) are compared across different semesters (i.e. Fa2006, Fa2008, & Sp2009) to identify possible variations. Some of these survey questions provide answer options that are categorical and others provide answer options that are rank-ordered categories. For survey questions which are categorical, but not rank ordered, we use a Chi-squared Test for Independence [Ref. 27, pg. 204] to identify variations [28]. For survey questions which are categorical and rank ordered, we use a Kruskal-Wallis Test for k-independent samples [Ref. 27, p. 288] to identify variations [29]. The statistical test used for each survey question is given in Appendix B, Table I (for Phys1) and Table III (for Phys2). For the initial statistical tests on the set of all semesters within a course, p-values less than or equal to 0.05 are taken to be statistically significant.

After survey questions that vary across Phys1 or Phys2 semesters are identified, the questions that show statistically significant differences are analyzed further. On questions which show statistically significant differences, we compare semesters in a pair-wise fashion (using either pair-wise Chi-squared Test for Independence or pair-wise Mann-Whitney U Test [Ref. 27, pg. 272], depending on whether the answer options are rank-ordered or not). When determining the statistical significance of the pair-wise comparisons, we decrease the threshold for statistical significance depending on the number of pair-wise semester comparisons for that question (p-value significance threshold equals 0.05 divided by the number of pairs). The resultant threshold for significance is given by question in Appendix B, Table I (for Phys1) & Table III (for Phys2). The pair-wise p-values by question (where pair-wise comparisons are justified) are given in Appendix B, Table II (for Phys1) and Table IV (for Phys2).

For statistical analyses of Survey 2, only two semesters of Phys1 and two semesters of Phys2 are being compared. For these comparisons, a Mann-Whitney test is used to determine if one semester tends to report more favorably than another semester. For the new survey 2 questions, a p-value less than 0.05 is considered significant. In order to give a sense of both the magnitude and the direction of the differences across semesters, a mean is calculated by applying a continuous numerical scale to the categorical answer options (1-being the least favorable and 5-being the most favorable). These mean values of students’ responses are provided as a general
heuristic, but the answer options are treated as categorical to determine statistically significant differences across semesters.

IV. Longitudinal Analyses of Students’ Broad Perceptions of Tutorials
A. Identifying Semester-to-Semester Variations in Students’ Responses to Survey 1 Questions

Within the data set of three Phys1 semesters, there are ten survey questions which were asked across at least two semesters. Eight were asked in all three semesters—five of which were identified as varying significantly from semester-to-semester (p<0.05 via Kruskal-Wallis Test). On the remaining two questions which were not asked over all semesters, only one was found to vary significantly (p<0.05 via Mann-Whitney U-Test). Overall, students’ survey responses were shown to vary significantly on six out of ten survey questions demonstrating statistically significant semester-to-semester variations within three semesters of Phys1. Specific p-values are list by question in Appendix B, Table I.

Within the data set of seven Phys2 semesters, there are thirteen survey questions which were asked across at least two semesters. Overall, students’ survey responses in Phys2 were shown to vary on eleven out of thirteen survey questions demonstrating significant and consistent semester-to-semester variations across seven semesters of Phys2. Three of these questions were tested using a Chi-squared test and one showed statistically significant semester-to-semester differences (p<0.05). The remaining ten questions all demonstrated statistically significant semester-to-semester variations (p<0.05 via Kruskal-Wallis). P-values are listed by question in Appendix B, Table III.

Although it was fairly straightforward to establish semester-to-semester variations in students’ survey responses, it is less straightforward to parse what factors may be contributing to these variations. Two survey questions are presented to provide concrete examples of the magnitude of the variation in students’ responses. In this data presentation (Figure 23 & Figure 24), the earliest implementation of Tutorials is plotted to the far left of each course listed and the most recent implementation to the far right. Hypothesized factors that may contribute to these semester-to-semester variations are explored in following sections of this chapter.

B. Examples of Semester-to-Semester Variations on Survey 1 Questions
1. Helpfulness of Tutorials for Student Learning (Q8)

Students were asked to “Please rank how much each of the following class activities helped your learning in the course of this semester: Tutorials” (1—No help, 2—A little help, 3—Moderate help, 4—Much help, 5—Very much help). Although these answer options are categorical, they are also rank-ordered. To simplify the presentation of these data, we have clustered the ‘No help’ category and the ‘A little help’ category into one group labeled ‘Low’ help (L). Similarly, the ‘Much help’ category and the ‘Very much help’ category were clustered into one group labeled ‘High’ help (H). Then the ‘Moderate help’ category was labeled ‘Medium’ help (M). The student responses for each semester of data are provided below in Figure 23.
Figure 23: Student Perceptions of the Degree of Helpfulness of the Tutorials for Student Learning (Q8). While the semesters of Phys1110 are statistically indistinguishable, there is significant variation in the perceived helpfulness of the Tutorials between semesters of Phys2.

Figure 23 shows that how students perceive the helpfulness of the Tutorials for their learning varies from semester to semester. At times, as many as 49% of students are reporting that the Tutorials were of ‘Low’ helpfulness (N&D semester of Phys1) while in other semesters as much as 51% of students are reporting that the Tutorials were of ‘High’ helpfulness (H&A semester of Phys2).

When we examine the three semesters of Phys1 data available, we see that there are no significant differences in the five-answer distribution of student responses (p=0.084 via Kruskal-Wallis Test). On this question, student responses from these three semesters of Phys1 are indistinguishable. When we examine the seven semesters of Phys2 data available, we see significant differences in students’ responses (p<0.001 via Kruskal-Wallis Test). Students from the A&F, H&A, and R&M semesters of Phys2 reported that the Tutorials were more helpful than students from the F&H, H, and H&N semesters of Phys2 semesters (all pair-wise p-values>0.001 via Mann-Whitney U-Test). In the A&F, H&A, and R&M semesters of Phys2, far more students are reporting that the Tutorials were of ‘High’ helpfulness than are reporting that the Tutorials were of ‘Low’ helpfulness (%H-%L ≥ 19%). However, in the F&H, H, and H&N semesters of Phys2, more students are reporting that the Tutorials were of ‘Low’ helpfulness than are reporting that the Tutorials were of ‘High’ helpfulness (%H-%L ≤ -5%). Specific pair-wise p-values are provided in Appendix B, Table IV.
2. Enjoyment of Tutorials (Q41)

Students were asked to “Please rank how much you enjoyed each of the following class activities in the course this semester: Tutorials (this refers to the Tues and Thurs recitation sections)” (1—Strongly disliked it, 2—Disliked it a bit, 3—Neutral, 4—Liked it a bit, 5—Very much enjoyed it). Although these answer options are categorical, they are also rank ordered. To simplify the presentation of these data we have clustered the ‘Strongly disliked it’ category and the ‘Disliked it a bit’ category into one group labeled ‘negative’ (-). Similarly, the ‘Liked it a bit’ category and the ‘Very much enjoyed it’ category were clustered into one group, labeled ‘positive’ (+). Then the ‘Neutral’ category label was left as ‘neutral’ (0). The student responses for each semester of data are provided below in Figure 24.

![Figure 24: Student Perceptions of the Enjoyment associated with the Tutorials (Q41). Most semesters at CU, students report disliking the Tutorials more than they report liking the Tutorials. Semesters of Phys1 are statistically indistinguishable, while there is significant variation in the reported enjoyment associated with Tutorials between semesters of Phys2.](image)

We can see from Figure 24 that at times as much as 57% of students are reporting that they disliked the Tutorials (F&W semester of Phys1) while in other semesters as little as 34% of students are reporting that they disliked the Tutorials (R&M semester of Phys2). All semesters of survey data but two (H&A and R&M) show that more (at least 10% more) students report disliking the Tutorials than report liking the Tutorials as shown in Figure 24. We see that in general students at CU usually report disliking the Tutorials more than they report liking them. In contrast to the utility question where we see an example where the majority of students report
finding the Tutorials helpful, in examining students’ perceived enjoyment the most positive semester shows 39% of students liking the Tutorials (R&M semester).

When we examine the three semesters of Phys1 data available, we see that there are no significant differences in the five-answer distribution of student responses (p=0.184 via Kruskal-Wallis Test). On this question, student responses from these three semesters of Phys1 are indistinguishable with more than half of the students reporting a dislike for the Tutorials. When we examine the seven semesters of Phys2 data available, we see significant differences in the five-answer distribution of student responses (p<0.001 via Kruskal-Wallis Test). Students from R&M and H&A semesters report more strongly liking the Tutorials than students in the F&H, H, H&N semesters (all pair-wise p-values ≤ 0.002 via Mann-Whitney U-Test). R&M’s students also report liking the Tutorials more than students in A&F’s semester (p=0.001 via Mann-Whitney U Test). Specific pair-wise p-values are provided in Appendix C.

3. Summary of Trends on other Survey 1 Questions

Students’ Perceptions of Discussing with Student Peers in Tutorials

Unlike the overall utility and enjoyment questions about the Tutorials in general, students’ reported utility of discussing with their peers in Tutorials varies in one of the three pairs of Phys1 semesters compared. In Phys1, the percent of students reporting that discussing with their peers was helpful or highly helpful ranged from 34%-43%. Phys1 students tend to report more favorably about the utility of discussing with their peers than they do about the Tutorials more generally (presented in Figure 23). Students’ reported utility of discussing with their peers in Tutorials varies in nine of the twenty-one pairs of Phys2 semesters compared. In Phys2, the percent of students reporting that discussing with their peers was helpful or highly helpful ranged from 36%-60%.

Students’ reported enjoyment of discussing with their peers in Tutorials varies in one of the three pairs of Phys1 semesters compared. In Phys1, the percent of students reporting that liked or strongly liked discussing with their peers in Tutorials ranged from 40%-52%. Students’ reported enjoyment of discussing with their peers in Tutorials varies in seven of the twenty-one pairs of Phys2 semesters compared. In Phys2, the percent of students reporting that liked or strongly liked discussing with their peers in Tutorials ranged from 37%-50%. In both Phys1 and Phys2, students tend to report more favorably about the enjoyment associated with peer discussions in Tutorials than they do about the Tutorials more generally (presented in Figure 24).

Students’ Perceived Utility of Contact with Tutorial Educators

Students’ reported utility of contact with their graduate TA in Tutorials varies in one of the three pairs of Phys1 semesters compared. In Phys1, the percent of students reporting that contact with their graduate TA was helpful or highly helpful ranged from 36%-44%. Students’ reported utility of contact with their undergraduate LA in Tutorials varies in three of the three pairs of Phys1 semesters compared. In Phys1, the percent of students reporting that contact with their undergraduate LA was helpful or highly helpful ranged from 36%-53%. Students’ perceived utility of LAs tends to vary more than students’ perceived utility of TAs across these three semesters of Phys1. Phys1 students tend to report more favorably about the utility of contact with their TAs and LAs than they do about the Tutorials more generally (presented in Figure 23).

Students’ reported utility of contact with their graduate TA in Tutorials varies in eight of the twenty-one pairs of Phys2 semesters compared. In Phys2, the percent of students reporting
that contact with their graduate TA was helpful or highly helpful ranged from 39%-64%.
Students’ reported utility of contact with their undergraduate LA in Tutorials varies in only one
of the twenty-one pairs of Phys2 semesters compared. In Phys2, the percent of students
reporting that contact with their undergraduate LA was helpful or highly helpful ranged from
44%-54%. Interestingly the converse trend is found in Phys2, students’ perceived utility of the
LAs tends to vary less than students’ perceived utility of TAs across the seven semesters of
Phys2. Phys2 students tend to report more favorably about the utility of contact with their TAs
and LAs than they do about the Tutorials more generally (presented in Figure 23).

Students’ Perceptions of Working with Equipment in Tutorials
Survey questions about equipment use in Tutorials were rarely asked in Phys1 due to the
rarity of equipment use. Students’ reported utility of working with equipment in Tutorials varies
in eleven of the twenty-one pairs of Phys2 semesters compared. In Phys2, the percent of
students reporting that working with equipment in Tutorials was helpful or highly helpful ranged
from 38%-70%. Students’ reported enjoyment of working with equipment in Tutorials varies in
eight of the twenty-one pairs of Phys2 semesters compared. In Phys2, the percent of students
reporting that they liked or strongly liked working with equipment in Tutorials ranged from
42%-66%.

Overall Patterns of Similarity or Difference between professors across many semesters
Here we address the question: Which teaching teams tend to differ from other teaching
teams over a large number of questions and which tend to be consistently the same? Now we will
pay less attention to the question-by-question survey results and focus on the entire Tutorial
survey. We will say that two semesters (and their associated team teaching pairs) are NOT
significantly different overall (i.e. similar) if less than 20% of the survey questions show
statistically significant differences. We will say that two semesters are significantly different
overall if 50% or more of the survey questions show statistically significant differences.

We examined 3 semesters of Phys1 [N&D, S&V, and F&W], resulting in 3 pair-wise
comparisons. In Phys1, student responses from the Nemo & Donald (N&D) semester are similar
to the student responses of both the Stuart & Violet (S&V) semester and the Felix & Wilbur
(F&W) semester. However the student responses from the Stuart & Violet (S&V) semester are
consistently different from the student responses of the Felix & Wilbur (F&W) semester (student
responses vary on five of ten questions).

resulting in 21 pair-wise comparisons. In Phys2, the Linus & Milo (L&M) semester, Felix &
Harold (F&H) semester, Harold semester (H), and the Harold & Nemo (H&N) semester are
similar to each other (fewer than 20% of questions show statistically significant differences).
Additionally, the Phys2 two semesters of Linus & Milo (L&M), Ron & Milo (R&M), and Harold
& Arthur (H&A) are similar (fewer than 20% of questions show statistically significant
differences). The student responses from the Linus & Milo (L&M) semester are similar to the
Arthur and Felix (A&F) semester (fewer than 20% of questions show statistically significant
differences). However, the Phys2 semesters of Felix & Harold (F&H), Harold (H), and Harold
& Nemo (H&N) each differs from Harold & Arnold (H&A) and Ron & Milo (R&M)
(statistically significant differences found on 50% or more of the survey questions).
Additionally, student responses from the Arnold and Felix (A&F) semester differ from both the
Felix & Harold (F&H) semester and the Harold (H) semester (statistically significant differences found on 50% or more of the survey questions).

Since there are so many pairs of Phys2 semesters being compared, we can represent these similarities and differences using a Venn diagram. If semesters are similar then they are represented within the same circle, if semesters are different then are represented in the non-overlapping region of the circles. These Venn diagrams are provided below in Figure 25. For example, the first diagram shows that 1) H&A, R&M and L&M are each similar to each other, 2) F&H, H, H&N, and L&M are each similar to each other, and 3) H&A and R&M are each different than F&H, H, and H&N.

![Figure 25: Venn diagrams showing similarities and differences between team teaching pairs across Phys2 semesters.](image)

This analysis has shown that there are significant semester-to-semester variations in students’ survey responses about the Tutorials as well as described the magnitude of these variations which at times vary by as much as 20%. The data analysis that follows investigates some possible sources of these variations. The following hypothesized influences are addressed in turn: possible variations in the backgrounds of the students’ enrolled in the course, possible variations in the TAs assigned to work with the course, variations in the primary instructor teaching the course, and variations in the secondary instructor co-teaching the course.

C. Examining Semester-to-Semester Variations in Student Background Factors

One possible hypothesis might be that the semester-to-semester variations in students’ survey responses are due to fluctuations in the demographics and prior experiences of the students’ enrolled in the course. This demographic information was gathered through CU’s Planning, Budget and Analysis. For the Phys1 populations, the following student background factors are explored: gender make-up of the course, ethnicity make-up of the course, major demographics, average number of years of physics taken in high school, Colorado Learning Attitudes about Science Survey (CLASS) [30] overall favorable pretest score, and Force and Motion Conceptual Evaluation (FMCE) [31] pretest scores. CLASS and FMCE pretest information is not available for all students, so will only roughly approximate the backgrounds of the entire student population of the courses. For the Phys2 populations, the following student background factors are explored: gender make-up of the course, ethnicity make-up of the course, major demographics, average number of years of physics taken in high school, average CLASS overall favorable pretest score, average incoming FMCE posttest scores, and average Brief
Electricity and Magnetism Assessment (BEMA) [32] pretest scores. As mentioned before, data on students’ FMCE posttest, CLASS and BEMA pretest information is not available however for all students, so will only roughly approximate the backgrounds of the entire student population of the courses. First the variations in student background factors across Phys1 semesters will be discussed, followed by a parallel discussion for the Phys2 semesters.

1. Comparing Student Background Factors across Semesters of Phys1

The Phys1 course tends to consist of approximately 24% female students and 76% male students and there are no statistically significant differences in the gender make-up of the student population between Phys1 semesters (p=0.167 via chi-squared test). The Phys1 course tends to consist of about 82% white students, 9% Asian students, 5% Hispanic students, 3% foreign students, 1% African American, and less than 1% Native American. There are no statistically significant differences in the ethnic make-up of the student population within Phys1 semesters (p=0.924 via chi-squared test).

In our sample of Phys1 semesters, there are significant differences in the distribution of student majors enrolled in the course (p<0.001 via chi-squared test). However, when we compare the two fall semesters (Fa2006—N&D & Fa2008—S&V) we find that the distribution of majors is indistinguishable. The only differences are between the fall semesters and the spring semester. During the spring semester, there tends to be a lower percentage of physics majors and a higher percentage of engineers as compared to fall semesters.

Three measures are used to compare students’ incoming physics-specific knowledge or prior experiences: their average pretest score on the FMCE, the average number of years of high school physics taken, and their average overall favorable CLASS pretest score. In our sample of Phys1 semesters, there are no statistically significant differences in students’ FMCE pretest scores (p=0.666 via ANOVA). There are significant differences in the mean number of semesters of HS Physics taken (p=0.035 via ANOVA). Through a Tukey post-hoc test, the only significant difference that was found was between the Fa2006 semester and the Sp2009 semester (p=0.029). All other semesters are statistically indistinguishable. Statistically significant differences between semesters were also found on students’ average overall favorable CLASS pretest score (p<0.001 via ANOVA). The average overall favorable CLASS pretest score for Fa2006, Fa2008, and Sp2009 were found to be 68%, 66%, and 63% respectively. Through a Tukey post-hoc test, significant differences were found between the Fa2006 and Sp2009 semesters (p<0.001) as well and between the Fa2006 and Fa2008 (p=0.046).

Based on these analyses, similar student populations can be identified as well as variations to keep in mind as course comparisons are made. In Phys1, the background factors of students in N&D’s course are statistically indistinguishable from S&V’s course on all factors investigated. Similarly in Phys1, the background factors of students in S&V’s course are statistically indistinguishable from F&W’s course on all factors except for students’ major and average overall favorable CLASS pretest score. More care must be taken when comparing N&D’s course to F&W’s course as multiple student background factors were found to vary significantly (including students’ major, HS physics experience, and average overall CLASS pretest scores). For semesters of Phys1, significant variations in student background factors are summarized in Appendix C, Table I.

2. Comparing Student Background Factors across Semesters of Phys2
The Phys2 course tends to consist of approximately 25% female students and 75% male students and there are no statistically significant differences in the gender make-up of the student population between Phys2 semesters (p=0.404 via chi-squared test). The ethnic make-up of the course does not fluctuate greatly. Phys 2 student population tends be between 76-85% white, between 5-14% Asian, 3-9% Hispanic, and less than 4% African American, Native American, or foreign. Since the number of students from many of these ethnic groups is so small, a Chi-squared test becomes inappropriate. Although statistics cannot be run on these differences, the distributions tend to vary by less than 10% which is likely to be a small effect.

In our sample of Phys2 semesters, there are significant differences in the distribution of student majors enrolled in the course (p<0.001 via chi-squared test). When semesters of Phys2 are compared in a pair-wise fashion and the p-value thresholds decreased accordingly, the only significant differences were found to be between Spring & Fall semesters (p ≤ 0.002). During the spring semesters, there tends to be a higher percentage of physics majors and other science majors and a lower percentage of engineers as compared to fall semesters.

Four measures are used to compare students’ incoming physics-specific knowledge or prior experiences: their average pretest score on the BEMA, the average number of years of high school physics taken, their average posttest score on the FMCE, and their average overall favorable CLASS pretest score. In our sample of Phys2 semesters, there are no statistically significant differences in students’ BEMA pretest scores (p=0.058 via ANOVA) or in the mean number of semesters of HS Physics taken (p=0.874 via ANOVA). Statistically significant differences between semesters were found on students’ average FMCE posttest score (p<0.001 via ANOVA) and students’ overall favorable CLASS pretest score (p<0.001 via ANOVA). These average scores by semester are given in Table Q.

### Table Q: Student background factors that vary from semester-to-semester in Phys2. The standard error on the mean is ~2% on FMCE posttest scores. The standard error on the mean is 1-2% on CLASS pretest scores.

<table>
<thead>
<tr>
<th>Term</th>
<th>Lead &amp; Secondary Professors</th>
<th>Survey</th>
<th>N students enrolled</th>
<th>Avg. FMCE Posttest Score (%)</th>
<th>Avg. Overall CLASS Pretest Score (% Fav.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa04</td>
<td>A&amp;F</td>
<td>1</td>
<td>461</td>
<td>77%</td>
<td>70%</td>
</tr>
<tr>
<td>Sp05</td>
<td>F&amp;H</td>
<td>1</td>
<td>322</td>
<td>75%</td>
<td>66%</td>
</tr>
<tr>
<td>Fa05</td>
<td>H</td>
<td>1</td>
<td>444</td>
<td>65%</td>
<td>61%</td>
</tr>
<tr>
<td>Sp06</td>
<td>L&amp;M</td>
<td>1</td>
<td>351</td>
<td>63%</td>
<td>64%</td>
</tr>
<tr>
<td>Fa06</td>
<td>H&amp;A</td>
<td>1</td>
<td>422</td>
<td>63%</td>
<td>64%</td>
</tr>
<tr>
<td>Fa08</td>
<td>R&amp;M</td>
<td>1&amp;2</td>
<td>451</td>
<td>58%</td>
<td>62%</td>
</tr>
<tr>
<td>Sp09</td>
<td>H&amp;N</td>
<td>1&amp;2</td>
<td>433</td>
<td>66%</td>
<td>NA</td>
</tr>
</tbody>
</table>

The A&F semester’s average FMCE posttest scores was found to be statistically higher than that from the H, L&M, H&A, R&M, and H&N semesters (p ≤ 0.05 via Tukey post-hoc test). The F&H semester’s average FMCE posttest score was found to be statistically higher than that from the H, L&M, H&A, and R&M semesters (p ≤ 0.05 via Tukey post-hoc test). The H&N semesters’ average FMCE posttest scores was found to be statistically higher than the R&M semester (p=0.004 via Tukey post-hoc test). All other pairs were found to have FMCE posttest score averages that were statistically indistinguishable. Upon further investigation, A&F semester’s average overall favorable CLASS pretest score was found to be significantly higher than that from the H, L&M, H&A, and R&M semesters (p ≤ 0.006 via Tukey post-hoc test). All
other pairs were found to have overall favorable CLASS pretest score averages that were statistically indistinguishable.

Based on these analyses, similar student populations can be identified as well as variations to keep in mind as course comparisons are made. Across Phys2 semesters, no significant differences were found concerning the gender make-up, average years of high school physics, or average BEMA pretest scores of the student population. Only minor variations in the ethnic make-up of the course were found. Variations in student majors were limited to spring/fall effects. Average overall CLASS pretest scores and average FMCE posttest scores were found to vary between particular pairs of semesters. Each pair-wise comparison will not be summarized here (since there are 21 pairs) however significant variations in student background factors for semesters of Phys2 are summarized in Appendix C, Table II. Variations in students’ backgrounds between specific implementations will be summarized as they become pertinent in upcoming discussions.

3. Can Variations in the Ethnicity of the Student Population enrolled in the Courses Account for the Observed Semester-to-Semester Variations in Students’ Perceptions of Tutorials?

The variations in the ethnic make-up of the introductory physics courses is small (varies by less than 10%). Small variations in the ethnic make-up of the Phys2 courses may be contributing to variations in students’ perceptions of Tutorials from semester-to-semester. To test this, a Kruskal-Wallis test was used to see if any particular ethnic group within a given course tended to report more favorably about the Tutorials across the ten survey questions asked. These results can be found in Table R. Since there are ten semesters being investigated, we reduce the p-value that will be considered significant by a factor of ten, such that $p \leq 0.005$ is considered statistically significant.

| Table R: Kruskal-Wallis Resultant P-values Comparing the Perceptions of Students from Different Ethnic Groups within a Given Semester. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | Phys1 | Phys2 |       |       |       |       |       |       |       |
|                 | N&D   | S&V   | F&W  | A&F  | F&H  | H    | L&M  | H&A  | R&M  | H&N  |
| HELPED TUTORIALS | 0.5   | 0.4   | 0.8  | 0.5  | 0.8  | 0.3  | 0.4  | 0.9  | 0.4  | 0.5  |
| ENJOY TUTORIALS  | 0.3   | 0.8   | 0.7  | 0.6  | 0.7  | 0.5  | 0.8  | 0.7  | 0.4  | 0.8  |
| FEEL TUT         | 0.6   | 0.8   | 0.4  | 0.7  | 0.5  | 0.4  | 0.2  | 0.3  | 0.5  | 0.9  |
| OFTEN ATTEND TUT | 0.8   | 0.04  | 0.8  | NA   | 0.4  | 0.2  | 0.1  | 0.6  | 0.2  | 0.09 |
| HELPED DISCUSS TUT | 0.5   | 0.2   | 0.3  | 0.1  | 0.7  | 0.9  | 0.7  | 0.9  | 0.5  | 0.7  |
| ENJOY DISCUSS TUT | 0.3   | 0.2   | 0.07 | 0.07 | 0.5  | 0.6  | 0.9  | 0.4  | 0.9  | 0.8  |
| HELPED TA        | 0.2   | 0.1   | 0.2  | 0.04 | 0.9  | 0.5  | 0.7  | 0.7  | 0.4  | 0.6  |
| HELPED LA        | 0.2   | 0.5   | 0.08 | 0.04 | 0.3  | 0.9  | 0.1  | 0.5  | 0.5  | 0.2  |
| HELPED EQUIP     | NA    | 0.3   | 0.7  | 0.5  | 0.8  | 0.8  | 0.5  | 0.9  | 0.1  | 0.6  |
| ENJOY EQUIP      | NA    | 0.01  | 0.6  | 0.07 | 0.9  | 0.8  | 0.8  | 0.3  | 0.5  | 0.9  |
There are no variations by ethnic group that are statistically significant at the $p \leq 0.005$ level. Overall, students’ perceptions of Tutorials within a given course do not seem to vary significantly by ethnic group. Based on these variations, it is unlikely that small changes in the ethnic make-up of the student population from semester-to-semester could account for the semester-to-semester variations that are observed in students’ perceptions of Tutorials.

4. Can Variations in the Declared Majors of Students’ from Spring to Fall Semesters Account for the Observed Semester-to-Semester Variations in Students’ Perceptions of Tutorials?

The variations in the distribution of majors enrolled in both Phys1 and Phys2 were found to vary from semester-to-semester particularly across spring and fall terms. Variations in the proportion of students with particular declared majors may be contributing to variations in students’ perceptions of Tutorials from semester-to-semester. The following question is addressed through this analysis: Within a given semester of physics, does a particular group of student major tend to respond more favorably on any Survey 1 questions about the Tutorials? Student majors were grouped into five categories: physics, engineering, other science, non-science, and undeclared. To test this, a Kruskal-Wallis test was used to see if any particular declared major group within a given course tended to report more favorably about the Tutorials across the ten survey questions asked. These results can be found in Table S. Since there are ten semesters being investigated, we reduce the p-value that will be considered significant by a factor of ten, such that $p \leq 0.005$ is considered statistically significant.

Table S: Kruskal-Wallis Resultant P-values Comparing the Perceptions of Students with Different Declared Majors within a Given Semester.

<table>
<thead>
<tr>
<th></th>
<th>Phys1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N&amp;D</td>
<td>S&amp;V</td>
<td>F&amp;W</td>
<td>A&amp;F</td>
<td>F&amp;H</td>
<td>H</td>
<td>L&amp;M</td>
<td>H&amp;A</td>
<td>R&amp;M</td>
<td>H&amp;N</td>
</tr>
<tr>
<td>HELPED TUTORIALS</td>
<td>0.02</td>
<td>0.1</td>
<td>0.8</td>
<td>0.3</td>
<td>0.3</td>
<td>0.8</td>
<td>0.4</td>
<td>$p &lt; 0.001$</td>
<td>0.6</td>
<td>0.05</td>
</tr>
<tr>
<td>ENJOY TUTORIALS</td>
<td>0.1</td>
<td>0.04</td>
<td>0.9</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.8</td>
<td>$p &lt; 0.001$</td>
<td>0.4</td>
<td>0.005</td>
</tr>
<tr>
<td>FEEL TUT</td>
<td>0.03</td>
<td>0.4</td>
<td>0.9</td>
<td>0.4</td>
<td>0.05</td>
<td>0.2</td>
<td>0.7</td>
<td>$p &lt; 0.001$</td>
<td>0.2</td>
<td>0.007</td>
</tr>
<tr>
<td>OFTEN ATTEND TUT</td>
<td>0.001</td>
<td>0.06</td>
<td>0.013</td>
<td>NA</td>
<td>0.5</td>
<td>$p &lt; 0.001$</td>
<td>0.1</td>
<td>0.2</td>
<td>0.03</td>
<td>0.3</td>
</tr>
<tr>
<td>HELPED DISCUSS TUT</td>
<td>0.2</td>
<td>0.01</td>
<td>0.7</td>
<td>0.03</td>
<td>0.04</td>
<td>0.7</td>
<td>0.1</td>
<td>0.01</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>ENJOY DISCUSS TUT</td>
<td>0.4</td>
<td>0.05</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.9</td>
<td>0.3</td>
<td>0.02</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td>HELPED TA</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
<td>0.6</td>
<td>0.2</td>
<td>0.6</td>
<td>0.7</td>
<td>0.65</td>
<td>0.5</td>
<td>0.09</td>
</tr>
<tr>
<td>HELPED LA</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.4</td>
<td>0.9</td>
<td>0.4</td>
<td>0.9</td>
<td>0.05</td>
</tr>
<tr>
<td>HELPED EQUIP</td>
<td>NA</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.03</td>
<td>0.2</td>
<td>0.1</td>
<td>0.09</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>ENJOY EQUIP</td>
<td>NA</td>
<td>0.9</td>
<td>0.9</td>
<td>0.5</td>
<td>0.3</td>
<td>0.07</td>
<td>0.8</td>
<td>0.02</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Across these ten semesters of Phys1 and Phys2, six semesters show no significant differences based on students’ particular declared major. Three semesters only show statistically significant differences on one question. We do not discuss the direction of these differences since these differences do not appear to be important across multiple questions and may be anomalous. Only one semester showed statistically significant differences on 3 questions out of the ten. On three questions that vary significantly by major in the H&A semester of Phys2, the physics majors respond the most favorably compared to other majors, followed by other science majors. Undeclared students, Engineering Majors, and Non-science majors tended to report the least favorably on these three questions. Perhaps this semester there are some meaningful differences by major, but they did not persist across the majority of the survey questions and we cannot estimate the degree of influence. Overall, students’ perceptions of Tutorials do not tend to vary strongly by major (statistically significant differences are rare within courses). It is unlikely that variations in students’ major can account for the observed semester-to-semester variations in students’ perceptions of Tutorials.

D. Examining the Influence of TAs

We assume that the population of TAs engaged in teaching this course does not vary significantly from semester-to-semester. We assume that the number of experienced (or returning) TAs remains fairly constant from semester-to-semester. This is justified since the policies for accepting graduate students into the physics department has not changed dramatically over the years and the policies for hiring graduate teaching assistants has also not changed over the years. This argument cannot however discard the possibility that fluctuations in the background experiences of the TAs assigned to teach the Tutorial course may influence students’ perceptions of the Tutorials.

At CU, the majority of TAs are first-year graduate students, many of whom have limited experience teaching beyond tutoring. Since most first year graduate students are awarded a full year-long teaching assistantship, TAs in the spring semester may have previous experience teaching at CU. An additional complexity is whether the TA taught with the Tutorials during the fall semester or with some other course. If the TA has prior experiences teaching with the Tutorials, then he or she may have developed ideas about how to teach with the Tutorials based both on their prior experiences in the classroom and through the training meetings with other secondary professors. TAs in the spring may tend to be more seasoned as compared to TAs in the fall, but the teaching norms that they have been exposed to may not always make them savvier at implementing the Tutorials curriculum. Variations in TAs and their background may contribute to semester-to-semester variations in students’ perceptions of Tutorials. Students’ perceptions of their TAs are explored in more detail in Survey 2.

E. Examining the Influence of the Lecture Professor on Students’ Survey Responses

One possible hypothesis might be that the semester-to-semester variations in students’ survey responses are due to differences in implementation decisions made by the primary instructor. In the case of Phys2 data available, there are two semesters of Tutorial use where the secondary professor is the same (Milo) and the primary professor varies (Linus and Ron). These two semesters of Phys2 are selected to investigate the potential influence of the primary professor on students’ perceptions of the Tutorials.

Looking at the L&M and the R&M semesters of Phys2, statistically significant differences are found on only one of the seven student background factors investigated—
students’ major. With the majority of student background factors being similar and the apparent similarities in students’ perceptions of Tutorial by major within a given course, it is unlikely that variations in students’ major would be responsible for variations in student perceptions of Tutorials. L&M semester is a Tutorial newcomer semester, while the R&M semester is not. Survey 1 consisted of thirteen questions. Three of these questions (Q51, Q52, and Q53) were not asked in either of these semesters. The remaining ten questions are examined in Table T where the Mann-Whitney U Test p-values are given as well as the semester which tended to yield more favorable student responses (for all questions with p<0.10). All statistically significant differences are shaded red.

Table T: Student Responses to Survey 1 Compared between two Phys2 semesters with Professor Milo as the Secondary Professor. (Statistical test used is Mann-Whitney U Test).

<table>
<thead>
<tr>
<th>Question</th>
<th>L&amp;M vs. R&amp;M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helped_Tutorials (Q8)</td>
<td>p=0.085 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Enjoy_Tutorials (Q41)</td>
<td>p=0.054 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Feel_Tutorials (Q48)</td>
<td>p=0.018 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Often_Attend_Tut (Q50)</td>
<td>p&lt;0.001 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Helped_Discuss_Peers (Q9)</td>
<td>p=0.417</td>
</tr>
<tr>
<td>Enjoy_Discuss_Peers (Q42)</td>
<td>p=0.005 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Helped_TA (Q23)</td>
<td>p&lt;0.001 (L&amp;M ↑)</td>
</tr>
<tr>
<td>Helped_LA (Q24)</td>
<td>p=0.048 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Helped_Equip (10)</td>
<td>p=0.048 (R&amp;M ↑)</td>
</tr>
<tr>
<td>Enjoy_Equip (Q43)</td>
<td>p=0.911</td>
</tr>
</tbody>
</table>

The Linus & Milo semester of Phys2 is found to be significantly different from the Ron & Milo semester of Phys2 on two out of ten questions. During the Ron & Milo semester, students reported attending the Tutorials more often (as compared to the L&M semester). During the Linus & Milo semester, students reported that their TAs were more helpful for their learning (as compared to the R&M semester). These variations may be due to differences in the lecture professor’s implementation decisions. Overall however, these two semesters do not vary dramatically in a consistent direction. Purely looking at these empirical results, this may suggest that the lecture professor does not influence students’ perceptions of Tutorials, or that these two lecture professors had similar approaches to Tutorial implementation.

F. Examining the Influence of the Back-up Professor across Three Phys2 Semesters

In this section, differences in the secondary instructor for the course are investigated. One possible hypothesis might be that the semester-to-semester variations in students’ survey responses are due to differences in implementation decisions made by the secondary instructor. In the case of Phys2 data available, there are three semesters of Tutorial use where the lecture professor is the same (Harold) and the secondary professor varies (Harold, Arthur, and Nemo). Although it is possible that Professor Harold changes across these three semesters, we have evidence from interviews that professors largely use previous course materials such as lecture notes, clicker questions, and homework problems from their prior implementations in designing their current courses and that there is likely to be significant overlap in their implementation.
decisions. These three semesters of Phys2 are selected to investigate the potential influence of the secondary professor on students’ perceptions of the Tutorials.

There are interesting differences in the prior experiences of these three instructors. Professor Arthur was the PER professor that first instigated the use of the Tutorials at CU and by the time he was involved in co-teaching with Professor Harold, he had already taught with the Tutorials in three prior semesters of Phys1 or Phys2. Professor Arthur had spent time at the University of Washington learning how the Tutorial designers implemented the Tutorials at their institution. Professor Harold and Nemo however do not conduct physics education research. Professor Harold was the secondary instructor once with another PER professor prior to teaching the Tutorials by himself. However, the first semester where Professor Harold was the primary instructor teaching with the Tutorials, he was left to also assume the responsibilities of the secondary professor due to a family emergency of his assigned secondary professor. When Professor Harold and Professor Nemo were co-teaching with the Tutorials, Professor Harold had been involved in teaching with the Tutorials a total of four different times and Professor Nemo had been involved in teaching with the Tutorials a total of twice (once as the primary instructor and once as the secondary professor).

First, similarities and variations in student background factors across these semesters are summarized. Only looking at the H, H&A, and H&N semesters of Phys2, we only see statistically significant differences on one of the seven student background factors investigated—students’ major. With the majority of student background factors being similar and the apparent similarities in students’ perceptions of Tutorial by major within a given course, it is unlikely that variations in students’ major would be responsible for the variations in student perceptions of Tutorials that are observed. The H semester is a Tutorial newcomer semester where the majority of students do not have prior experiences learning physics in a Tutorials environment. However, in both the H&A and the H&N semesters of Phys2 the majority of students experienced learning in a Tutorial environment in the previous Phys1 course.

Survey 1 consisted of thirteen questions. One question (Q53) was not asked in any of these three semesters. Two additional questions (Q51 & Q52) are not considered here since there were no significant differences found via Kruskal-Wallis test in the set of all Phys2 semesters and pair-wise comparisons were therefore unjustified. The remaining ten questions are examined in Table U where the Mann-Whitney U Test p-values are given as well as the semester which tended to yield more favorable student responses (for all questions with p<0.10). All statistically significant differences are shaded red.
Table U: Student Responses to Survey 1 Compared between Three Phys2 semesters with Professor Harold as the Lecture Professor.

<table>
<thead>
<tr>
<th>Semester Comparison</th>
<th>H vs. H&amp;A</th>
<th>H vs. H&amp;N</th>
<th>H&amp;A vs. H&amp;N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helped_Tutorials (Q8)</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
<td>p=0.10 (H&amp;N ↑)</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Enjoy_Tutorials (Q41)</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
<td>p=0.09 (H&amp;N ↑)</td>
<td>p=0.002 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Feel_Tutorials (Q48)</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
<td>p=0.08 (H&amp;N ↑)</td>
<td>p=0.001 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Often_Attend_Tut (Q50)</td>
<td>p=0.005 (H&amp;A ↑)</td>
<td>p=0.7</td>
<td>p=0.02 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Helped_Discuss_Peers (Q9)</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
<td>p=0.9</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Enjoy_Discuss_Peers (Q42)</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
<td>p&lt;0.2 (H&amp;A ↑)</td>
<td>p=0.009 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Helped_TA (Q23)</td>
<td>p=0.001 (H&amp;A ↑)</td>
<td>p=0.8</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Helped_LA (Q24)</td>
<td>p=0.6</td>
<td>p=0.04 (H&amp;N ↑)</td>
<td>p=0.1</td>
</tr>
<tr>
<td>Helped_Equip (10)</td>
<td>p=0.001 (H&amp;A ↑)</td>
<td>p=0.9</td>
<td>p&lt;0.001 (H&amp;A ↑)</td>
</tr>
<tr>
<td>Enjoy_Equip (Q43)</td>
<td>p=0.001 (H&amp;A ↑)</td>
<td>p=0.006 (H&amp;N ↑)</td>
<td>p=0.8</td>
</tr>
</tbody>
</table>

The Harold semester of Phys2 is found to be significantly different from the Harold & Arthur semester of Phys2 on eight of the remaining ten questions. The Harold & Nemo semester of Phys2 is also found to be significantly different from the Harold & Arthur semester of Phys2 on six of the remaining ten questions. However, the Harold semester of Phys2 and the Harold & Nemo semester of Phys2 are not found to vary significantly on any questions. There is significant evidence that the secondary professor can influence students’ perceptions of Tutorials. Significant differences were found between semesters in which Professor Arthur was the secondary instructor and semesters where either Professor Harold or Professor Nemo was the secondary instructor.

G. Summarizing the Influence of Instructors on Students’ Perceptions of Tutorials

These analyses have shown that students’ perceptions of Tutorials vary from semester-to-semester and that variations in student background factors are unlikely to account for variations in students’ perceptions of Tutorials on a semester-to-semester basis. After carefully examining particular semesters of Phys2 where one of the two co-teachers doesn’t change, it was discovered that the secondary professor can influence students’ perceptions of Tutorials. In comparing three semesters of Phys2 with the same lecture professor, we found that when Professor Arthur was in the back-up instructor role, students tended to report more favorably about the tutorials as compared to other semesters where Professor Arthur was not involved. One might wonder if semester-to-semester variations that are observed in Phys2 can be accounted for merely by Professor Arthur’s involvement—a PER professor. Although variations in students’ perceptions of Tutorials are common when comparing semesters where Professor Arthur is involved and semesters where Professor Arthur is not involved, there are many other variations across pairs of
Phys2 semesters where Arthur was not involved in either semester (see for example comparisons of F&HvsR&M or R&MvsH&N in Appendix B, Table IV). This suggests that instructors’ decisions in implementing Tutorials (in either the primary or secondary roles) may influence students’ perceptions of Tutorials.

V. Eliciting Students’ Finer-grained Perceptions of Tutorials with Survey 2

By analyzing Survey 2 responses, we hope to unpack and disaggregate aspects of the course which were commonly referenced in students’ comments. By examining themes of implementation that students noticed, we hope to identify aspects of course implementation that are worth further ethnographic investigations. Some aspects of course implementation may be contributing more to semester-to-semester variations that are observed in broad perceptions data from Survey 1 and these analyses of Survey 2 will help us to identify aspects of the course to investigate further. These analyses may help us to prioritize where to begin our future investigations into the efficacy and variations in Tutorials implementation.

Since initial analyses of Survey 1 suggested that aspects of Tutorial implementation may vary from semester-to-semester, Survey 2 was designed to see if particular aspects of Tutorial implementation were perceived differently by students. Using themes from students long answer responses on Survey 1, Survey 2 probes students’ perspectives on student-student interactions, student-TA interactions, and the coordination of the Tutorials with other course components. Survey 2 questions are provided in Appendix A, Table II.

Survey 2 was administered in two semesters (Fall 2008 & Spring 2009) of Phys1 and Phys2. As with Survey 1, comparisons are only made within each course across different semesters. As discussed in the Setting and Methods section, the S&V semester of Phys1 had a response rate of 77% (N=466) and the F&W semester of Phys1 had a response rate of 73% (N=430). When the students’ background factors were analyzed across these two semesters of Phys1, the student populations were only found to differ by the distribution of student majors. However, students’ perceptions of Tutorials were not found to vary according to students’ declared major. Similarly, the R&M semester of Phys2 had a response rate of 71% (N=322) and the H&N semester of Phys2 had a response rate of 65% (N=280). When six student background factors were analyzed across these two semesters of Phys2, the student populations were found to differ by the distribution of student majors and students’ incoming FMCE posttest score. However, students’ perceptions of Tutorials were not found to vary according to students’ declared major.

A. Student-Student Interactions

There are eight survey questions which target students’ reflections on peer interactions during the Tutorials. Two of these questions were taken from the original Survey 1 questions and an additional six questions were added based on students’ long answer responses. Most of the newly added survey 2 items include statements that students are asked to agree or disagree with. Examples of these statements include: “During physics discussions with my peers in the Tutorials, I usually share my ideas,” and “In the Tutorials, I feel like my peers value my ideas.” Two semesters of Phys1 and two semesters of Phys2 are compared in Table V. A Mann-Whitney test is used to determine if one semester tends to report more favorably than another. For all p<0.1, the semester that tended to yield more favorable responses is indicated with an arrow. Statistically significant differences are highlighted in red for p<0.05. Although these answer options are technically categorical, to give a quick sense of the magnitude and direction
of the difference, mean values are reported for each semester by applying a continuous numerical score to these categories such that a higher number denotes a more favorable response. For each survey question, p-values and mean responses for each course can be found in Table O.

Table V: Comparison of Students’ Survey 2 Responses about Peer Interactions in Tutorials: Mean Values given for each semester and associated p-value based on comparison by Mann-Whitney Test.

<table>
<thead>
<tr>
<th>Specific Question</th>
<th>S&amp;V Mean</th>
<th>F&amp;W Mean</th>
<th>MW p-values</th>
<th>R&amp;M Mean</th>
<th>H&amp;N Mean</th>
<th>MW p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELPED_DISCUSS_TUT†</td>
<td>3.18 ±0.06</td>
<td>2.99 ±0.06</td>
<td>p=0.011</td>
<td>3.48 ±0.07</td>
<td>3.11 ±0.07</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>ENJOY_DISCUSS_TUT†</td>
<td>3.35 ±0.05</td>
<td>3.13 ±0.05</td>
<td>p=0.001</td>
<td>3.57 ±0.06</td>
<td>3.17 ±0.07</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>COMFORT_Peers_TUT</td>
<td>4.41 ±0.04</td>
<td>4.39 ±0.04</td>
<td>p=0.3</td>
<td>4.47 ±0.05</td>
<td>4.40 ±0.06</td>
<td>p=0.3</td>
</tr>
<tr>
<td>SHARE_IDEAS_TUT</td>
<td>4.14 ±0.05</td>
<td>4.20 ±0.05</td>
<td>p=0.6</td>
<td>4.19 ±0.06</td>
<td>4.13 ±0.06</td>
<td>p=0.2</td>
</tr>
<tr>
<td>PEERS_VALUE_IDEAS_TUT</td>
<td>3.93 ±0.05</td>
<td>4.03 ±0.04</td>
<td>p=0.167</td>
<td>4.02 ±0.05</td>
<td>3.97 ±0.06</td>
<td>p=0.6</td>
</tr>
<tr>
<td>PEERS_LISTEN_CAREFULLY</td>
<td>3.83 ±0.05</td>
<td>3.91 ±0.04</td>
<td>p=0.2</td>
<td>3.98 ±0.05</td>
<td>3.90 ±0.06</td>
<td>p=0.2</td>
</tr>
<tr>
<td>STUDENTS_DONT_LISTEN</td>
<td>3.91 ±0.05</td>
<td>4.07 ±0.05</td>
<td>p=0.039</td>
<td>4.08 ±0.06</td>
<td>3.99 ±0.06</td>
<td>p=0.2</td>
</tr>
<tr>
<td>STUDENTS_DONT_TAKE_SERIOUS</td>
<td>4.03 ±0.05</td>
<td>4.11 ±0.05</td>
<td>p=0.10</td>
<td>4.21 ±0.06</td>
<td>4.03 ±0.06</td>
<td>p=0.030</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>3.85 ±0.05</td>
<td>3.85 ±0.05</td>
<td>(F&amp;W↑)</td>
<td>4.00 ±0.06</td>
<td>3.84 ±0.06</td>
<td>(R&amp;M↑)</td>
</tr>
</tbody>
</table>

†Question taken from survey 1 items significance taken to be p≤0.002 for Phys2 semester and p≤0.017 for Phys1 semesters based on the different number of semester being compared.

Students tend to report favorably about their interactions with their peers in Tutorials (each semester the student-student interaction cluster averages from between 3.84 to 4.00). The two semesters of Phys1 vary on three of the eight questions and the semesters of Phys2 vary on three of the eight questions. Students perceptions of student-student interactions do not seem to vary strongly from semester to semester (cluster averages are indistinguishable between Phys1 semesters and vary by 0.16 across semesters of Phys2).

B. Student-TA Interactions

Students’ perceptions of their TAs were expected to vary since aspects of the TA/LA training meetings across these semesters tended to vary. For these four semesters, the Lead Graduate teacher was asked to write one paragraph summaries at the end of the term describing the typical practices in the TA/LA training meetings emphasizing the role that the secondary professor played in these meetings. The summaries written by the Lead TA are provided in Table W.
| Physl Fall 2008—S&V semester: Professor Violet, Secondary Professor | “This group was particularly large due to an influx of ‘LAs’ from the Teaching & Learning Physics course, which often hindered large group discussion. The facilitator also provided large amounts of snacks, and circulation of the snacks occasionally distracted from the task at hand. About 10 minutes were spent working the pretests, 5 minutes looking at sample responses, and another less than 5 discussing observations in large group. Both the facilitator and the lecturer (Professor Stuart) circulated and asked questions pertaining to the tutorial. Meetings were generally 75-90 minutes in length.” |
| Physl Spring 2009—F&W semester: Professor Wilbur, Secondary Professor | “Meetings began with TAs/LAs completing pretests in an informal manner, often discussing pretest questions as they were being completed. Sample responses were handed out and discussed in small groups, but some tables would spend more time discussing than others. Groups would often finish at different times since some groups started the tutorial about 10 minutes earlier, leading to some rushed discussion of the last pages. Facilitator circulated and modeled TA behavior by asking about particular tutorial questions. A faculty TA (the department chair) participated in the small groups. Meetings were generally 60-75 minutes in length.” |
| Phys2 Fall 2008—R&M semester: Professor Milo, Secondary Professor | “About 10 minutes were spent completing the pretests, and 5 minutes discussing the pretest in small groups. Sample responses were not given. Facilitator did not circulate continually and occasionally sat off to the side while the groups worked. Facilitator interacted with the groups primarily to check progress and accuracy on the tutorials. Meetings were generally 45-60 minutes in length.” |
| Phys2 Spring 2009—H&N semester: Professor Nemo, Secondary Professor | “Meetings began with TAs/LAs completing pretests in a formal manner, individually. Facilitator then instructed TAs/LAs to each predict and write down two common student difficulties, handed out sample responses, and asked TAs/LAs to compare the observed difficulties with their prediction. This process often took at least 30 minutes. Facilitator occasionally identified what he considered “trouble spots” and discussed how to handle these in large group before starting the tutorial. Facilitator commonly posed “challenge” questions to probe TA/LA conceptual understanding, and discussed these questions at length with each table. Facilitator provided a few small bags of candy at each meeting. Meetings were generally at least 90 minutes in length.” |

Across the training meeting of the two semesters of Physl, Professor Violet led a whole group discussion about common student ideas and associated difficulties on the pretest while Professor Wilbur did not. Both professors handed out student responses to the pretest and expected TAs and LAs to reflect on students’ comments—although Professor Wilbur did not expect TAs and LAs to describe and discuss these with the entire group. During the S&V semester of Physl, the lecture professor (Professor Stuart) was consistently present at the TA/LA meeting also modeling TA behavior. Although the lecture professor was not present at the training meetings for the F&W semester of Physl, another professor (the department chair) was present these meetings sitting with TAs and completing the Tutorial activities with them. Based on my own observations in these environments, Professor Violet provided significant scaffolding for TAs about exactly what TAs might ask their students at particular points in the tutorial—
expecting TAs and LAs to reflect on what they will do to support students in completing and learning from this tutorial.

Across the training meeting of the two semesters of Phys2, Professor Milo didn’t hand out examples of students’ responses to the Tutorial pretest questions while Professor Nemo did. In Professor Milo’s training meeting, there was no discussion of the ideas that students might bring to the class. In contrast, Professor Nemo asked TAs and LAs to predict what ideas students might bring to class and to then compare their predictions with actual student responses to the pretest. Professor Milo rarely interacted with the TA and LA groups as they completed the Tutorials. Professor Nemo discussed what he perceive as ‘trouble spots’ and how to address these spots with students. Professor Nemo seemed to provide more scaffolding to TAs and LAs about their role in the Tutorial environments.

It is not well understood how differences in the TA/LA training meetings might affect students in the Tutorial recitations. The differences in TA/LA training meeting practices do not directly affect students, rather they potentially influence TAs and LAs and the choices that TAs and LAs make about how to engage in the Tutorial recitations. Assuming that the TA/LA training meetings impact the instructional decisions of TAs, one might expect to see differences in how students perceive their TAs since different instructional practices were encouraged in the various TA/LA training meetings.

There are seven survey questions which target students’ reflections on interactions with their TA during the Tutorials. One of these questions was taken from the original Survey 1 questions and an additional six questions were added based on students’ long answer responses. Most of the newly added survey 2 items include statements that students are asked to agree or disagree with. Examples of these statements include: “My TA encouraged me to further investigate physics during Tutorials,” and “My TA listened carefully to my ideas in the Tutorials.” As with the previous cluster of survey 2 questions, two semesters of Phys1 and two semesters of Phys2 are compared in Table X. Statistics were calculated in a similar fashion to the previous cluster. For each survey question about student-TA interactions, p-values and mean responses for each course can be found in Table X.
Students are somewhat favorable about their interaction with their TAs in Tutorials (although less favorable than working with their peers). Students’ average responses across this cluster range from 3.52 to 3.71. In all semesters, the Student-TA cluster average is lower than the student-student cluster average. The two semester of Phys 1 vary on six of the seven questions demonstrating that students’ perceptions of their TAs can vary significantly from semester-to-semester. Students’ perceptions of TAs do not always vary however since the two semesters of Phys2 only vary on one out of the seven questions.

Based on the significant differences observed in the TA/LA training meeting across the two Phys2 semesters, it is surprising to see such limited variations in students’ perceptions of their TA. Equally puzzling is the significant differences in students’ perceptions of their TAs across the two semester of Phys1. In the context of these Tutorials survey questions, it becomes less clear if students disaggregate course components. Understanding students’ perspectives on their interactions with TAs may be particularly complex due to the unfamiliar structure of the Tutorial learning environment. Students may not evaluate ‘good teaching’ or ‘helpful questions’ in ways which align with the Tutorial designers’ intentions.

C. Coordination of Course Components

There are five survey questions which target students’ reflections on the coordination of the Tutorials with other course components. None of these questions were taken from the original Survey 1 questions. However, many of students’ long answer responses emphasized the lack of coordination between the Tutorials and other aspects of the course. Two questions investigated the coordination of the Tutorials with exams and three questions investigated the

---

Table X: Comparison of Students’ Survey 2 Responses about Student-TA Interactions in Tutorials: Mean Values given for each semester and associated p-value based on comparison by Mann-Whitney Test.

<table>
<thead>
<tr>
<th>Specific Question</th>
<th>Phys1 S&amp;V Mean</th>
<th>Phys1 F&amp;W Mean</th>
<th>Phys1 MW p-values</th>
<th>Phys2 R&amp;M Mean</th>
<th>Phys2 H&amp;N Mean</th>
<th>Phys2 MW p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELPED_TA†</td>
<td>3.02 ±0.05</td>
<td>3.23 ±0.05</td>
<td>p=0.006 (F&amp;W↑)</td>
<td>3.08 ±0.07</td>
<td>3.19 ±0.07</td>
<td>p=0.3</td>
</tr>
<tr>
<td>COMFORT_TA_TUT</td>
<td>4.09 ±0.05</td>
<td>4.10 ±0.05</td>
<td>p=0.9</td>
<td>4.12 ±0.06</td>
<td>4.18 ±0.06</td>
<td>p=0.3</td>
</tr>
<tr>
<td>TA_RIGHT_ANSWER</td>
<td>2.92 ±0.06</td>
<td>3.09 ±0.06</td>
<td>p=0.04 (F&amp;W↑)</td>
<td>3.09 ±0.07</td>
<td>2.96 ±0.07</td>
<td>p=0.2</td>
</tr>
<tr>
<td>TA ENCOURAGE</td>
<td>3.47 ±0.05</td>
<td>3.66 ±0.06</td>
<td>p=0.01 (F&amp;W↑)</td>
<td>3.40 ±0.07</td>
<td>3.53 ±0.07</td>
<td>p=0.2</td>
</tr>
<tr>
<td>QUESTIONS_TA_HELPED</td>
<td>3.55 ±0.05</td>
<td>3.88 ±0.05</td>
<td>p&lt;0.001 (F&amp;W↑)</td>
<td>3.61 ±0.07</td>
<td>3.82 ±0.06</td>
<td>p=0.05 (H&amp;N↑)</td>
</tr>
<tr>
<td>QUESTIONS_FACILITATE_DISCUSS</td>
<td>3.66 ±0.05</td>
<td>3.82 ±0.05</td>
<td>p=0.01 (F&amp;W↑)</td>
<td>3.50 ±0.06</td>
<td>3.72 ±0.06</td>
<td>p=0.02 (H&amp;N↑)</td>
</tr>
<tr>
<td>TA_LISTEN_CAREFULLY</td>
<td>3.94 ±0.05</td>
<td>4.18 ±0.04</td>
<td>p&lt;0.001 (F&amp;W↑)</td>
<td>3.98 ±0.06</td>
<td>4.12 ±0.06</td>
<td>p=0.3</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>3.52</td>
<td>3.71</td>
<td></td>
<td>3.54</td>
<td>3.65</td>
<td></td>
</tr>
</tbody>
</table>

†Question taken from survey 1 items significance taken to be p≤0.002 for Phys2 semester and p≤0.017 for Phys1 semesters based on the different number of semester being compared.
connections made between the Tutorials and the lecture portion of the course. The newly added survey 2 items include statements that students are asked to agree or disagree with. Examples of these statements include: “The material tested on the exams was related to the Tutorial activities,” and “The Tutorial activities helped me make sense of the lectures.” As with the previous cluster of survey 2 questions, two semesters of Phys1 and two semesters of Phys2 are compared in Table Y. Statistics were calculated in a similar fashion to the previous cluster. For each survey question about the coordination of course components, p-values and mean responses for each course can be found in Table Y.

Table Y: Comparison of Students’ Survey 2 Responses about the Coordination of the Tutorials with Other Course Components: Mean Values given for each semester and associated p-value based on comparison by Mann-Whitney Test.

<table>
<thead>
<tr>
<th>Specific Question</th>
<th>Phys1</th>
<th>Phys2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S&amp;V Mean</td>
<td>F&amp;W Mean</td>
</tr>
<tr>
<td>TUT_SUCCEED</td>
<td>3.20 ± 0.06</td>
<td>2.73 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>(S&amp;V↑)</td>
<td>(S&amp;V↑)</td>
</tr>
<tr>
<td>EXAMS_RELATED_TUT</td>
<td>3.80 ± 0.05</td>
<td>3.07 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>(S&amp;V↑)</td>
<td>(S&amp;V↑)</td>
</tr>
<tr>
<td>SENSE_LEC</td>
<td>3.06 ± 0.05</td>
<td>3.04 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>(R&amp;M↑)</td>
<td>(R&amp;M↑)</td>
</tr>
<tr>
<td>LECTURE_MAKE_SENSE</td>
<td>2.95 ± 0.05</td>
<td>3.60 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>(F&amp;W↑)</td>
<td>(F&amp;W↑)</td>
</tr>
<tr>
<td>SYNCHRONIZED</td>
<td>3.08 ± 0.05</td>
<td>2.87 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>(S&amp;V↑)</td>
<td>(S&amp;V↑)</td>
</tr>
<tr>
<td>Average</td>
<td>3.22</td>
<td>3.06</td>
</tr>
</tbody>
</table>

Students are by far the most negative about the connections between the Tutorials and other course components (cluster averages range from 3.06 to 3.36). Significant variations across semesters are found in the coordination cluster for both Phys1 and Phys2. Interestingly, the variations across questions however do not always occur in the same direction. For example, in the R&M semester of Phys2, students are more favorable about Tutorials helping them make sense of the lectures than the H&N semester of Phys2. While in the H&N semester of Phys2, students are more favorable about lecturing helping them make sense of the Tutorials than students in the R&M semester. There were not dramatically different approaches across these two semester about when topics where first introduced—either in lecture or in Tutorial recitation. This may be some indication that students may be evaluating course component relative to each other. For example, if the lecture is perceived as less clear or difficult to understand, students may value the Tutorial experience more. Or the converse may be true that if students feel that the lecture is clear and understandable, student may value the Tutorials less.

VI. Study Limitations and Directions for Future Study

The first limitation of these studies is that the survey instruments utilized have not been rigorously validated. To better understand these survey responses, students should be interviewed to establish that there is a consistent interpretation of the survey statements. However, Survey 1 questions were written in a style consistent with many other instruments
including familiar end-of-course questionnaires. Survey 2 items were written based on language from student long answer responses to Survey 1. Additionally, student long answer responses coupled to a subset of Survey 2 survey items are consistent with researchers’ interpretations of the survey items.

The second limitation to these survey studies is concerns that students’ perceptions of Tutorials may be tightly coupled to students’ overall perceptions of the course more broadly. Students’ perceptions of Tutorials may be coupled to assessment in the course, practices in the lecture, etc. Students may not clearly disaggregate course components. Further investigations into what students perceive as important in Tutorials may benefit from use of Q-methodology where students would rank order statements relative to each other rather than just agreeing or disagreeing with statements.

A third limitation of these survey studies is that students may be evaluating course components relative to each other. For example, if the lecturer is perceived as unclear or difficult to understand, students may value the Tutorial experience more. Similarly, if the lecturer is perceived as clear and easily understandable, students may not value struggling to make sense of the material themselves in Tutorials. This suggests that Tutorials must be understood relative to other aspects of the course which also suggests the potential utility of using an instrument based on Q-methodology.

VII. Conclusions

Through analysis of Survey 1, significant semester-to-semester variations in students’ survey responses about the Tutorials were found. The absolute magnitude of these variations was found at times to be as much as 20%. For example, some semesters (such as the F&W semester of Phys1) we see only 25% of students reporting that Tutorials are useful for their learning while other semesters (such as the H&A semester of Phys2) as many as 50% of students are reporting that Tutorials are useful for their learning. Students were sometimes found to be quite favorable about the utility of the Tutorials, although students were often quite negative about the enjoyment associated with the Tutorials. Students consistently responded more favorably about working with their peers in Tutorials than they did about the Tutorials more broadly. It was shown that variations in student background factors are unlikely to account for variations in students’ perceptions of Tutorials on a semester-to-semester basis. The potential impact of variations in the population of teaching assistants should be investigated further. After carefully examining particular semesters of Phys2 where one of the two co-teachers didn’t change, it was discovered that the secondary professor can influence students’ perceptions of Tutorials. In comparing three semesters of Phys2 with the same lecture professor, we found that when Professor Arthur was in the back-up instructor role, students tended to report more favorably about the tutorials as compared to other semesters where Professor Arthur was not involved. Although variations in students’ perceptions of Tutorials are common when comparing semesters where Professor Arthur is involved and semesters where Professor Arthur is not involved, there are many other variations across pairs of Phys2 semesters where Arthur was not involved in either semester (see for example comparisons of F&H vs R&M or R&M vs H&N in Appendix B, Table IV). This suggests that instructors’ decisions in implementing Tutorials (in either the primary or secondary roles) may influence students’ perceptions of Tutorials.

By analyzing students’ long answer responses on Survey 1, three themes emerged that seemed particularly salient to students concerning their experiences in Tutorials: student-student
interactions, student-TA interactions, and the coordination of course components. Survey 2 included items added to particularly probe these dimensions. It was found that students are overall quite positive about working with their peers and that these perspectives do not tend to vary significantly from semester-to-semester (cluster averages range from 3.84 to 4.00). Students were found to be somewhat favorable about their interactions with their TAs in Tutorials although less so than about interactions with their peers (cluster averages range from 3.52 to 3.71). In all semesters, the Student-TA cluster average is lower than the student-student cluster average. Significant variations on the Student-TA clusters were found between the two semesters of Phys1, but not between the two semesters of Phys2. However, the differences perceived by students do not clearly associate with differences instructional practices as was found in the case of Peer Instruction. Students were found to be the most negative about connections between the Tutorials and other course components (cluster averages range from 3.06 to 3.36). Significant variations across semesters are found in the coordination cluster for both Phys1 and Phys2. Interestingly, the variations across questions do not always occur in the same direction.

It appears that by the end of the term, students’ perspectives on Tutorials are favorable overall unlike the initial fears of physics faculty members. On average, students tend to be neutral or positive about many aspects of Tutorials including working with their peers and working with educators in the Tutorial environments. However, students tend to not enjoy the Tutorials. Significant semester-to-semester variations in students’ perceptions of Tutorials were identified, suggesting that semester-to-semester variations in implementation may influence students’ experiences in Tutorials. It may still be worth further investigation to see if and how students’ perspectives on the Tutorials may change over time throughout the course. The concerns of faculty about students’ dislike of the Tutorial learning environment may accurately represent students’ experiences during earlier stages of the term.

Understanding student-TA interactions from the students’ perspective and associating these views with observed practices in the TA/LA Tutorial Training meeting was found to be particularly complex. It may be more fruitful to examine the practices of TAs (across multiple TAs within the same course) and attempt to associate these practices with students’ perspectives on their TA. It is hypothesized that the TA/LA training meetings may influence the decisions that TAs make in the classroom, but this assumption is worth further examination. The nature and extent of the influence of the TA/LA training meetings on TAs’ instructional practices is unclear. Using students’ survey responses to understand their thought about their TAs were particularly problematic since it wasn’t clear if students’ ideas about ‘good practice’ aligned with the intentions of the Tutorial designers.

Further investigations into understanding the use of Tutorials should investigate the ongoing practical routines of students and their TAs in Tutorials recitations as well as how connections are made across course components in the routines of the course. Students and TAs are simultaneously working to engage in a learning environment that is structured in unfamiliar ways. Interviews with students and TAs may be the best way to proceed in making sense of these participants’ experiences.
Peer-reviewed Publications based on this work:

References (Chapter 5)


[26] University of Colorado Course catalog; Available at [http://www.colorado.edu/catalog/catalog08-09/search.pl?abbr=PHYS&num=](http://www.colorado.edu/catalog/catalog08-09/search.pl?abbr=PHYS&num=)


[28] Chi-squared Test for Independence determines if we are justified in rejecting the null hypothesis which states that the row variables (semester) and the column variables (possible answer options) are unrelated (that is, only randomly related).

[29] Kruskal-Wallis Test for k-independent samples determines if we are justified in rejecting the null hypothesis which states all of the populations are identical versus the alternative that some of the populations tend to furnish greater observed values than other populations.


Chapter 6: Conclusions

This thesis demonstrates that educational change does not occur by simply replicating effective curricular materials or by arguing for fixed instructional strategies to be used by individual instructors. For example, the complexity of adopting and implementing PI is not captured by a simple seven step process [1,2]. Through the use of a socio-cultural theoretical perspective (i.e. cultural historical activity theory [3,4] and a frames-of-context model [5,6]) in this thesis, we have shown that simple linear models of educational change are insufficient. Enacting educational change is not simply making available instructional tools (clickers), nor simply implementing prescribed instructional practices (the seven steps of Peer Instruction). Accurately representing educational change requires a systems view that captures the nested nature of educational systems and the complex feedback that occurs across levels of the educational system. This thesis demonstrates that wide-scale change is possible, but documenting what exactly changes requires a broader unit of analysis than solely focusing on individual instructors or individual students.

In the three detailed case studies of PI implementation in this thesis, varying instructional decisions (implicit or explicit) are shown to create different classroom cultures (reflected in different classroom norms and practices). From a Vygotskian perspective [7], learning is defined as the internalization of social norms and practices, and hence classroom cultures are critical to understand. Students learn different things in classes with different social norms and practices. According to Vygotsky’s general genetic law of cultural development [7], all higher psychological functions appear twice, first on the social level (between people) and later on an individual level. Therefore, the social forms of activity that occur in a classroom are precursors to what eventually may become internalized by individual students. Focusing on studying the interpersonal or social forms of higher psychological processes provides a valuable research lens since these functions are still distributed in an observable arena. A Vygotskian focus on the social norms and practices of the classroom may provide new opportunities for understanding what has previously been called the “hidden curriculum” [8]. For example, if a student agrees that “A significant problem in learning physics is being able to memorize all of the information I need to know” (A CLASS Statement) [9], this response may be a result of the student’s prior participation in the activities of a classroom where memorization was a perceived norm, rather than a deficiency of the student. Focusing on the cultural practices within ongoing classroom activities also provides the opportunity to investigate how previous theoretical constructs such as student content knowledge, student affect, student attitudes and beliefs, student engagement, and student identity may be interrelated.

A detailed descriptive analysis characterizing the social norms and practices of practical classroom activity shines new light on what exactly might be learned through students’ engagement in these activities. For example, our analyses of PI classrooms suggest additional student learning outcomes that have yet to be examined, such as students’ facility with formulating questions and verbally communicating their physical reasoning (which are valued disciplinary activities). With these possible student outcomes in mind, new avenues for formative assessment and instructional intervention become apparent. If an instructor does not provide opportunities for students to engage in scientific practices during class, then there are few opportunities for the instructor to give feedback to students on how they are engaging in scientific practices. Students engaging in scientific practices must happen in class if the
The research studies in this thesis about Tutorial adoption and implementation demonstrate that Tutorial use at CU did not occur by individual instructors simply deciding to adopt the Tutorials activity book. The use of the Tutorials at CU resulted in changes across multiple scales of the educational system (the physics department, partnerships between CU programs and departments, the university administration, and partnerships between CU programs and external organizations). These investigations suggest that different levels of the educational system influence each other through dynamic feedback loops which would call into question typical change models, such as RDDE discussed in Chapter 1 [10,11]. This case study in successful institutional change suggests that change efforts might be more successful if they focused on 1) addressing and coordinating multiple levels of the educational system, 2) developing local structures that support reflective practice (i.e. the sharing of course materials and assessment results), and 3) dedicating infrastructure and resources to the change process. Additional case studies of change efforts at additional institutions would help in further identifying critical aspects of these changes processes.

In this final discussion, we take the opportunity to look across the two case studies of Peer Instruction and Tutorials. We describe how the integration of these instructional strategies at CU has led to fundamental changes in how students engage in learning physics as compared to traditional lecture courses. The complexities of change processes are summarized drawing from concrete examples of classroom routines. The discussion of implications is organized around the over-arching research questions presented in the introduction.

Research Question #1: Do the daily classroom routines of a traditional lecture vary significantly from the classroom routines surrounding the use of Peer Instruction and the Tutorials by professors at CU? Are the students engaging in doing physics in new ways?

In Chapter 2, we saw that the use of PI by physics faculty at CU has led to some consistent transformations in the lecture portion of the course. In all of the introductory physics courses studied where PI was run, students were found to be trying out and applying physical concepts, discussing physics content with their peers, justifying their reasoning to their peers, and debating physical reasoning with their peers. These modifications to the daily routines of the physics classroom are significant since students now have opportunities to actively work out physics problems in class making their thinking about physics visible to peers and the professor for immediate feedback. Classes provided varied opportunities to students in other ways discussed under the subsequent research question.

Similarly, in Chapter 4, we saw that the changes associated with the adoption of the Tutorials span multiple scales of the educational system. Students involved in working on the Tutorials in recitations are now found to be engaged in conducting “thought-experiments,” experimenting with physical equipment, hypothesizing the consequences of physical phenomena, and actively solving problems in contrast to the traditional recitation where students were commonly found to passively watch graduate teaching assistants solve problems at the chalkboard. The impacts of the use of Tutorials however are not limited to undergraduate students enrolled in the course. Undergraduate STEM majors can now apply for and engage in early teaching experiences as learning assistants in undergraduate physics courses (which exposes them to a potential career in teaching as well as improves their mastery of the physics content [12]). Graduate student teaching assistants are now spending more time preparing for
teaching with the guidance of the lead instructors as well as being exposed to more student-
centered instructional strategies.

These changes in the daily routines of instructors at CU are not trivial. In fact prior
research directed at understanding the beliefs of faculty would have posited that change would
require extensive interventions or professional development with faculty, if not claim that change
would not be possible [13]. We have shown, however, that many physics faculty (even beyond
those in the PER community) have dramatically shifted many of their instructional practices in
the introductory physics classroom. At CU, change occurred through largely ecological means
and informal community consensus building which were met by top-down support; these change
processes are particularly illustrated in the Tutorial studies discussed in more detail later.

Research Question #2: Are there significant differences amongst CU physics professors’
implementations of Peer Instruction? What are the impacts of these differences on how students
are observed to be engaged in doing physics?

Although some consistency across PI implementations at CU was found, significant
variations were also identified. The degree to which instructional environments supported
students’ expression of their physical reasoning in the public discussion of the CQ solution
varied from less than 10% of the time to 100% of the time. The degree of interaction that
occurred between students and professor during PI was also found to vary. Across the PI
classrooms observed, there were large differences in students’ opportunities to engage in
formulating and asking questions, evaluating the correctness and completeness of problem
solutions, interacting with physicists, identifying themselves as sources of solutions,
explanations, or answers, and communicating scientific ideas in a public arena.

We demonstrate that concrete actions in PI implementation lead to differences in
classroom norms as perceived by both students and the observing researcher. Students are
shown to perceive PI classrooms differently (reflected in student survey responses) in ways that
were associated with PI implementation. We specified collections of classroom practices that
appeared to support students’ sense that PI was about faculty-student collaboration, student-
student collaboration, and sense-making. These claims are supported through successive
examinations of case studies of Professor Red and Green present in Chapters 2 and 3. We briefly
describe these case studies here.

In Chapter 2, case studies of Professor Red and Professor Green demonstrated how PI
implementation differences may create different classroom norms. The most significant
differences between Green and Red become apparent during the CQ solution discussion stage.
Although both professors asked conceptual questions and elicited verbal student explanations,
the professors spent different amounts of time discussing the solution, and different types of
interactions occurred between the professor and students during the solution discussion. In
Green’s case, only a single student explanation was elicited, and this student’s explanation was
typically clear and correct. Following this correct student explanation, the professor
communicated the correctness of this explanation and did not elicit additional student comments,
although more than 25% of the students had answered the question incorrectly. In Red’s case,
we saw that multiple students contributed explanations, and some correct and some incorrect
ideas were presented publicly. In this example, the student explanations built on fellow students’
answers. In Red’s class, students were responsible for evaluating the correctness and
completeness of the problem solution proposed by their peers. Students in these classrooms
were given different opportunities to practice identifying themselves as sources and evaluators of solutions, explanations, or answers.

These differences resulted in different kinds of faculty-student collaboration and differences in the use of student prior knowledge. Additionally, these implementation differences contributed to varying degrees of emphasis on reasoning and sense making (one particular norm more fully discussed in Chapter 3). It appears that although students did have a significant amount of voice in Green’s class, flawed student reasoning was not voiced equally in this class even on questions where there was a significant fraction of students incorrectly answering the CQ. Since incorrect ideas were not as likely to be shared, the importance of reasoning and sense making in this class was reduced. It was the answer that was predominantly valued. Red’s course, on the other hand, further emphasized the importance of reasoning through the professor’s management of disagreement among his students. From this case study, we saw how Red encouraged student-to-faculty dialogue by asking clarifying questions and structured student-to-student dialogue by positioning students to respond to or comment on other students’ contributions. In this way, the professor structured the students’ interactions with other students such that they were debating, providing alternative explanations, arguing, defending, challenging, or clarifying each others ideas.

Research Question #3: How do students perceive courses using Peer Instruction and the Tutorials? Do students perceive the use of the educational tool and the classroom in different ways depending on the tool’s use in classroom activities?

As discussed in Chapter 3, students perceive PI differently based on specific PI implementation practices. In the context of PI, detailed observations of PI practices allowed for the construction of definitions of classroom norms that were closely tied with concrete classroom practices. Survey questions targeting these norms revealed that students were more likely to report that they were comfortable discussing with and asking questions of their professor in courses where there were many types of faculty-student interaction that occur often during class and in which students and faculty took on and moved amongst a diversity of roles in the class.

In Chapter 5, we also saw that students’ perceptions of Tutorials significantly varied from semester to semester. For example, some semesters (such as the F&W semester of Phys1) we saw that only 25% of students reported that Tutorials were useful for their learning while other semesters (such as the H&A semester of Phys2) as many as 50% of students reported that Tutorials were useful for their learning. Students were sometimes found to be quite favorable about the utility of the Tutorials, although students were often quite negative about the enjoyment associated with the Tutorials. Students consistently responded more favorably about working with their peers in Tutorials than they did about the Tutorials more broadly. It was shown that variations in student background factors are unlikely to account for variations in students’ perceptions of Tutorials on a semester-to-semester basis. Results of these investigations suggest that instructors’ implementation decisions (in either the primary or secondary roles) may influence students’ perceptions of (and likely engagement with) Tutorials. This finding would parallel finding from our related PI studies.

The coupled PI studies illustrated the utility of detailed examinations of classroom practices to uncover largely implicit cultural norms. That is, while many professors and students at CU know that PI helps to make students more engaged and active, they may be less able to articulate the role that peer discussion played in the course and how concrete classroom practices may encourage or discourage sense-making. Uncovering variations in classroom norms in the
case of the Tutorials is more complicated. In our investigations, this is largely due to the differences in limited data sources that were used to generate descriptions of norms in the case of the Tutorials. In attempting to draw from students’ self-reported descriptions of their Tutorial experiences to develop norms, we found that students were largely unable to articulate implicit cultural practices in short answer survey responses.

Research Question #4: How has the CU-Boulder physics department created a sustained use of the Tutorials? What changes occurred throughout the educational system that allowed for the sustained use of the Tutorials?

The research studies and associated analysis described in Chapter 4 draws attention to the fact that purely describing the changes that individual faculty members’ experience would not fully capture the complexity of the change processes associated with Tutorials adoption. Individual faculty members—the initial PER adopter, additional non-PER faculty members participating in implementing the Tutorials, and faculty advocates—played important roles in the adoption and institutionalization process, but other simultaneous changes throughout the educational system should not be overlooked. These simultaneous changes throughout the education system included shifts in external funding, shifts in the dedication of departmental resources, shifts in support from university administrators, expanding partnerships across STEM departments, expanding partnerships between the School of Education and School of Arts and Sciences, and expanding partnerships between the physics department and other university programs across campus. By broadening our lens and examining faculty within the situations that they work, we saw that structural changes in how things were done were coupled with changes in how individual faculty members’ engaged in teaching their courses. These findings call into question the common assumption of dissemination approaches that focus solely on individual faculty members’ adoption and individual use of curricular materials. These findings suggest that individual faculty members involved in these change processes may also need to learn to build partnerships and act as advocates within their institutional contexts to achieve sustainable change. Furthermore, approaches to educational change must coordinate and address multiple levels of the educational system simultaneously.

Based on a frames-of-context analysis and our preliminary observations in Tutorial recitations, TA/LA Tutorial Training Meetings and associated course lectures (presented in Chapter 4), we were able to define what it means to implement the Tutorials at CU and the particular ways in which we might expect implementation to vary. For example, there are choices that faculty make at the course level such as syllabus descriptions and the design of assessments that effect Tutorial implementation. There are choices at the level of the TA/LA training meetings where the secondary instructor’s practices frame what the purpose of the Tutorial activities is for TAs and LAs. There are also activities that exist at the same scale as the Tutorial recitations such as the lecture activity and out-of-class homework activities that may support the goals of the Tutorials or sit in tension with the Tutorials. All of these decisions (both nuanced and nested) by physics faculty influence Tutorials implementation.

Research Question #5: What justifications do physics professors give for initially using the Tutorials?

As discussed in Chapter 4, faculty members’ early decisions to adopt the Tutorials were complex due to team-teaching assignments. At times, a single faculty member decided to adopt the Tutorials independently from his/her co-teacher, while at other times team-teachers
collaborated in deciding to adopt the Tutorials. Many faculty members openly admitted not having looked at the actual Tutorial activities prior to deciding to use them. Faculty members who felt involved in the decision to adopt the Tutorials expressed the following reasons for their decision: 1) lecturing in recitation is largely ineffective, 2) local data on student learning were compelling, 3) tutorials relieve pressure from under-prepared TAs, 4) specific pedagogical aspects of the curriculum such as peer discussion and attention to student misconceptions, and 5) the availability of help and resources in using the Tutorials. These results suggest that some initial individual dissatisfaction with current instructional methods may be critical in decisions to try something new, locally collected data maybe particularly compelling for other faculty members, and support and resources offered for the new methods may be persuasive (although the timing of these offers may be critical). Corroborating the findings from the analysis of institutional change, the results of this interview study suggests the importance of developing local structures that support reflective practice of faculty—specifically the sharing of assessment results. Encouraging reflective practices and the sharing of assessment results may be particularly supported by the presence of a physics education research group, although participation of and communication with the broader faculty also seems important.

Directions for Future Work

Our investigations highlight possible benefits of particular implementations of Peer Instruction that have yet to be explored and assessed. The assessment of students’ facility with scientific practices as a result of different instructional interventions is a fruitful direction for future research. Possible research questions in the context of Peer Instruction include

- Can we measure other possible impacts of PI beyond conceptual understanding—e.g. how students justify their reasoning or verbally communicate their scientific reasoning?
- Do particular aspects of PI implementation impact students’ proficiency at particular scientific practices such as asking scientific questions, evaluating the correctness or completeness of problem solutions, communicating in a public arena?
- How does PI implementation vary across multiple institutions or across various types of institutions?
- In parallel to PI studies, how might one develop definitions of norms based on classroom practices in the context of Tutorials?

This thesis develops an observation rubric and a coding framework for capturing PI implementation practices as well as identifies critical classroom norms related to PI. The research presented in this thesis sets the stage for addressing these research questions in future work.

In another vein of research, additional case studies of educational change examining multiple levels of the educational system would be helpful in drawing attention to critical elements of change processes more generally. In hopes of better understanding Tutorial implementation, possible research studies could examine the following questions:

- How have the Tutorials been adopted and sustained at other institutions?
- What are the common elements of change processes associated with successful (and unsuccessful) Tutorial adoption and institutionalization?
- Do students’ at other institutions enjoy working on the Tutorials? If so, what associated contextual differences may account for students’ enjoyment?
- How are TAs and LAs influenced by the Tutorial training meetings?
• How are the Tutorials coordinated with other elements of the course?

Again, this thesis provides tools for investigating this second set of future research questions through the presentation of a framework for documenting Tutorial implementation and changes across multiple levels of the educational system.

At a more theoretical level, this thesis project suggests how a socio-cultural lens may address outstanding theoretical concerns raised by current studies:

• Can we expand our focus on student learning beyond “content knowledge” to include identity development, facility with scientific practices, student affect, attitudes and beliefs, etc.?
• Do current research-based instructional strategies support students’ development of scientific practices?
• Can we develop mechanistic models of instructional activities influence students’ development?
• Can we develop research-based models to guide educational changes efforts?

There is a common adage in physics education research, “It is not about our teaching; it is about students’ learning.” Although this adage did productive work in shifting the focus in education (and education research) onto what was going on with the students, we now need to establish not only a theory of student development, but a theory of how instructional activities influence students' developmental trajectories. As a community, we have serious work ahead of us to bring together theories of student development with theories of instructional design.
References (Chapter 6)

Bibliography


D. Ball & A. Schoenfeld (Eds.), *A Study of Teaching: Multiple Lenses, Multiple Views, JRME Monograph #14*. (Journal for Research in Mathematics Education Monograph Series, Reston, VA, 2008).


D.M. Hanson, *Instructor’s guide to process-oriented guided-inquiry learning* (Pacific Crest, Lisle, IL, 2006).


L.R. Herrenkohl and M.R. Guerra, “Participant structures, scientific discourse, and student engagement in fourth grade.” Cognition and Instruction, 16(4), 431-473.


I. Horn, “Accountable argumentation as a participant structure to support learning through disagreement.” In D. Ball & A. Schoenfeld (Eds.) A Study of Teaching:


N. Lasry, “Clickers or flashcards: Is there really a difference?” *Physics Teacher* 46, 242-244 (2008).


C. Turpen and N. Finkelstein, “Not all Interactive Engagement is the same: Variation in physics professors’ implementation of Peer Instruction.” Physical Review Special Topics, Physics Education Research 5, 020101 (2009).


University of Colorado Course catalog: Available at http://www.colorado.edu/catalog/catalog08-09/search.pl?abbr=PHYS&num=


### CHAPTER 2 APPENDICES

Appendix A: Instructor Attributes and Characteristics of Instructor Practices by Stage

#### Table I: Course and Professor Attributes

<table>
<thead>
<tr>
<th>Course</th>
<th>Primary student major</th>
<th>Experienced or novice clicker user</th>
<th>Temporary Instructor or Faculty</th>
<th>Phys. Ed. Researcher (Y/N)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow Calc-based</td>
<td>Science</td>
<td>Experienced</td>
<td>Faculty</td>
<td>N</td>
<td>Previously mentored by Blue</td>
</tr>
<tr>
<td>Phys1 Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First time teaching a large-enrollment course</td>
</tr>
<tr>
<td>Green Calc-based</td>
<td>Science</td>
<td>Novice</td>
<td>Instructor</td>
<td>N</td>
<td>First time teaching a large-enrollment course</td>
</tr>
<tr>
<td>Phys2 Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nationally recognized for teaching excellence</td>
</tr>
<tr>
<td>White Alg-based</td>
<td>Science</td>
<td>Novice</td>
<td>Instructor</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Phys1 Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue Alg-based</td>
<td>Science</td>
<td>Experienced</td>
<td>Faculty</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Phys2 Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purple Physics</td>
<td>Non-science</td>
<td>Experienced</td>
<td>Faculty</td>
<td>Y</td>
<td>None</td>
</tr>
<tr>
<td>elective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red Calc-based</td>
<td>Science</td>
<td>Experienced</td>
<td>Faculty</td>
<td>Y</td>
<td>None</td>
</tr>
<tr>
<td>Phys3 Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Table II: Characteristics of Practice—CQ Set Up Stage

<table>
<thead>
<tr>
<th></th>
<th>Yellow</th>
<th>Green</th>
<th>White</th>
<th>Blue</th>
<th>Purple</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clickers part of course grade</td>
<td>No</td>
<td>No</td>
<td>Yes (5%)</td>
<td>Yes (1%)</td>
<td>Yes (5%)</td>
<td>Yes (15%)</td>
</tr>
<tr>
<td>CQs points (for grade) based on correctness</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>Yes (3:1)</td>
<td>No</td>
<td>Rarely</td>
</tr>
<tr>
<td>Extra Credit (EC) point from clickers</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>EC points based on correctness</td>
<td>Yes (2:1)</td>
<td>Yes (3:1)</td>
<td>NA</td>
<td>NA</td>
<td>Yes (2:1)</td>
<td>NA</td>
</tr>
<tr>
<td>CQs with answers available</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CQs answer explanations provided</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>&lt;No. CQs&gt;/hr</td>
<td>5.9 ± 0.5</td>
<td>3.2 ± 0.1</td>
<td>8.2 ± 0.6</td>
<td>6.5 ± 0.4</td>
<td>5.0 ± 0.3</td>
<td>5.3 ± 0.4</td>
</tr>
<tr>
<td>% of students correct</td>
<td>69 ± 2%</td>
<td>68 ± 2%</td>
<td>64 ± 3%</td>
<td>76 ± 1%</td>
<td>70 ± 2%</td>
<td>72 ± 4%</td>
</tr>
<tr>
<td>Median % of students correct</td>
<td>73%</td>
<td>72%</td>
<td>65%</td>
<td>82%</td>
<td>75%</td>
<td>80%</td>
</tr>
<tr>
<td>Fraction of logistical CQs</td>
<td>0.06</td>
<td>0.00</td>
<td>0.02</td>
<td>0.10</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>Fraction of recall CQs</td>
<td>0.20</td>
<td>0.04</td>
<td>0.15</td>
<td>0.06</td>
<td>0.09</td>
<td>0.24</td>
</tr>
<tr>
<td>Fraction of algorithmic CQs</td>
<td>0.06</td>
<td>0.11</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Fraction of conceptual CQs</td>
<td>0.69</td>
<td>0.85</td>
<td>0.75</td>
<td>0.77</td>
<td>0.79</td>
<td>0.64</td>
</tr>
</tbody>
</table>
### Table III: Characteristics of Practice—CQ Response Stage

<table>
<thead>
<tr>
<th></th>
<th>Yellow</th>
<th>Green</th>
<th>White</th>
<th>Blue</th>
<th>Purple</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of CQs where the professor left the stage</td>
<td>12 ± 4%</td>
<td>11 ± 6%</td>
<td>57 ± 7%</td>
<td>10 ± 5%</td>
<td>81 ± 7%</td>
<td>69 ± 8%</td>
</tr>
<tr>
<td>Percent of CQs where the professor answered a student question</td>
<td>19 ± 5%</td>
<td>25 ± 8%</td>
<td>17 ± 5%</td>
<td>23 ± 7%</td>
<td>74 ± 8%</td>
<td>63 ± 9%</td>
</tr>
<tr>
<td>Percent of CQs where the professor discussed with the students</td>
<td>8 ± 4%</td>
<td>0%</td>
<td>41 ± 6%</td>
<td>15 ± 6%</td>
<td>55 ± 9%</td>
<td>84 ± 6%</td>
</tr>
<tr>
<td>Average time given to students to answer the CQ (seconds)</td>
<td>133 ± 4</td>
<td>149 ± 6</td>
<td>100 ± 5</td>
<td>124 ± 7</td>
<td>124 ± 10</td>
<td>153 ± 10</td>
</tr>
<tr>
<td>Standard deviation in avg. time given to respond (seconds)</td>
<td>59</td>
<td>61</td>
<td>69</td>
<td>89</td>
<td>103</td>
<td>116</td>
</tr>
</tbody>
</table>

### Table IV: Characteristics of Practice—CQ Solution Discussion Stage

<table>
<thead>
<tr>
<th></th>
<th>Yellow</th>
<th>Green</th>
<th>White</th>
<th>Blue</th>
<th>Purple</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average time discussing solution (minutes:seconds)</td>
<td>3:04 ± 0:18</td>
<td>3:16 ± 0:30</td>
<td>1:10 ± 0:10</td>
<td>2:28 ± 0:18</td>
<td>3:26 ± 0:35</td>
<td>3:42 ± 0:26</td>
</tr>
<tr>
<td>Percent of CQs where the incorrect answers were discussed</td>
<td>19 ± 5%</td>
<td>35 ± 9%</td>
<td>18 ± 5%</td>
<td>26 ± 7%</td>
<td>61 ± 9%</td>
<td>58 ± 9%</td>
</tr>
<tr>
<td>Percent of CQs where student explanations were heard</td>
<td>17 ± 5%</td>
<td>100%</td>
<td>2 ± 2%</td>
<td>10 ± 5%</td>
<td>50 ± 9%</td>
<td>55 ± 9%</td>
</tr>
<tr>
<td>When student explanations are requested, average number of student explanations heard per CQ</td>
<td>2.2 ± 0.2</td>
<td>1.4 ± 0.1</td>
<td>NA</td>
<td>1.3 ± 0.3</td>
<td>2.3 ± 0.5</td>
<td>2.4 ± 0.4</td>
</tr>
<tr>
<td>Average number of student explanations heard per hour of class</td>
<td>2.4 ± 0.6</td>
<td>4.6 ± 0.6</td>
<td>0.1 ± 0.1</td>
<td>0.6 ± 0.4</td>
<td>4.2 ± 0.5</td>
<td>4.8 ± 1.3</td>
</tr>
</tbody>
</table>
Appendix B: Descriptions of Dimensions of Practice

Defining the Academic Task

1. The Role of Clicker Use within the Organization of the Course: This dimension of practice describes how the use of clickers is integrated or coordinated with broader course activities and evaluation. As described in How People Learn, “A systems approach to promote coordination among activities is needed to design effective learning environments” (pg. 152).\(^1\) Sometimes the role or function of clickers within the course is explicitly described in the course syllabus, and at other times professors spend class time during the first few days of the semester verbally describing to the students how and why clickers will be used. Beyond these occasional explicit statements, the role of clickers may be reflected through classroom practices in terms of the relative amount of time dedicated to CQs or the degree to which CQs are aligned with other course components such as exam assessments, recitation activities, or homework assignments. Since course grades hold significant weight in our schooling systems, the function of clickers may also be described by its relative contribution to students’ overall course grade. In addition, how clickers are graded may help to communicate what is valued about the activity, i.e. participation or the correctness of the response.\(^{ii,iii,iv}\) These descriptions concerning the role of clicker use within the course can help researchers determine the importance of clicker use relative to other course components.

2. Types of Clicker Questions: We have observed physics instructors using clickers to ask two very different types of questions: logistical questions and content questions. Logistical questions include: gathering student input on homework due dates, polling student availability for scheduling office hours, inquiring about whether students have completed some logistical task, or checking students’ awareness of resources. These questions are not necessarily specific to the physics content, but provide an opportunity for professors to use student input to inform the structure and direction of the course. Content questions are CQs that involve physics and these may be divided into various subtypes. Subtypes of content questions have been discussed elsewhere and many different schemes have been proposed.\(^{v,vi,vii,viii,ix}\) These subtypes are usually defined based on the literal meaning of the question statement removed from the context of the question’s specific implementation.

---


According to Holyoak, “a task analysis of problems can provide information about constraints that the nature of the problem imposes on the nature of the problem solver” (pg. 268).\textsuperscript{x} The wording of the question partially defines the task that the student is expected to accomplish, i.e. calculate x, interpret graph 1, etc. We have only investigated a coarse-grained categorization based on whether the question was a recall, algorithmic, or conceptual question because we found that the way the question was asked in practice strongly impacted the nature of the students’ task depending, for example, on the information that was provided directly prior to asking the question. A recall question is a question in which the primary task of the students is to remember a definition, equation, or single fact without requiring multiple steps of reasoning. An algorithmic question primarily asked the students to complete a mathematical procedure which required linking mathematical definitions with symbols and possibly multi-step symbolic manipulation. Usually requiring qualitative reasoning, conceptual questions asked student to complete tasks such as a multi-step analysis, an application of a concept, interpretation of graphs, translating words into symbolic representations, or selecting between possible solution strategies. According to this categorization, a subset of the conceptual questions resemble ConcepTests as discussed by Mazur et al.\textsuperscript{xi} Our interest in the types of questions students are asked stems from the National Research Council’s finding that “in science, existing curricula tend to overemphasize facts” (pg. 136) and further warning that “stress on isolated parts can train students in a series of routines without educating them to understand the overall picture that will ensure the development of integrated knowledge structures and information about the conditions of applicability” (pg. 139).\textsuperscript{xii}

3. Introduction of the Clicker Question: In addition to the structure provided by the wording of the CQ itself, the CQ is framed by how the professor talks about and introduces the CQ. One important way that the question is defined is by the activity that precedes the question. Introductory verbal comments by the instructor appear to be particularly important in framing the question for the students. For further discussion on the teacher’s role in initiating changes from one activity to another and its impacts on students, see Lemke’s discussion in Talking Science.\textsuperscript{xiii} Through introductory comments, the professor frames what he/she sees as the purpose of the CQ: A check for understanding question, an application question, a prediction question, a real-world connection question. For example, professors may introduce a CQ in one of the following ways: “This is a quickie.” “This isn’t even a physics question; it’s just a math question to see if we are on the same page.” “This is a tricky one so make sure you check with your neighbors.” At times the professor may first put up the statement of the CQ without the multiple-choice answer options to avoid an immediate focus on the answers provided and try to avoid the students’ urge to reason by process of elimination. At other times the professor may verbally remind students that their reasoning is what is important and not just the answer. Based on our observations, the CQ task and how it is perceived by


the students can change based on what information the professor has introduced to the students prior to the question.

4. **Student-Student Collaboration:** Student-student collaboration delineates the degree to which students are allowed and/or encouraged to discuss their responses to the CQ with their peers. We note that this is an explicitly stated step in *Peer Instruction* as described by Mazur.\textsuperscript{xiv,xv} Additionally, it describes the level of actual student discussion as gauged by the volume of noise in the lecture during the CQs and samples of student conversation overheard locally by the classroom observer. According to the recent article in Science magazine, Smith et al report, “Our results indicate that peer discussion enhances understanding, even when none of the students in a discussion group originally knows the correct answer.”\textsuperscript{xvi} Due to the impacts of peer discussion on student understanding, we are interested in when peer discussion is allowed and encouraged by the instructor.

5. **Determining Time Constraints:** One task that the professor is generally engaged in during the implementation of clickers is determining time constraints. In LeCompte’s qualitative study of four fourth grade classroom, she found that teachers spent approximately 20% of their non-instructional time verbally signaling the beginnings and endings of activities and maintaining a schedule in the classroom.\textsuperscript{xvii} During the CQ response time, the professor determines the length of time within which the students must respond to the question. Some professors state verbally at the beginning of the question the amount of time the students will be given to respond. Other professors watch the votes come in on an electronic receiver and begin warning the students to finish up when approximately seventy-five percent of the students have voted. As reflected in interviews, instructors continually take into account what they intend to accomplish during that class period and how the CQ is positioned relative to the end of the class period. In addition to determining the length of the voting time, the professor must also decide how long to spend discussing the solution to the CQ. These decisions include managing whether to hear student explanations, how many students to hear from, and how to respond to disagreement amongst the students.

**Student-Professor Interactions**

6. **Creating or Reducing Spatial Boundaries between Students and Instructor:** Historically, most university classrooms have a ‘front’ of the classroom that is almost entirely occupied by the instructor and an ‘audience’ section that is entirely occupied by the students. In large enrollment courses, this ‘front’ area is even more pronounced and appears as a stage area so that all students can see the instructor.\textsuperscript{xviii} Some professors use the CQ response time to leave the front stage area of the classroom and walk amongst the students in the audience seating. Wandering outside of the stage area has the result of reducing spatial boundaries between the instructor and the students. Reducing spatial boundaries between students and the professor

---


makes the professor more accessible to students. Rather than having to get the professor’s attention in a loud or attention drawing manner, a student might have opportunities to get the professor’s attention in passing when the professor and the student are in close proximity. This also provides an opportunity for students to have a ‘private’ conversation with the professor without the pressure of a public dialogue in a large enrollment course. Physical distance between people has been shown to connote varying degrees of formality and familiarity as well as impede or facilitate verbal interaction and the perception of facial expressions.\textsuperscript{xix}

7. \textit{Listening to Student Explanations and Comments:} During the CQ response time, the professor can listen to student reasoning in many ways. Some professors choose to not leave the stage, but still pause to intentionally overhear the discussions of students in the first few rows near the stage. Other professors leave the stage area and pursue opportunities to listen to student discussions across the room. By listening to student explanations during the CQ response time, the professor has an opportunity to hear examples of student reasoning. This may be useful if the professor has limited knowledge of common student ideas prior to instruction or research on student conceptions that can impede the development of scientifically accurate conceptions. Listening to students’ explanations during peer discussion can expose professors to the language that students use to express physical ideas and also key the professor into potential gaps in student reasoning. On other occasions, listening to students’ explanations can expose professors to ways that students can get the correct answer based on incorrect or incomplete reasoning. The professor may also listen to student explanations during the solution discussion stage. Through these examples of student reasoning the professor can lead a discussion of multiple pathways to the solution. Each of these practices make student thinking visible to the educator, as encouraged in How People Learn (pg. 140).\textsuperscript{xx}

8. \textit{Faculty-Student Collaboration:} Beyond simply listening to students’ explanations, the professor may decide to interact directly with the students during the CQ response time and the CQ solution discussion stage. This collaboration may take many forms: Socratic dialogue, clarification of CQ wording, argumentation for alternative answer options with the students, or requests for the students to clarify or elaborate on their explanations. Through modeling by the professor, the students may benefit from collaborating with the professor in many ways: learn to debate, argue, or defend ideas, reason logically, question assumptions, and to be skeptical or critical of reasoning processes or ideas.\textsuperscript{xxi} Research on upper elementary mathematics classrooms has shown that different types of educator-student collaboration can result in dramatic differences in student learning.\textsuperscript{xxii}

9. **Instructor’s Use of Student Prior Knowledge:** The professor may connect the current CQ discussion to common experiences from the students’ everyday life, or students’ knowledge from prior instruction. The professor may have access to student prior knowledge based on previous experiences teaching this course or this student population, awareness of or exposure to literature on common student ideas surrounding a specific physics concept, immediate examples of student thinking from listening during the response time, or requests for and use of student explanations during the solution construction time. In this way, the instructor may use student prior knowledge by having the students contribute their ideas directly during class, or the professor may use his/her own knowledge of student reasoning in his/her description of the solution. The professor may demonstrate his/her awareness of student prior knowledge through the use of informal, colloquial language, or common student language. Additionally, the professor’s description of the correct CQ answer or any possible discussion of incorrect answers may reflect specific aspects of students’ reasoning and prior understanding. The importance of building on students’ prior knowledge was featured as a key finding in the National Research Council’s How People Learn publication: “Students come to the classroom with preconceptions about how the world works. If their initial understanding is not engaged, they will fail to grasp new concepts and information that are taught…” (pg. 14).

10. **Use of student voice:** Student voice may arise in multiple ways during CQs. The professor may actively encourage students to talk to each other while they are arriving at a CQ answer. In this way, an ‘allowed’ space is made for students to be able to talk to each other about the content. Another important aspect of student voice is whether students participate in the construction of the solution during the whole class CQ solution discussion stage through the professor’s requests for students to provide explanations of reasoning for their selected answer options or requests for students to comment or ask questions about topics related to the CQ. Giving students a role in the explanation of the CQ solution provides opportunities to empower students, give students alternative wordings or framings on solution strategies, and engage students in providing descriptions of their reasoning process to their fellow students. This can potentially frame students as sources of scientific answers and explanations. In K-12 mathematics classrooms, research have shown that students develop different levels of autonomy with mathematics depending on the classroom opportunities for authentic mathematical participation.

11. **Management of disagreement amongst students:** Some professors only emphasize the correct answer to the CQ even in cases when there are competing answer options (i.e., incorrect answer options that 10% or more of the students have chosen). Usually this practice seems to be motivated by a need to either reduce the time spent discussing the topic or to avoid occasions that may confuse the students. In other cases, professors address alternative answer options in a limited way. Usually by describing a common mistake, the professor

---


illustrates how the mistake leads to a particular incorrect answer. Some professors provide space for ‘multiple’ plausible interpretations of the CQ which at times justifies different correct answers. Some professors utilize disagreement amongst the students as a rich opportunity to engage students in active debate and scientific argumentation, withholding expert evaluation until significant progress and consensus has developed through class discussion. This tactic may be employed to engage students in ‘authentic’ scientific practices of public discussion, debate, and argumentation. The benefits of these kinds of discussions and arguments for students have been discussed in the literature.\textsuperscript{xxvi} Whole class discussions can, however, take significant amounts of class time and, in practice, professors must balance the benefits of arriving at a jointly constructed class consensus with the perceived risks of not covering sufficient amounts of material.

12. \textit{Formative Use of Students’ Ideas and Aggregate Student Responses}: After listening to student explanations or collaborating with students, the professor can spend time during the response time considering what he/she has learned from these interactions and how to use this knowledge in the immediate progression of the class. In this way, the professor can consider various strategies for addressing common ideas in the class, whether these ideas are correct or incorrect. The professor may even ask students in private interactions if they would be willing to share their ideas with the whole class so that other students could benefit from hearing their explanation. The professor may decide to spend more or less time on the concept related to the question at hand based on his/her knowledge of the state of student understanding. Some professors also use the real-time histograms available from the clicker software program to anticipate challenges that are present within the aggregate student responses. Formative use of student responses is usually reflected in improvisations or spontaneous changes to the flow of classroom activity. There seems to be at least two different common types of student response distributions: questions where 90\% or more of the students are getting the question correct or questions where at least 10\% of the students are choosing each of multiple competing answer options. Formative use of these types of results can be manifestly different. When the overwhelming majority of students are answering the question correctly, the professor may decide to not describe the solution, to have a student briefly describe their reasoning, to move on to a new topic, to change the pace of the course, or to ask a more challenging related question. In the case where there are multiple competing answer options, the professor may alter the flow of classroom activity by: asking students to share or clarify their reasoning, asking the same question again after some clarification, asking another related question to verify student understanding, spending additional class time discussing or lecturing on a core concept related to the poor CQ performance. Occasionally, we have observed professors making more substantial changes to the flow of the entire course by: adding or changing the assigned homework problems, altering the recitation or tutorial schedule, or placing additional questions on reviews or exams. A research review by Black and Wiliam has shown that, “innovations that include

strengthening the practice of formative assessment produce significant and often substantial learning gains” (pg. 2).xxvii

13. **Summative Use of Aggregate Student Responses:** Professors also use aggregate student responses as summative assessment to gather a concise summary of student cumulative understanding. Sometimes this is done explicitly through evaluative reading quizzes or unit quizzes, but often clickers are implicitly used for summative assessment. The professor may finish a topic with a few summary or review questions that require students to pull multiple small ideas together. In this way, a summative use of CQ responses may not necessarily mean that the question is graded or evaluated. The professor may use these questions to close one topic and signal that the class will be moving on to new topics.

---

## Appendix C: Classroom Observation Rubric

### Introduction of Question:

- **Time Log**
  - Start Intro:
  - Start CQ Time:
  - End CQ Time:
  - End Discussion:

### Description of Question:

### CQ Response Stage:
- What did professor do or say?
- What did students do or say?

### CQ Solution Discussion Stage:
- How professor conducts explanation:
  - Student explanations heard
  - Number of Volunteers to explain
  - Number of students heard
  - Incorrect options discussed

### Distribution of Student Responses:
- **A:**
- **B:**
- **C:**
- **D:**
- **E:**

### Dialogic Interactions (Student-student and Student-professor):
- **A:**
- **B:**
- **C:**
- **D:**
- **E:**
Appendix D: Classroom Observation Rubric User’s Guide

In this framing, the implementation of a clicker question (CQ) happens in three stages:

- The CQ Set Up Stage: During which the professor introduces the CQ to the class,
- The CQ Response Stage: During which the students (and/or the professor) are constructing an answer to the question,
- The CQ Solution Discussion Stage: During which the class comes to a public solution to the question.

It is useful to describe the activity which immediately preceded and followed the implementation of a CQ. These descriptions can help the researcher to infer what the purpose of the CQ might have been.

Materials

For each observation, bring the following materials: a digital recorder, an extra set of batteries, approximately 10 copies of the observation rubric per course, one page for aggregate notes, and additional sheets of blank paper for notes. (The blank sheets of paper are used to write down descriptions of other interactive engagement techniques used by the professor, even if clickers are not being used.)

Preparation

Fill in the title information on multiple sheets before the beginning of class.

Explicit Descriptions of rubric

Note: The following notes are collected for all CQ for which the students vote however, some of the codes do not apply for logistical questions.

Time Log: Start the audio-recorder when the professor first instigates an interaction with the entire class. Note the actual time that the recorder was started on the top of the first CQ sheet. Use the running time on the recorder to record the time log. For example, a researcher would write that the CQ intro started at 2m23s, and the beginning of the student response time was at 2m53s, etc. This way the researcher can go back to the digital recording and find that section of interaction on the tape. It is not important to get these numbers completely accurately, because the researcher can go back with the audio recording and adjust them for accuracy, but having a rough idea of the beginning and end of a question is useful. Write down that the discussion ended after the conversation has moved away from the specifics of the CQ.

Introduction of Question: Primarily focus on if the professor reads the question verbatim, if the professor elaborates the question as he presents it, if the professor asks explicitly if the students understand the question, etc. Also note if the professor verbally reminds students about what they should be doing during the response time: talk to your neighbors, focus on the reasoning… These descriptions can be brief.

Description of Question: Here write down the question that is asked. It was useful for the researcher to know if professor puts their CQs online, because then their recording of this section can be less thorough.
Description of Answer Options: Here record the answer options provided to the students.

Distribution of Student Responses: Here record how many students or the percent of students which fall into each answer bin, depending on how the information is displayed. Again depending on the clicker software, it may be more or less important to capture this information here. Many clicker software programs capture this information for the researcher to consult at a later time.

CQ Response Stage: What did the professor do?: Here describe in words what the professor does during the question response period. Where does the professor stand? What does he do with his hands? Is he arranging other logistical course stuff? Also check the boxes below to show if the professor leaves the stage area during the response time. If the professor answers a student question during the response time, start tallying how many students ask questions. Student questions are occasions where the student initiates an interaction with the professor usually by having his/her hand raised. If the professor initiates an interaction with students, start tallying the number of groups that the professor interacts with. The professor discussing with students are occasions where the professor initiates an interaction with a small group of students. The nature of the dialogue during these interactions is not captured, but at least the interaction is not started on the premise of getting some information or assistance from the professor.

CQ Response Stage: What did the students do?: Check either the individual box or the group box depending on if the question is answered by students together or if they are told to answer the question on their own. ‘Answered as a group’ was checked if there was any observed discussion between students (even if they could, in the end, disagree and vote differently). Then, describe anything striking about the groups sitting immediately around the researcher.

CQ Solution Discussion Stage: How the professor conducts the explanation?: Check the student explanations heard box if the professor calls on students for student explanations. Tally the number of students who have their hands in the air when the professor calls for student volunteers in the next box labeled, Number of Volunteers to explain. Then, tally the number of students that are called on to speak out for a given question. Also check the box, incorrect options discussed, if the professor or the students address reasoning for alternative answer options. The professor or students had to spend some time explicitly discussing a reason for choosing an alternative answer option, for this box to be checked. If they do, start tallying the number of different answer options that the professor and the class talk about explicitly.

Dialogic Interactions: (Student-student or student-professor interaction diagrams) For each student that speaks, draw a circle and label that circle S1 to represent student one. Then, use arrows only to show a direct comment to another person in the classroom. If the student is just espousing an explanation to the whole class that isn’t followed up on by the professor, then do not draw an arrow connecting to it. Most interactions involve the professor addressing the student to give an explanation. The professor may then respond to the student by paraphrasing what the student said or may ask the student a question so that the professor better understands the student’s ideas. Each of these direct addresses is captured in the diagram by indicating an arrow. When the professor calls on another student, that student may couch their ideas by either
agreeing or disagreeing with a previous student; use an arrow to show this connection. See the example below.

The idea behind this diagramming is to document who is talking to whom or verbally referencing other class participants. One way to see this type of conversation is to see if what is said in class is building on the ideas, explanations, or arguments of others. Again, this diagramming can be filled in somewhat from the audiotape file. However it seems to be important to accurately capture the number of students contributing on the first pass, since it can be difficult to tell students’ voices apart on the recording.

1. Professor calls on S1
2. S1 gives a short explanation
3. Professor asks S1 a short clarifying question
4. S1 replies shortly
5. Professor calls on S2
6. S2 directly says that he disagrees with S1 and gives a reason
7. Professor calls on S3, S3 gives their reason for choosing b
8. Professor calls on S4
9. S4 says that she disagree with S2 and gives a short reason.

The class seems to have arrived at a consensus that S1 and S4 are correct concerning answer a.
CHAPTER 3 APPENDICES
Appendix A: Survey questions used to document students’ perceptions of Peer Instruction

For how many semesters have you used clickers at CU?
○ 1 (this is my first semester)
○ 2 - 4 semesters
○ 5 or more semesters

(Q2) Overall, how do you feel about this course?
○ Awful course
○ Not a good course
○ Neutral
○ Good course
○ Great course

(Q3) Regardless of how you feel about it, how much did you learn in this course?
○ Almost nothing
○ A small amount
○ A reasonable amount
○ Quite a bit
○ A great deal

(Q4a) Please rate how useful for your learning each of the following class activities are in this course:
Pure Lecture Presentation (only instructor talking)
○ Completely useless
○ Mostly useless
○ Somewhat Useful
○ Useful
○ Very Useful

(Q4b) Please rate how useful for your learning each of the following class activities are in this course:
Use of Clickers
○ Completely useless
○ Mostly useless
○ Somewhat Useful
○ Useful
○ Very Useful

(Q4c) Please rate how useful for your learning each of the following class activities are in this course:
Textbook or Course Readings
○ Completely useless
○ Mostly useless
○ Somewhat Useful
○ Useful
○ Very Useful

(Q4d) Please rate how useful for your learning each of the following class activities are in this course:
Homework and/or Course Assignments
○ Completely useless
○ Mostly useless
○ Somewhat Useful
○ Useful
○ Very Useful

(Q5a) Please rate how much you enjoyed each of the following class activities in this course [where "enjoyed" is meant in the sense of academic or intellectual pleasure]:
Pure Lecture Presentation (only instructor talking)
○ Strongly Dislike
○ Dislike
○ Neutral
○ Like
○ Strongly Like
(Q5b) Please rate how much you enjoyed each of the following class activities in this course [where "enjoyed" is meant in the sense of academic or intellectual pleasure]:

- Use of Clickers
  - Strongly Dislike
  - Dislike
  - Neutral
  - Like
  - Strongly Like

(Q5c) Please rate how much you enjoyed each of the following class activities in this course [where "enjoyed" is meant in the sense of academic or intellectual pleasure]:

- Textbook or Course Readings
  - Strongly Dislike
  - Dislike
  - Neutral
  - Like
  - Strongly Like

(Q5d) Please rate how much you enjoyed each of the following class activities in this course [where "enjoyed" is meant in the sense of academic or intellectual pleasure]:

- Homework and/or Course Assignments
  - Strongly Dislike
  - Dislike
  - Neutral
  - Like
  - Strongly Like

(Q6) To what extent does your instructor usually encourage student-to-student discussion about clicker questions in class?

- Does not allow discussion
- Allows discussion, but does not encourage it, and a small fraction of students discuss
- Allows discussion, but does not encourage it, and a large fraction of students discuss
- Encourages discussion, and a small fraction of students discuss
- Encourages discussion, and a large fraction of students discuss

(Q7) When your instructor gives your class a typical clicker question AND you are allowed to talk with others, what do you usually do?

- Does not apply--we are usually not allowed to talk with other students
- I rarely use a clicker in this course
- I guess the answer and do not check with other students
- I actively think about the question independently and arrive at an answer without speaking or listening to other students
- I listen to other students' answers and/or reasoning
- I actively participate in discussions with other students around me

(Optional:) Why? Please explain in the space below.

(Q8a) Please rate how useful for your learning the following types of clicker questions are:

- Challenging conceptual questions, where I have to think and apply the ideas presented
  - Completely useless
  - Mostly useless
  - Somewhat Useful
  - Useful
  - Very Useful

(Q8b) Please rate how useful for your learning the following types of clicker questions are:

- Questions that require me to recall a fact from the text or a prior lecture
  - Completely useless
  - Mostly useless
  - Somewhat Useful
  - Useful
  - Very Useful

(Q8c) Please rate how useful for your learning the following types of clicker questions are:

- Questions that require me to recall a fact that was just stated
(Q8d) Please rate how useful for your learning the following types of clicker questions are:
Questions where I have to plug numbers into an equation
○ Completely useless
○ Mostly useless
○ Somewhat Useful
○ Useful
○ Very Useful
(Optional:) Please comment in the space below with as much detail as possible.

(Q9) Would you recommend that other instructors, who teach this course in the future, use clickers?
○ Definitely Not Recommend
○ Not Recommend
○ Neutral
○ Recommend
○ Definitely Recommend
(Optional:) Please comment in the space below with as much detail as possible.

(Q10) If my professor were to approach me in class during a clicker question, I would be comfortable discussing the content with my professor.
○ Strongly disagree
○ Somewhat disagree
○ Not sure
○ Somewhat agree
○ Strongly agree

(Q11) It is awkward to ask my professor questions during class.
○ Strongly disagree
○ Somewhat disagree
○ Not sure
○ Somewhat agree
○ Strongly agree
(Optional:) Why? Please explain in the space below.

(Q12) How often do you speak directly to the professor during class?
○ Never
○ Once or twice a semester
○ Once every few weeks
○ Nearly every week
○ Nearly every class

(Q13) Knowing the right answers is the only important part of the clicker questions.
○ Strongly disagree
○ Somewhat disagree
○ Not sure
○ Somewhat agree
○ Strongly agree
(Optional:) Why? Please explain in the space below.

(Q14) How comfortable do you feel discussing the course content with your peers during clicker questions?
○ Not comfortable
○ Somewhat uncomfortable
○ Not sure
○ Somewhat comfortable
○ Very comfortable

(Q15) On average, what fraction of class time do students speak either with each other, or to the professor?
○ Less than 5 minutes out of a 50-minute class
○ About 5 minutes out of a 50-minute class
○ About 10 minutes out of a 50-minute class
○ About 15 minutes out of a 50-minute class
○ About 20 minutes out of a 50-minute class
○ More than 20 minutes out of a 50-minute class

(Q16) How often do you raise your hand or ask questions in class?
○ Never
○ About once a semester
○ About once a month
○ Nearly every week
○ Nearly every class

(Q17) In class, how important is it for you to articulate your reasoning either to your peers or during whole class discussion?
○ Not important at all
○ Not very important
○ Somewhat important
○ Important
○ Very Important
(Optional:) Why? Please explain in the space below.
**Appendix B: Kruskal-Wallis p-values for set of three semesters of data**

<table>
<thead>
<tr>
<th>Question Number</th>
<th>p-value</th>
<th>Pair-wise significance thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q3</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q4a</td>
<td>p=0.002</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q4b</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q5a</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q5b</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q6a</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q6b</td>
<td>p=0.03</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Q7</td>
<td>p=0.006</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q8a</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q8d</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q10</td>
<td>p=0.003</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q11</td>
<td>p=0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q12</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q13</td>
<td>p=0.002</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q14</td>
<td>p=0.002</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q15</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q16</td>
<td>p&lt;0.001</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Q17</td>
<td>p=0.005</td>
<td>Marginally Significant= 0.033&gt;p&gt;0.003; Significant= p&lt;0.003</td>
</tr>
<tr>
<td>Question Number</td>
<td>Associated p-values for Instructor/Course Pair-wise comparisons</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yellow vs. Red</td>
<td>Green vs. Red</td>
</tr>
<tr>
<td></td>
<td>Phys1 vs. Phys3</td>
<td>Phys2 vs. Phys3</td>
</tr>
<tr>
<td>Overall Course Perceptions</td>
<td>Q2 ( p&lt;0.001 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q3 ( p&lt;0.001 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td>Broad Perceptions of Clickers</td>
<td>Q4b ( p=0.9 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q5b ( p=0.9 )</td>
<td>( p=0.002 )</td>
</tr>
<tr>
<td></td>
<td>Q8a ( p&lt;0.001 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q8d ( p=0.001 )</td>
<td>( p=0.9 )</td>
</tr>
<tr>
<td>Emphasis on Sense-making</td>
<td>Q10 ( p=0.03 )</td>
<td>( p=0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q11 ( p&lt;0.001 )</td>
<td>( p=0.002 )</td>
</tr>
<tr>
<td></td>
<td>Q12 ( p&lt;0.001 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q16 ( p&lt;0.001 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td>Faculty-Student Collaboration</td>
<td>Q6a ( p=0.4 )</td>
<td>( p=0.003 )</td>
</tr>
<tr>
<td></td>
<td>Q6b NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Q7 ( p=0.04 )</td>
<td>( p=0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q14 ( p=0.009 )</td>
<td>( p&lt;0.001 )</td>
</tr>
<tr>
<td></td>
<td>Q15 ( p&lt;0.001 )</td>
<td>( p&lt;0.001 )</td>
</tr>
</tbody>
</table>

NS: No statistically significant difference found via Kruskal-Wallis Test
NA: Not Available (Data for one of the semesters, not available for comparison)
Bold type face: p-values below the threshold for marginal statistical significance
Shading: Color of the shading denotes the professor whose students responded more favorably (shading shown when p<0.1).
<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question Variable Name</th>
<th>Question Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8</td>
<td>HELPED_TUTORIALS</td>
<td>Please rate how much each of the following class activities helped your learning in Physics 1110 this semester: Tutorials (the Thursday recitation section) [No Help, A little help, Moderate help, Much help, Very much help]</td>
</tr>
<tr>
<td>Q51</td>
<td>TUTVSTRAD_USEFUL</td>
<td>Which do you think are more useful for you to learn physics: tutorials, or recitation in which a TA shows you how to do problems at the blackboard, and answers questions about homework? [Tutorials are much more useful, Tutorials are a little more useful, It's be about the same, Tutorials are a little less useful, Tutorials are much less useful]</td>
</tr>
<tr>
<td>Q41</td>
<td>ENJOY_TUTORIALS</td>
<td>Please rate how much you enjoyed each of the following class activities in Physics 1110 this semester: Tutorials (the Thursday recitation section) [Strongly Dislike, Dislike, Neutral, Like, Strongly Like]</td>
</tr>
<tr>
<td>Q52</td>
<td>TUTVSTRAD_ENJOY</td>
<td>Which would you enjoy more: tutorials or a recitation (as described in the previous question?) [I'd enjoy Tutorials much more, I'd enjoy Tutorials a little more, It'd be about the same, I'd enjoy Tutorials a little less, I'd enjoy the Tutorials much less]</td>
</tr>
<tr>
<td>Q48</td>
<td>FEEL_TUT</td>
<td>Now that the course is over, how do you feel about the use of Tutorials in Phys 1110? [they're very good, I'm mildly positive about them, I'm completely neutral, I'm mildly negative about them, they're very bad]</td>
</tr>
<tr>
<td>Q50</td>
<td>OFTEN_ATTEND_TUT</td>
<td>How often did you attend Thursday Tutorial? [Perfect attendance!, I only missed one, I've missed 2 or 3 so far, I've missed a lot (4 or more)]</td>
</tr>
<tr>
<td>Q53</td>
<td>TUT_REACTION_CHANGED</td>
<td>Think about how your reaction to the Tutorials has changed over the course of the term: what is your reaction to them now, compared with the start of the term? [I feel more positively now, I feel about the same: positive all along, I feel about the same: neutral all along, I feel about the same: negative all along, I feel more negatively now]</td>
</tr>
<tr>
<td>Q9</td>
<td>HELPED_DISCUSS_TUT</td>
<td>Please rate how much each of the following class activities helped your learning in Physics 1110 this semester: Working/ Discussing with fellow students in Tutorials [No Help, A little help, Moderate help, Much help, Very much help]</td>
</tr>
<tr>
<td>Q42</td>
<td>ENJOY_DISCUSS_TUT</td>
<td>Please rate how much you enjoyed each of the following class activities in Physics 1110 this semester: Working/Discussing with fellow students in Tutorials [Strongly Dislike, Dislike, Neutral, Like, Strongly Like]</td>
</tr>
<tr>
<td>Q23</td>
<td>HELPED_TA</td>
<td>Please rate how much each of the following individual support mechanisms helped your learning in Physics 1110 this semester: Contact with graduate TA's [No Help, A little help, Moderate help, Much help, Very much help]</td>
</tr>
<tr>
<td>Q24</td>
<td>HELPED_LA</td>
<td>Please rate how much each of the following individual support mechanisms helped your learning in Physics 1110 this semester: Contact with Learning Assistants [No Help, A little help, Moderate help, Much help, Very much help]</td>
</tr>
<tr>
<td>Q10</td>
<td>HELPED_EQUIP</td>
<td>Please rate how much each of the following class activities helped your learning in Physics 1110 this semester: Use of hands-on equipment in Tutorials [No Help, A little help, Moderate help, Much help, Very much help]</td>
</tr>
<tr>
<td>Q43</td>
<td>ENJOY_EQUIP</td>
<td>Please rate how much you enjoyed each of the following class activities in Physics 1110 this semester: Use of hands-on equipment in Tutorials [Strongly Dislike, Dislike, Neutral, Like, Strongly Like]</td>
</tr>
<tr>
<td>Table II: Survey 2 Tutorial-Specific Questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HELPED_DISCUSS_TUT</strong></td>
<td>Please rate how much each of the following class activities <em>helped your learning</em> in Physics 1110 this semester: Working/Discussing with fellow students in Tutorials [No Help, A little help, Moderate help, Much help, Very much help]</td>
<td></td>
</tr>
<tr>
<td><strong>ENJOY_DISCUSS_TUT</strong></td>
<td>Please rate how much you <em>enjoyed</em> each of the following class activities in Physics 1110 this semester: Working/Discussing with fellow students in Tutorials [Strongly Dislike, Dislike, Neutral, Like, Strongly Like]</td>
<td></td>
</tr>
<tr>
<td><strong>COMFORT_PEERS_TUT</strong></td>
<td>Please rate how <em>comfortable</em> you felt (or would feel) discussing physics in the following situations: With your peers during Tutorials (the Thursday recitation section) [Very uncomfortable, Somewhat uncomfortable, Neutral, Somewhat comfortable, Very comfortable]</td>
<td></td>
</tr>
<tr>
<td><strong>SHARE_IDEAS_TUT</strong></td>
<td>During physics discussions with my peers in the Tutorials, I usually share my ideas. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>PEERS_VALUE_IDEAS_TUT</strong></td>
<td>In the Tutorials, I feel like my peers value my ideas. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>PEERS_LISTEN_CAREFULLY</strong></td>
<td>My peers listen carefully to my ideas in the Tutorials. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>STUDENTS_DONT_LISTEN</strong></td>
<td>I don't feel like other students listen to me during the Tutorial discussions. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>STUDENTS_DONT_TAKE_SERIOUS</strong></td>
<td>Other students don't take my ideas seriously during the Tutorial discussions. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>HELPED_TA</strong></td>
<td>Please rate how much each of the following individual support mechanisms <em>helped your learning</em> in Physics 1110 this semester: Contact with graduate TA's [No Help, A little help, Moderate help, Much help, Very much help]</td>
<td></td>
</tr>
<tr>
<td><strong>COMFORT_TA_TUT</strong></td>
<td>Please rate how <em>comfortable</em> you felt (or would feel) discussing physics in the following situations: With your TA during Tutorials (the Thursday recitation section) [Very uncomfortable, Somewhat uncomfortable, Neutral, Somewhat comfortable, Very comfortable]</td>
<td></td>
</tr>
<tr>
<td><strong>TA_RIGHT_ANSWER</strong></td>
<td>I can learn in Tutorial even when the TA doesn't give me the right answer. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>TA_ENCOURAGE</strong></td>
<td>My TA encouraged me to further investigate physics during Tutorials. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>QUESTIONS_TA_HELPED</strong></td>
<td>The questions that my TA asked us usually helped me make sense of the physics. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>QUESTIONS_FACILITATE_DISCUSS</strong></td>
<td>My TA's questions facilitated peer discussion in my Tutorial group. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>TA_LISTEN_CAREFULLY</strong></td>
<td>My TA listened carefully to my ideas in the Tutorials. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td>Coordination of Course Components</td>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td><strong>TUT_SUCCEED</strong></td>
<td>The Tutorial activities helped me succeed on the exams. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>EXAMS_RELATED_TUT</strong></td>
<td>The material tested on the exams was related to the Tutorial activities. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>SENSE_LEC</strong></td>
<td>The Tutorial activities helped me make sense of the lectures. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>LECTURE_MAKE_SENSE</strong></td>
<td>The lectures helped me make sense of the Tutorial activities. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>SYNCHRONIZED</strong></td>
<td>The lectures were well synchronized with the Tutorial activities. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>TUT_MAKE_SENSE_CAPA</strong></td>
<td>The Tutorial activities helped me make sense of the CAPA homework assignments. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
<tr>
<td><strong>EXAMS_RELATED_CAPA</strong></td>
<td>The material tested on the exams was closely related to the CAPA homework assignments. [Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree]</td>
<td></td>
</tr>
</tbody>
</table>

†Questions taken from Survey 1 Questions which relate to new Survey 2 Clusters

Table III: Examples of Student Long Answer Responses from Survey 1 that Informed the Design of Survey 2 Questions

<table>
<thead>
<tr>
<th>Themes Identified</th>
<th>Examples of Students’ Long Answer Responses from Survey 1</th>
</tr>
</thead>
</table>
| Student-Student Interactions | • “I’ve found that recitation (if you have a good group of students who are working with you and who are all willing to discuss about the topics) can be really enjoyable and actually fun.”  
  • “I’ve started enjoying doing the tutorials because I’m not afraid of saying something wrong and having someone else bring up another point that I didn't consider.”  
  • “Tutorials are very helpful if you get in a group with people you enjoy and are willing to work.”  
  • “Discussing with fellow students only helps when the students sitting around you are serious and truly looking to learn.”  
  • “In tutorial, I ended up with a group whose understanding was quite a bit ahead of mine. So, they would fly through the tutorial while I still had questions. When I did ask them questions, they couldn't really explain it in a way I could understand well.”  
  • “Talking with other students is a very good way of measuring your own knowledge on the subjects.” |
| Student-TA Interactions | • “This tutorial would be more useful if done with a TA directing the class through the exercise.”  
  • “Everyone was too cryptic when they tried to help.”  
  • “I wish they would have directly explained the material.” |
| Coordination of Course Components | “I never learned anything from Tutorials that helped me out on an exam.”
| | “The format of this class tried to combine the curriculums from three different classes, which made it hard to really gain a concrete understanding.”
| | “I think the lectures and recitation should go over the material that was tested.”
| | “Tutorials would have helped if they followed with the lectures.”
| | “It would have been nice if we did the tutorials that related to what we were learning in class while we were learning it.”
| | “Have it [Tutorials] coincide better with lecture.”
| | “The tutorials weren't useful because they didn't fit with what we were learning in class.”

- “Working with others does not help me learn physics.”
- “I've found that the TA's all are pretty helpful and knowledgeable.”
- “Our TA knew little to nothing about the subject at hand and was more of a hindrance than a help.”
- “I believe that the TA has the potential in his/her role to be very effective for a student's learning process. My TA was not. My LA however was very approachable and always encouraged me to further investigate physics.”
### Table I: Examining Statistical Differences across Four Semesters of Phys1—Survey 1

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Course</th>
<th>Statistical Test</th>
<th>p-value</th>
<th>Number of semesters of data</th>
<th>Pair-wise significance threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8</td>
<td>Phys1</td>
<td>K.W.</td>
<td>p=0.08</td>
<td>3 (3 pairs)</td>
<td>Not Significant (NS)</td>
</tr>
<tr>
<td>Q51</td>
<td>Phys1</td>
<td>X²</td>
<td>NA</td>
<td>1 (No pairs)</td>
<td>NA</td>
</tr>
<tr>
<td>Q41</td>
<td>Phys1</td>
<td>K.W.</td>
<td>p=0.1</td>
<td>3 (3 pairs)</td>
<td>Not Significant (NS)</td>
</tr>
<tr>
<td>Q52</td>
<td>Phys1</td>
<td>X²</td>
<td>NA</td>
<td>1 (No pairs)</td>
<td>NA</td>
</tr>
<tr>
<td>Q48</td>
<td>Phys1</td>
<td>K.W.</td>
<td>p=0.1</td>
<td>3 (3 pairs)</td>
<td>Not Significant (NS)</td>
</tr>
<tr>
<td>Q50</td>
<td>Phys1</td>
<td>K.W.</td>
<td>p=0.040</td>
<td>3 (3 pairs)</td>
<td>p ≤ 0.017</td>
</tr>
<tr>
<td>Q53</td>
<td>Phys1</td>
<td>X²</td>
<td>NA</td>
<td>1 (No pairs)</td>
<td>NA</td>
</tr>
</tbody>
</table>

#### Overall Tutorial Perceptions
- Q9: Phys1, K.W. (p=0.040, 3 (3 pairs), p ≤ 0.017)
- Q42: Phys1, K.W. (p=0.006, 3 (3 pairs), p ≤ 0.017)

#### Working with Peers
- Q23: Phys1, K.W. (p=0.025, 3 (3 pairs), p ≤ 0.017)
- Q24: Phys1, K.W. (p<0.001, 3 (3 pairs), p ≤ 0.017)

#### Working with Educators
- Q10: Phys1, M.W. (p=0.2, 2 (1 pair), p ≤ 0.05)
- Q43: Phys1, M.W. (p=0.004, 2 (1 pair), p ≤ 0.05)

### Table II: Examining Pair-wise Statistical Differences across Four Semesters of Phys1—Survey 1

<table>
<thead>
<tr>
<th>Semesters Comparison</th>
<th>Question Number (statistical test used and associated p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q8 (MWU)</td>
</tr>
<tr>
<td>N&amp;DvsS&amp;V</td>
<td>p=0.04</td>
</tr>
<tr>
<td>N&amp;DvsF&amp;W</td>
<td>p=0.6</td>
</tr>
<tr>
<td>S&amp;VvsF&amp;W</td>
<td>p=0.09</td>
</tr>
<tr>
<td>Question Number</td>
<td>Course</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------</td>
</tr>
<tr>
<td>Q8</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q51</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q41</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q52</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q48</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q50</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q53</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q9</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q42</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q23</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q24</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q10</td>
<td>Phys2</td>
</tr>
<tr>
<td>Q43</td>
<td>Phys2</td>
</tr>
</tbody>
</table>
Table IV: Examining Pair-wise Statistical Differences across Seven Semesters of Phys2—Survey 1

<table>
<thead>
<tr>
<th>Semesters Comparison</th>
<th>Q8 (MWU)</th>
<th>Q41 (MWU)</th>
<th>Q48 (MWU)</th>
<th>Q50 (MWU)</th>
<th>Q53 (X²)</th>
<th>Q9 (MWU)</th>
<th>Q42 (MWU)</th>
<th>Q23 (MWU)</th>
<th>Q24 (MWU)</th>
<th>Q10 (MWU)</th>
<th>Q43 (MWU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&amp;F vs F&amp;H</td>
<td>p&lt;0.001</td>
<td>p=0.1</td>
<td>p=0.001</td>
<td>NA</td>
<td>NA</td>
<td>p&lt;0.001</td>
<td>p=0.001</td>
<td>p=0.4</td>
<td>p=0.2</td>
<td>p&lt;0.001</td>
<td>p=0.01</td>
</tr>
<tr>
<td>A&amp;F vs H</td>
<td>p&lt;0.001</td>
<td>p=0.004</td>
<td>p&lt;0.001</td>
<td>NA</td>
<td>NA</td>
<td>p&lt;0.001</td>
<td>p=0.006</td>
<td>p=0.2</td>
<td>p=0.05</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>A&amp;F vs L&amp;M</td>
<td>p=0.03</td>
<td>p=0.6</td>
<td>p=0.06</td>
<td>NA</td>
<td>NA</td>
<td>p&lt;0.001</td>
<td>p=0.04</td>
<td>p=0.4</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.01</td>
</tr>
<tr>
<td>A&amp;F vs H &amp; A</td>
<td>p=0.3</td>
<td>p=0.02</td>
<td>p=0.6</td>
<td>NA</td>
<td>NA</td>
<td>p=0.2</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.01</td>
<td>p=0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>A&amp;F vs R&amp;M</td>
<td>p=0.01</td>
<td>p=0.001</td>
<td>p=0.5</td>
<td>NA</td>
<td>NA</td>
<td>p=0.03</td>
<td>p=0.002</td>
<td>p=0.004</td>
<td>p=0.01</td>
<td>p=0.001</td>
<td>p=0.002</td>
</tr>
<tr>
<td>A&amp;F vs H &amp; N</td>
<td>p&lt;0.001</td>
<td>p=0.3</td>
<td>p=0.004</td>
<td>NA</td>
<td>NA</td>
<td>p&lt;0.001</td>
<td>p=0.1</td>
<td>p=0.08</td>
<td>p=0.8</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>F &amp; H vs H</td>
<td>p=0.2</td>
<td>p=0.2</td>
<td>p=0.3</td>
<td>p&lt;0.03</td>
<td>p=0.3</td>
<td>p=0.06</td>
<td>p=0.05</td>
<td>p=0.6</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>F &amp; H vs F &amp; M</td>
<td>p=0.1</td>
<td>p=0.1</td>
<td>p=0.3</td>
<td>p=0.03</td>
<td>p=0.02</td>
<td>p=0.2</td>
<td>p=0.7</td>
<td>p=0.4</td>
<td>p=0.8</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>F &amp; H vs H &amp; A</td>
<td>p&lt;0.001</td>
<td>p=0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.001</td>
<td>p=0.001</td>
<td>p=0.002</td>
<td>p=0.9</td>
<td>p=0.2</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>F &amp; H vs H &amp; R &amp; M</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.01</td>
<td>p=0.003</td>
<td>p=0.6</td>
<td>p&lt;0.001</td>
<td>p=0.01</td>
</tr>
<tr>
<td>F &amp; H vs H &amp; N</td>
<td>p=0.8</td>
<td>p=0.6</td>
<td>p=0.5</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>p=0.06</td>
<td>p=0.02</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.1</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>H vs L &amp; M</td>
<td>p=0.009</td>
<td>p=0.006</td>
<td>p=0.03</td>
<td>p=0.09</td>
<td>p=0.008</td>
<td>p=0.06</td>
<td>p=0.003</td>
<td>p=0.4</td>
<td>p=0.007</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>H vs H &amp; A</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.005</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.6</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>H vs H &amp; R &amp; M</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.03</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.001</td>
<td>p=0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>H vs H &amp; N</td>
<td>p=0.1</td>
<td>p=0.09</td>
<td>p=0.08</td>
<td>p=0.7</td>
<td>NA</td>
<td>p=0.9</td>
<td>p=0.2</td>
<td>p=0.8</td>
<td>p=0.04</td>
<td>p=0.9</td>
<td>p=0.006</td>
</tr>
<tr>
<td>L &amp; M vs H &amp; A</td>
<td>p=0.003</td>
<td>p=0.02</td>
<td>p=0.01</td>
<td>p=0.018</td>
<td>p&lt;0.010</td>
<td>p&lt;0.018</td>
<td>p=0.09</td>
<td>p=0.2</td>
<td>p=0.09</td>
<td>p=0.7</td>
<td>p=0.4</td>
</tr>
<tr>
<td>L &amp; M vs H &amp; R &amp; M</td>
<td>p=0.09</td>
<td>p=0.05</td>
<td>p=0.02</td>
<td>p&lt;0.001</td>
<td>NA</td>
<td>p=0.4</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.05</td>
<td>p=0.05</td>
<td>p=0.9</td>
</tr>
<tr>
<td>L &amp; M vs H &amp; N</td>
<td>p=0.2</td>
<td>p=0.2</td>
<td>p=0.5</td>
<td>p=0.04</td>
<td>NA</td>
<td>p&lt;0.007</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.009</td>
<td>p=0.2</td>
</tr>
<tr>
<td>H &amp; M vs H &amp; R &amp; M</td>
<td>p=0.2</td>
<td>p=0.4</td>
<td>p=0.8</td>
<td>p=0.5</td>
<td>NA</td>
<td>p=0.3</td>
<td>p=0.1</td>
<td>p=0.001</td>
<td>p&lt;0.009</td>
<td>p=0.001</td>
<td>p=0.3</td>
</tr>
<tr>
<td>H &amp; M vs H &amp; N</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.02</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.1</td>
<td>p&lt;0.001</td>
<td>p=0.8</td>
</tr>
<tr>
<td>R &amp; M vs H &amp; N</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p=0.3</td>
<td>p=0.3</td>
<td>p=0.4</td>
</tr>
</tbody>
</table>
## Appendix C: Summaries of Variations in Student Background Factors

### Table I: Summary of Semester-to-semester Variations in Student Background Factors in Phys1

<table>
<thead>
<tr>
<th>Variable</th>
<th>N&amp;D (Fa2006)</th>
<th>S&amp;V (Fa2008)</th>
<th>F&amp;W (Sp2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage by Gender</td>
<td>no significant differences</td>
<td>no significant differences</td>
<td>no significant differences</td>
</tr>
<tr>
<td>Percentage by Ethnicity</td>
<td>no significant differences</td>
<td>no significant differences</td>
<td>no significant differences</td>
</tr>
<tr>
<td>Percentage by Major</td>
<td>significantly different from 20091</td>
<td>significantly different from 20091</td>
<td>significantly different from 20067 &amp; 20087</td>
</tr>
<tr>
<td>HS Physics</td>
<td>significantly different (Higher) than 20091</td>
<td>no significant differences</td>
<td>significantly different (Lower) than 20067</td>
</tr>
<tr>
<td>FMCE Pretest</td>
<td>no significant differences</td>
<td>no significant differences</td>
<td>no significant differences</td>
</tr>
<tr>
<td>CLASS Overall Favorable Pretest</td>
<td>significantly different (Higher) than 20087 &amp; 20091</td>
<td>significantly different (Lower) than 20067</td>
<td>significantly different (Lower) than 20067</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Percentage by Gender</td>
<td>no significant differences</td>
<td>no significant differences</td>
<td>no significant differences</td>
</tr>
<tr>
<td>Percentage by Ethnicity</td>
<td>small variations &lt;10%</td>
<td>small variations &lt;10%</td>
<td>small variations &lt;10%</td>
</tr>
<tr>
<td>Percentage by Major</td>
<td>only fall/spring differences</td>
<td>only fall/spring differences</td>
<td>only fall/spring differences</td>
</tr>
<tr>
<td>HS Physics</td>
<td>no significant differences</td>
<td>no significant differences</td>
<td>no significant differences</td>
</tr>
<tr>
<td>BEMA Pretest</td>
<td>no significant differences</td>
<td>no significant differences</td>
<td>no significant differences</td>
</tr>
<tr>
<td>FMCE Posttest</td>
<td>Significantly different (Higher) than 20057, 20061, 20067, &amp; 20087</td>
<td>Significantly different (Higher) than 20057, 20061, 20067, &amp; 20087</td>
<td>Significantly different (Lower) than 20047 &amp; 20051</td>
</tr>
<tr>
<td>CLASS Overall Favorable Pretest</td>
<td>Significantly different (Higher) than 20057, 20061, 20067, &amp; 20087</td>
<td>no significant differences</td>
<td>Significantly different (Lower) than 20047</td>
</tr>
</tbody>
</table>