Not all interactive engagement is the same: Variations in physics professors' implementation of *Peer Instruction*

Chandra Turpen and Noah D. Finkelstein

Department of Physics, University of Colorado at Boulder, Boulder, Colorado 80309, USA

While educational reforms in introductory physics are becoming more widespread, how these reforms are implemented is less well understood. This paper examines the variation in faculty practices surrounding the implementation of educational reform in introductory physics courses. Through observations of classroom practice, we find that professors' actual practices differ strikingly. We present a framework for describing and capturing instructional choices and resulting variations in enacted practices for faculty who are implementing *Peer Instruction*. Based on our observations, there are a variety of scientific practices that are supported and modeled in 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. Case studies of six professors demonstrate how these variations in classroom practices, in aggregate, create different classroom norms, such as the relative emphasis on student sense-making vs answer-making during *Peer Instruction*.

I. INTRODUCTION

Consider two representative vignettes of interactive engagement in a large-enrollment introductory physics course:

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 their answers. 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] We note that 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 implemen-

tations. 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. The present paper begins to address this question by developing a system for describing and measuring classroom practices that contribute to the construction of different classroom norms,[3–7] 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 variation in teacher practice results in disparate opportunities for students to practice conceptual reasoning,[1,8,9] skills at talking physics,[10,11] agency,[12–14] 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.[17–19] 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 cen-

tered 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] This research has also noted that these organizational aspects of curricular change are usually not evaluated and left implicit in discussions of the curricular design, if addressed at all.[20] This paper provides a metric for documenting and discussing these organizational changes associated with the use of Peer Instruction. In this paper, 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, either: (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 University of Colorado (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. 27)] 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, and calls 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.[29–33] 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 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[29–31] 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, we investigate 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. This work compares six university faculty members implementing *Peer Instruction* in six different introductory physics courses, and documents 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 13 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). Subsequent work 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 min 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:[1,34]

General format of ConcepTest
(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 11 separate higher education institutions including 30 introductory physics courses in which *Peer Instruction* has been implemented.[35] According to this study, professors that 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 paper, 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. The stages we use are: 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.[36–38] We seek to characterize faculty use of Peer Instruction which requires a slight broadening of the definition of Peer Instruction 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: observable characteristics of practice and dimensions of practice (DoP). 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 class-

room 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,[42–45] 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,[29–33,46] descriptive studies of classroom response system implementation,[1,34,35,47–50] and broad educational research results.[51–55] 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[56–58] 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, Part 1 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 certain aspects of the classroom norms, obligations, and expectations about the participant roles and the rules of interaction[3–7] 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;[59–63] however, this set was beyond the scope of this study. The dimensions that we have identified are summarized in Table I. The dimensions described below are not completely independent, but

TABLE I. Summary of dimensions of practice.

Dimensions of Practice

Defining the Academic Task

(1) The Role of Clicker Use within the Organization of the Course: *Do I integrate or coordinate the use of clickers within broader course activities and course evaluation? If so how?*

(2) Types of Clicker Questions: *What do I want students to learn and what do I want to learn from the students?*

(3) Introduction of the Clicker Question: *As I pose the question, how do I convey to the students what this question is about and what the students should be doing?*

(4) Student-Student Collaboration: *Do I allow, support, or encourage student discussion during the CQ?*

(5) Determining Time Constraints: *Given the nature of this question and what I expect students to be doing, how long should this take?*

Student-Professor Interactions

(6) Creating or Reducing Spatial Boundaries between Students and Instructor: *Should I walk around the room?*

(7) Listening to Student Explanations and Comments: *Do I need to listen to students' ideas and reasoning and what are the benefits of listening to students' ideas?*

(8) Faculty-Student Collaboration: *What kinds of interactions should I have with the students during the CQ?*

(9) Instructor's Use of Student Prior Knowledge: *Do I build on students' prior knowledge and in what ways?*

(10) Use of student voice: *Should students voice their understanding and reasoning during class? If so when and how should this happen?*

(11) Management of disagreement among students: *How do I respond to the CQ results when there is a split response among the students?*

(12) Formative Use of Students' Ideas and Aggregate Student Responses: *Is the information that I am gathering useful in determining what happens next? If so how is this information useful?*

(13) Summative Use of Aggregate Student Responses: *Do I want to know where students got to? Where did students get to?*

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.

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 nonscience 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 educa-

tion 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 A 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 paper should be loosely interpreted.

Ethnographic research methods[56–58] 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 B, Part 2 for the observation rubric and Appendix B, Part 3 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 their students do during a typical CQ, purposes of clicker use, and the importance of various student activities associated with CQs such as articulating their 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 13 DoPs, presented in Table I, 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. 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

TABLE II. Average number of clicker questions per hour of class.

	Yellow	Green	White	Blue	Purple	Red
$\langle CQ \rangle/h$	5.9 ± 0.5[a]	3.2 ± 0.1	8.2 ± 0.6	6.5 ± 0.4	5.0 ± 0.3	5.3 ± 0.4
No. of classes	38	40	31	38	27	35
No. of CQs	187	108	213	211	170	155

[a]In all tables the error stated is the standard error on the mean, unless otherwise stated.

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.

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 II. 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.

Similarly, we were able to calculate the average fraction of students getting the CQ correct. These data are summarized in Table III. 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. This means for courses such as Blue and Red's 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 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.

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, Part 1, logistical questions were questions used to poll students' perspectives or opinions about the course. 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, was occasionally using CQs to give graded reading quizzes.

Many educators and researchers have proposed various schemes for categorizing types of content-based CQs.[1,72–76] 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, Part 1 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 IV. 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

TABLE III. Percent of students getting the clicker question correct on average.

	Yellow	Green	White	Blue	Purple	Red
Percent correct	$69 \pm 2\%$	$68 \pm 2\%$	$64 \pm 3\%$	$76 \pm 1\%$	$70 \pm 2\%$	$72 \pm 4\%$
Median Percent correct	73%	72%	65%	82%	75%	80%
No. of CQs	197	117	56[a]	190	146	29[a]

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

TABLE IV. Fraction of Clicker Questions that were: Logistical, Recall, Algorithmic, or Conceptual

	Fraction of Logistical CQs	Fraction of Recall CQs	Fraction of Algorithmic CQs	Fraction of Conceptual CQs
Yellow ($N=35$)	0.06	0.20	0.06	0.69
Green ($N=27$)	0.00	0.04	0.11	0.85
White ($N=60$)	0.02	0.15	0.08	0.75
Blue ($N=31$)	0.10	0.06	0.06	0.77
Purple ($N=34$)	0.06	0.09	0.06	0.79
Red ($N=33$)	0.12	0.24	0.00	0.64

asking a relatively higher fraction of recall questions compared to the other professors. It is interesting to note however that half of the recall questions (2 out of 4) asked by Red were reading quizzes.

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's classroom was the only classroom which occasionally asked questions that were intended to be answered individually and these 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 B.

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. We also note that 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 Fig. 1 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 either instructor or student) varied, as shown in columns two and three of Fig. 1, respectively.

From Fig. 1, 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, they chose to have very limited interactions
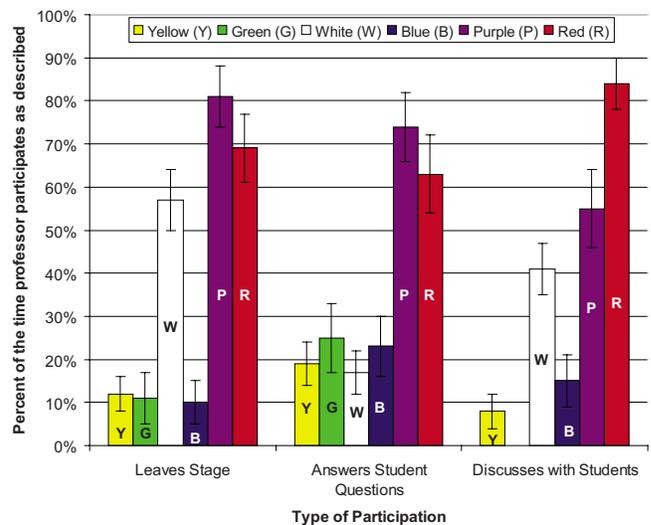

FIG. 1. (Color) The percentage of observed CQs where the professor was observed to participate with the students during the response time by leaving the stage (column 1), answering student questions (column 2), or actively discussing with the students (column 3). The error bars shown are the standard error on the proportion.
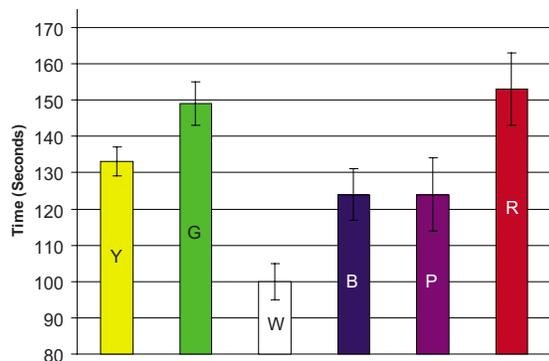
FIG. 2. (Color) Average time given for students to respond (seconds). The error bars shown are the standard error on the mean.

FIG. 3. (Color) Histogram of time elapsed (in minutes) for students to respond to CQs for the classrooms of Professor Red and Professor Yellow.

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, but also 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 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. We also note that 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 Fig. 2. 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.
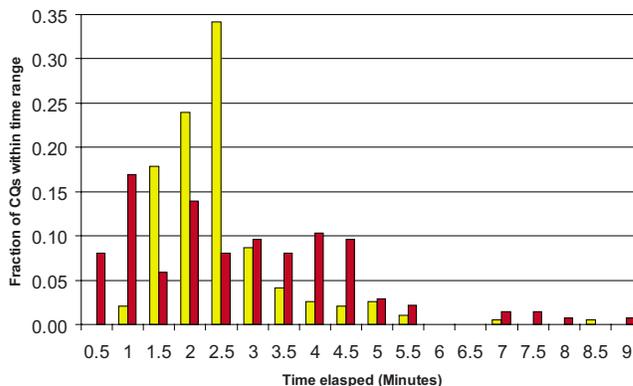
By looking at the distribution of clicker timing data, we can begin to describe further differences between these professors' practices. The time elapsed for each question, as captured by the clicker software, is binned into 30 s intervals and shown in Fig. 3 for two professors, Red and Yellow. These 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. These data from this stage are summarized in Appendix A, Table C.

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.

TABLE V. Average time spent discussing the CQ solution (minutes:seconds, N=number of questions)

	Average time discussing solution
Yellow	$3:04 \pm 0:18$ ($N=25$)
Green	$3:16 \pm 0:30$ ($N=22$)
White	$1:10 \pm 0:10$ ($N=29$)
Blue	$2:28 \pm 0:18$ ($N=25$)
Purple	$3:26 \pm 0:35$ ($N=28$)
Red	$3:42 \pm 0:26$ ($N=25$)

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 V.

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 Fig. 4 shows the fraction of observed CQs where incorrect answers were discussed during the description of the solution. The second column of data in Fig. 4 shows the

fraction of the observed CQs where student explanation(s) of the CQ solution were heard during the whole class solution discussion. From Fig. 4 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 Fig. 4 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 students 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 D.

We find that faculty-student collaboration (DoP 8) is even more varied now that we have examined the CQ solution discussion stage. For example in Green's class students' are always contributing their own descriptions of the CQ solution, while in Blue's class student explanations of the solutions are heard only 10% of the time. Similarly, we see that the use of student voice (DoP 10) varies from course to course. Students in these courses are given different opportunities to practice communicating in public. Even when student explanations are heard, it varies how many different students are heard. We also see in Purple, Green, and Red's class average the most number of students sharing ($\sim$4.5) while in Blue and White's classrooms less than a single student contributes a solution publicly in each class period. Students are given different opportunities to practice identifying
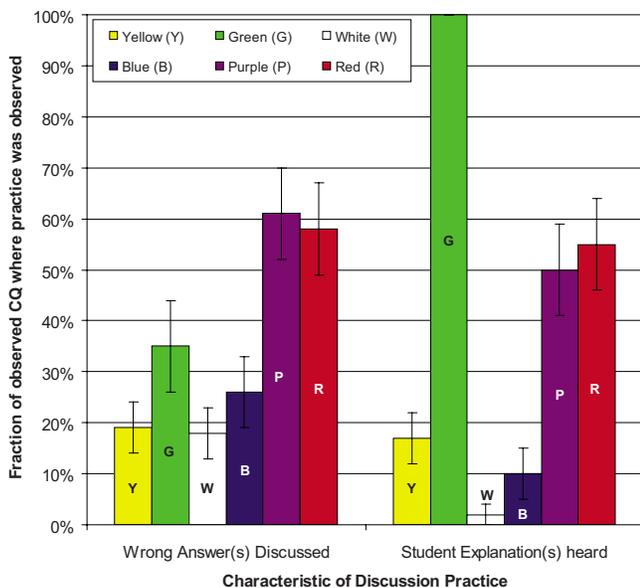

FIG. 4. (Color) The percentage of observed CQs where the wrong answers were discussed and student explanations were heard. The error bars shown are the standard error on the proportion.

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.[78–80] 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 then the educator gives evaluation (E) or feedback (F) based on the responses. We note that 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, it is worth noting 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 II). All of the professors in this study asked conceptual questions the majority of the time (DoP. 2; Table IV). 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. We also note that 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; Fig. 1 and 4). 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, Figs. 1 and 4). 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 student-student interactions, but not new opportunities for students-professor collaboration. 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, Fig. 4). 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 (Fig. 4). 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, Fig. 1), sometimes answering student questions, and occasionally discussing with the students (DoP. 7 and 8, Fig. 1). However these interactions with students were usually brief (DoP. 8 and 10, Fig. 2). The professor almost never requested student explanations during the solution discussion (DoP. 7 and 8, Fig. 4). 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 III, Fig. 2, and Table V). 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 that, "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, Fig. 1). 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, Fig. 4). 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, Fig. 1). Professor Purple usually asked students to contribute explanations of the CQ solution (DoP. 7 and 8, Fig. 4). 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 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 value for our students. Our studies demonstrate the potential for Peer Instruction to support the development of these scientific prac-
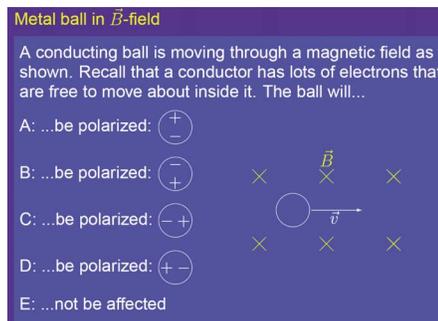

FIG. 5. (Color) Screen shot of a conceptual CQ in Green's classroom (correct answer: A).

tices; however, it 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 26 min into the 50 min class period. Prior to this CQ the professor had briefly discussed the domain model of magnetism and described permanent magnets. A student asks a question about particular materials and their magnetic properties which the professor addressed. The professor then said, "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 Fig. 5) 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 and they are affected by a magnetic field that is going into the board so the force on the positive charges would be up, so they would move up. But for the negative charges in the ball their velocity would be negative so there the force would be pointing down on the negative charges… so those forces would force the positive charges to the top of the ball and the negative charges to the bottom of the ball." The professor responds, "Okay, Joe says that the positive charges in the ball 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 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 is an example of electromagnetic induction using examples from demonstration equipment.
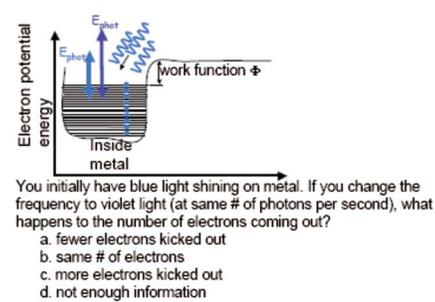

FIG. 6. (Color) Screen shot of a conceptual CQ in Red's classroom (correct answer: C).

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 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 Fig. 6). 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.

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 was not obvious if he had 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 was 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 discussed 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 had 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). Student 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, Fig. 3). However, the professors participate in the CQ response time differently (DoP. 6 and 8, Fig. 1). 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, Fig. 1). Red inquires and listens to what the students are thinking during this voting time (DoP.7 and 8, Fig. 1). 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 are varied depending 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, Fig. 4). 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 D). 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 D). In this example, the student explanations build on fellow students' answers (DoP. 11). Furthermore, each student contribution includes reasoning for their 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 that 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, this reduces the importance of reasoning and sense making in this class. 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 this 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 stu-

dents 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. 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. 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,[12–14] 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. In future work, we investigate how students from these classrooms perceive the norms of the classroom and the use of PI differently.

ACKNOWLEDGMENTS

APPENDIX A: REFERENCE TABLES-INSTRUCTOR ATTRIBUTES AND CHARACTERISTICS OF INSTRUCTOR PRACTICE BY STAGE

TABLE A. Course and professor attributes.

	Course	Primary student major	Experienced or novice clicker user	Temporary instructor or faculty	Phys. Ed. Researcher (Y/N)	Comments
Yellow	calc-based Phys 1	Science	Experienced	Faculty	N	Previously mentored by Blue
Green	calc-based Phys 2	Science	Novice	Instructor	N	First time teaching a large-enrollment course
White	alg-based Phys 1	Science	Novice	Instructor	N	First time teaching a large-enrollment course
Blue	alg-based Phys 2	Science	Experienced	Faculty	Y	Nationally recognized for teaching excellence
Purple	physics elective	Nonscience	Experienced	Faculty	Y	None
Red	calc-based Phys 3	Science	Experienced	Faculty	Y	None

TABLE B. Characteristics of Practice-CQ Set Up Stage.

	Yellow	Green	White	Blue	Purple	Red
Clickers part of course grade	No	No	Yes (5%)	Yes (1%)	Yes (5%)	Yes (15%)
CQs points (for grade) based on correctness	NA	NA	No	Yes (3:1)	No	Rarely
Extra credit (EC) points from clickers	Yes	Yes	No	No	Yes	No
EC points based on correctness	Yes (2:1)	Yes (3:1)	NA	NA	Yes (2:1)	NA
CQs with answers available	Yes	Yes	Yes	Yes	Yes	Yes
CQs answer explanations provided	No	Yes	Yes	Yes	Yes	Yes
$<CQ>$/hr	5.9 ± 0.5	3.2 ± 0.1	8.2 ± 0.6	6.5 ± 0.4	5.0 ± 0.3	5.3 ± 0.4
% of students correct	$69\pm2\%$	$68\pm2\%$	$64\pm3\%$	$76\pm1\%$	$70\pm2\%$	$72\pm4\%$
Median % of students correct	73%	72%	65%	82%	75%	80%
Fraction of logistical CQs	0.06	0.00	0.02	0.10	0.06	0.12
Fraction of recall CQs	0.20	0.04	0.15	0.06	0.09	0.24
Fraction of algorithmic CQs	0.06	0.11	0.08	0.06	0.06	0.00
Fraction of conceptual CQs	0.69	0.85	0.75	0.77	0.79	0.64

TABLE C. Characteristics of Practice-CQ Response Stage.

	Yellow	Green	White	Blue	Purple	Red
Percent of CQs where the professor left the stage	$12\pm4\%$	$11\pm6\%$	$57\pm7\%$	$10\pm5\%$	$81\pm7\%$	$69\pm8\%$
Percent of CQs where the professor answered a student question	$19\pm5\%$	$25\pm8\%$	$17\pm5\%$	$23\pm7\%$	$74\pm8\%$	$63\pm9\%$
Percent of CQs where the professor discussed with the students	$8\pm4\%$	0%	$41\pm6\%$	$15\pm6\%$	$55\pm9\%$	$84\pm6\%$
Average time given to students to answer the CQ (seconds)	133 ± 4	149 ± 6	100 ± 5	124 ± 7	124 ± 10	153 ± 10
Standard deviation in avg. time given to respond (seconds)	59	61	69	89	103	116

TABLE D. Characteristics of Practice-CQ Solution Discussion Stage.

	Yellow	Green	White	Blue	Purple	Red
Average time discussing solution (minutes:seconds)	$3:04\pm0:18$	$3:16\pm0:30$	$1:10\pm0:10$	$2:28\pm0:18$	$3:26\pm0:35$	$3:42\pm0:26$
Percent of CQs where incorrect answers were discussed	$19\pm5\%$	$35\pm9\%$	$18\pm5\%$	$26\pm7\%$	$61\pm9\%$	$58\pm9\%$
Percent of CQs where student explanations were heard	$17\pm5\%$	100%	$2\pm2\%$	$10\pm5\%$	$50\pm9\%$	$55\pm9\%$
When student explanations are requested, average number of student explanations heard per CQ	2.2 ± 0.2	1.4 ± 0.1	NA	1.3 ± 0.3	2.3 ± 0.5	2.4 ± 0.4
Average number of student explanations heard per hour of class	2.4 ± 0.6	4.6 ± 0.6	0.1 ± 0.1	0.6 ± 0.4	4.2 ± 0.5	4.8 ± 1.3

APPENDIX B: MORE DETAILED DESCRIPTIONS

See separate auxiliary material for more detailed descriptions of the DoPs (Part 1), the observation rubric (Part 2), and a user's guide to accompany this rubric (Part 3).

---

[1] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).

[2] M. H. Dancy and C. Henderson, "Preliminary Results From A Survey Of Physics Faculty On Use Of Research-Based Instructional Strategies." Contributed Poster at the Physics Teacher Education Coalition Conference, Pittsburgh, PA, March 13–14, 2009. Available at: http://homepages.wmich.edu/~chenders/Publications/DancyPTEC2009Poster.pdf

[3] M. Cole, *Cultural Psychology: A Once and Future Discipline* (Harvard University Press, Cambridge, MA, 1996).

[4] Y. Engeström, in *Understanding Practice: Perspectives on Activity and Context*, edited by S. Chaiklin and J. Lave (Cambridge University Press, New York, NY, 1993).

[5] *Perspectives on Activity Theory*, edited by Y. Engeström, R. Miettinen, and R. Punamäki (Cambridge University Press, New York, NY, 1999).

[6] P. Cobb, M. Stephan, K. McClain, and K. Gravemeijer, Participating in Classroom Mathematical Practices, J. Learn. Sci. **10**(1&2), 113 (2001).

[7] N. Finkelstein, Learning Physics in Context: A study of student learning about electricity and magnetism, Int. J. Sci. Educ. **27**, 1187 (2005).

[8] A. B. Arons, *Teaching Introductory Physics* (John Wiley and Sons, Inc., New York, NY, 1997).

[9] M. L. Rosenquist and L. C. McDermott, A Conceptual Approach to Teaching Kinematics, Am. J. Phys. **55**(5), 407 (1987).

[10] J. L. Lemke, *Talking Science* (Ablex Publishing Company, Norwood, NJ, 1990).

[11] W. M. Roth, *Talking Science: Language and Learning in Science Classrooms* (Rowman and Littlefield Publishers, Lanham, MD, 2005).

[12] A. C. Barton, J. L. Ermer, T. A. Burkett, and M. D. Osborne, *Teaching Science for Social Justice* (Teachers College Press, New York, NY, 2003).

[13] P. Freire, *Pedagogy of the Oppressed* (Continuum International Publishing Group, New York, NY, 2000).

[14] W. M. Roth and A. C. Barton, *Rethinking Scientific Literacy* (Routledge, New York, NY, 2004).

[15] *Inquiry and the National Science Standards: A Guide for Teaching and Learning* (The National Academies Press, Washington, DC, 2000).

[16] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).

[17] H. Borko and R. T. Putnam, in *Handbook of Educational Psychology*, edited by D. Berliner and R. Calfee (Macmillan, New York, NY, 1996), pp. 673–708.

[18] J. H. van Driel, N. Verloop, and W. de Vos, Developing Science Teachers' Pedagogical Content Knowledge, J. Res. Sci. Teach. **35**(6), 673 (1998).

[19] B. Y. White and J. R. Frederiksen, Inquiry, Modeling, and Meta-cognition: Making Science Accessible to All Students, Cogn. Instruct. **16**(1), 3 (1998).

[20] M. Fullan and A. Pomfret, Research on curriculum and instruction implementation, Rev. Educ. Res. **47**(2), 335 (1977).

[21] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**(1), 64 (1998).

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

[23] P. W. Laws, *Workshop Physics Activity Guide Modules 1–4*, 2nd ed. (John Wiley and Sons, New York, 2004); http://physics.dickinson.edu/~wp_web/wp_overview.html

[24] J. M. Saul and E. F. Redish, "Final Evaluation Report for FIPSE Grant #P116P50026: Evaluation of the Workshop Physics Dissemination Project." Physics Education Research Group, University of Maryland, 1997 (unpublished).

[25] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, Phys. Teach. **30**, 141 (1992); Available at http://modeling.asu.edu/R&E/Research.html.

[26] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 2002).

[27] S. J. Pollock and N. D. Finkelstein, Sustaining educational reforms in introductory physics, Phys. Rev. ST Phys. Educ. Res. **4**, 010110 (2008).

[28] R. Thornton and D. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation, Am. J. Phys. **66**(4), 338 (1998).

[29] C. Henderson and M. Dancy, Physics Faculty and Educational Researchers: Divergent Expectations as Barriers to the Diffusion of Innovations, Am. J. Phys. **76**(1), 79 (2008).

[30] M. Dancy and C. Henderson, Framework for articulating instructional practices and conceptions, Phys. Rev. ST Phys. Educ. Res. **3**, 010103 (2007).

[31] C. Henderson and M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, Phys. Rev. ST Phys. Educ. Res. **3**, 020102 (2007).

[32] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, Phys. Rev. ST Phys. Educ. Res. **3**, 020109 (2007).

[33] C. Henderson, E. Yerushalmi, V. H. Kuo, K. Heller, and P. Heller, Physics faculty beliefs and values about the teaching and learning of problem solving. II. Procedures for measurement and analysis, Phys. Rev. ST Phys. Educ. Res. **3**, 020110 (2007).

[34] C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur, Peer Instruction: Engaging Students One-on-One, All At Once, Re-

views in Physics Education Research 1, 1 (2007).

[35] A. P. Fagen, C. H. Crouch, and E. Mazur, Peer Instruction: Results from a Range of Classrooms, Phys. Teach. **40**, 206 (2002).

[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 card, or simple hand-raising…) [See Refs. 37,38]. We do not, however, debate the different affordances of these technologies here.

[37] N. Lasry, Clickers or flashcards: Is there really a difference? Phys. Teach. **46**, 242 (2008).

[38] R. E. Mayer, A. Stull, K. DeLeeuw, K. Almeroth, B. Bimber, D. Chun, M. Bulger, J. Campbell, A. Knight, and H. Zhang, Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes, Contemp. Educ. Psychol. **34**, 51 (2009).

[39] L. S. Vygotsky, *Mind in Society: The development of higher psychological processes* (Harvard University Press, Cambridge, MA, 1978).

[40] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**(3), 212 (1998).

[41] E. F. Redish, *Teaching Physics* (John Wiley and Sons, Danvers, MA, 2003), pp. 51–68.

[42] W. K. Adams, K. K. Perkins, N. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, A new instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **2**, 010101 (2006).

[43] K. K. Perkins, M. M. Gratny, W. K. Adams, N. D. Finkelstein, and C. E. Wieman, *PERC Proceedings 2005* (AIP Press, Melville NY, 2006).

[44] W. K. Adams, K. K. Perkins, M. Dubson, N. D. Finkelstein, and C. E. Wieman, *PERC Proceedings 2004* (AIP Press, Melville NY, 2005).

[45] K. K. Perkins, W. K. Adams, N. D. Finkelstein, S. J. Pollock, and C. E. Wieman, *PERC Proceedings 2004* (AIP Press, Melville NY, 2005).

[46] M. F. Pajares, Teachers' Beliefs and Educational Research: Cleaning up a Messy Construct, Rev. Educ. Res. **62**(3), 307 (1992).

[47] C. Fies and J. Marshall, Classroom Response Systems: A Review of the Literature, J. Sci. Educ. Technol. **15**(1), 101 (2006).

[48] W. R. Penuel, J. Roschelle, and L. Abrahamson, Proceedings of the 2005 IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'05)(unpublished). Available at: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1579270&isnumber=33360

[49] J. Roschelle, W. R. Penuel, and L. Abrahamson, Annual Meeting of the American Educational Research Association. San Diego, CA, April 2004 (unpublished). Available at: http://ubiqcomputing.org/CATAALYST_AERA_Proposal.pdf

[50] D. E. Meltzer and K. Manivannan, Transforming the lecture-hall environment: The fully interactive physics lecture, Am. J. Phys. **70**(6), 639 (2002).

[51] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn* (National Academy Press, Washington, DC, 2000).

[52] National Research Council, *National Science Education Standards* (National Academy Press, Washington, DC, 1996).

[53] American Association for the Advancement of Science, *Project 2061, Benchmarks for Scientific Literacy* (Oxford University Press, New York, NY, 1993).

[54] *The Cambridge Handbook of the Learning Sciences*, edited by R.

K. Sawyer (Cambridge University Press, New York, NY, 2006).

[55] D. Sawada, M. D. Piburn, E. Judson, J. Turley *et al.*, Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol, Sch. Sci. Math. **102**(6), 245 (2002).

[56] J. Creswell, *Educational Research* (Pearson Prentice Hall, Upper Saddle River, NJ 2005).

[57] J. Maxwell, *Qualitative Research Design* (Sage Publications, London, 2005).

[58] F. Erickson, *Handbook of Research on Teaching*, edited by M. C. Wittrock, 3rd ed. (Macmillan, New York, 1986) pp.119–161.

[59] A. Hoekstra, Vibrant student voices: exploring effects of the use of clickers in large college courses, Learning Media & Technology **33**(4), 329 (2008).

[60] M. K. Smith, W. B. Wood, W. K. Adams, C. Wieman, J. K. Knight, N. Gulid, and T. T. Stu, Why peer discussion improves student performance on in-class concept questions, Science **323**, 122 (2009).

[61] M. C. James, The effect of grading incentive on student discourse in peer instruction, Am. J. Phys. **74**(8), 689 (2006).

[62] M. C. James, F. Barbieri, and P. Garcia, What are they talking about? Lessons learned from a study of peer instruction, Astron. Educ. Rev. **7**(1), 37 (2008).

[63] S. D. Willoughby and E. Gustafson, Technology talks: Clickers and grading incentive in the large lecture hall, Am. J. Phys. **77**(2), 180 (2009).

[64] B. Rogoff, *The Cultural Nature of Human Development* (Oxford University Press, New York, NY, 2003), pp. 3–36.

[65] J. Spradley, *Participant observation* (Holt, Rinehart and Winston, New York, NY, 1980).

[66] C. Geertz, in *The Interpretation of Cultures: Selected Essays*, edited by C. Geertz (Basic Books, New York, 1973), pp. 3–30.

[67] R. Burgess, in *In The Field: An Introduction To Field Research*, edited by R. Burgess (Allen & Unwin, London, 1984), pp. 101–122.

[68] S. Mathieson, Why triangulate? Educ. Res. **17**(2), 13 (1988).

[69] G. Novak, A. Gavrin, W. Christian, and E. Patterson, *Just-in-Time Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Upper Saddle River, NJ, 1999).

[70] D. R. Sokoloff and R. K. Thornton, Using interactive lecture demonstrations to create an active learning environment, Phys. Teach. **35**, 340 (1997).

[71] None of the instructors were observed to consistently ask questions in a sequences progression as suggested by Meltzer *et al.* [Ref. 50].

[72] S. J. Pollock and K. Perkins, *APS Fed Newsletter*, Spring, 2004.

[73] Marcy Towns, "Using Clickers With 2500 Students: An Analysis of Question Type and Student Response." Invited Talk through the Science Education Initiative at the University of Colorado (2007) (unpublished).

[74] I. D. Beatty, W. J. Gerace, W. J. Leonard, and R. J. Dufresne, Designing effective questions for classroom response system teaching, Am. J. Phys. **74**(1), 31 (2006).

[75] N. W. Reay, L. Bao, P. Li, R. Warnakulasooriya, and G. Baugh, Towards the effective use of voting machines in physics lectures, Am. J. Phys. **73**(6), 554 (2005).

[76] J. Poulis, C. Massen, E. Robens, and M. Gilbert, Physics lecturing with audience paced feedback, Am. J. Phys. **66**(5), 439 (1998).

[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).

[78] C. Cazden, *Classroom Discourse: The Language of Teaching and Learning*, 2nd ed. (Heinemann, Portsmouth NH, 2001).

[79] M. C. O'Connor and S. Michaels, Aligning Academic Task and Participation Status through Revoicing: Analysis of a Classroom Discourse Strategy, Anthropol. Educ. Q. **24**(4), 318 (1993).

[80] H. Mehan, *Learning Lessons* (Harvard University Press, Cambridge, MA, 1979).

[81] C. C. Pappas and A. Barry, in *Reader Response in Elementary Classrooms*, edited by N. J. Karolides (Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997), pp. 215–237.