

The construction of different classroom norms during Peer Instruction: Students perceive differences

Chandra Turpen and Noah D. Finkelstein

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

(Received 7 October 2009; published 24 November 2010)

This paper summarizes variations in instructors' implementation practices during Peer Instruction (PI) and shows how these differences in practices shape different norms of classroom interaction. We describe variations in classroom norms along three dimensions of classroom culture that are integral to Peer Instruction, emphasis on: (1) faculty-student collaboration, (2) student-student collaboration, and (3) sense-making vs answer-making. Based on interpretations by an observing researcher, we place three different PI classrooms along a continuum representing a set of possible norms. We then check these interpretations against students' perceptions of these environments from surveys collected at the end of the term. We find significant correspondence between the researchers' interpretations and students' perceptions of Peer Instruction in these environments. We find that variation in faculty practices can set up what students perceive as discernibly different norms. For interested instructors, concrete classroom practices are described that appear to encourage or discourage these norms.

DOI: [10.1103/PhysRevSTPER.6.020123](https://doi.org/10.1103/PhysRevSTPER.6.020123)

PACS number(s): 01.40.Fk, 01.40.Di, 01.40.Ha

I. INTRODUCTION

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

In order to link PI implementation practices to students' perceptions of PI, it is necessary to aggregate particular fine-grained PI implementation practices into factors that would have meaning to students. Toward this end, we show how researchers can construct descriptions of classroom norms from collections of observed practices. We argue that collections of fine-grained classroom practices over time lead to the construction of classroom norms: a shared meaning system that gives sense or coherence to the community's collective activity. We focus on three particular dimensions of

classroom culture (continua of norms) integral to PI, emphasis on: faculty-student collaboration, student-student collaboration, and sense-making vs answer-making. The first two of these dimensions are particularly important because they are mechanisms by which student thinking is made visible and available for formative feedback from the professor or fellow peers. Understanding the emphasis on sense-making versus answer-making in the class allows us to describe the kinds of tasks or activities students are engaged in during PI. After presenting definitions of these norms along each dimension in terms of concrete classroom practices, we then investigate whether students' perceptions are consistent with the researchers' inferred norms, by surveying students' perceptions. We see strong consistency between the researchers' inferred classroom norms (what we believe to be going on in class) and students' perceptions of key elements of these classroom norms. These results suggest that how faculty (and students) engage in class establishes different rules and roles (norms) for participants. We show that what instructors do in the classroom correlates with what students perceive as valued in the activity of PI.

II. BACKGROUND

A. Description of Peer Instruction

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

- (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

If the percent of students getting the question correct is low after peer discussion, the concept is discussed again and another question cycle follows. In this way, the class adapts to the level of student understanding in the class. Mazur does not specify a particular technology (hands raised, colored cards, or personal response systems) to be used to collect students' votes in his descriptions of PI. This pedagogical strategy has many components, even within this short description.

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

We approach understanding curriculum adoption (such as the adoption of Peer Instruction) from a situative or sociocultural perspective which has significant implications for the way that we talk about curricula and tools. As Ball and Cohen explain, "While 'curriculum' is often taken to refer strictly to the textbook or curriculum materials, the enacted curriculum is actually jointly constructed by teachers, students, and materials in particular contexts" (Ref. [4], p. 7). We choose to focus on tool use in the classroom to capture enacted curriculum as distributed across the tools and the participants engaging in the activity. In this way, we see the material tools (i.e., clickers, or ConcepTests) and faculty as co-constituting the enacted curriculum. This perspective is described in additional detail in the theoretical approach section.

B. Research study purpose and research questions

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

- (1) Can differences in PI norms be delineated by using definitions of norms that are tightly linked to observable classroom practices?
Associated analytical questions:
 - (i) What types of faculty-student interaction are happening?
 - (ii) What roles and associated responsibilities are available to the classroom participants (faculty or students)?
 - (iii) What instructional practices foster or constrain student-student collaboration?
- (2) Do students notice differences in PI norms? And are these perceptions associated with what the researchers observe to be going on?

III. DEFINING CLASSROOM NORMS

A. Broad theoretical approach

We consider classrooms to be cultural systems which are constituted by norms of behavior that arise out of the repeated use of shared practices [5,6]. As other sociocultural researchers have claimed, "Every continuing social group, such as members of a classroom or workplace, develops a culture or set of social relationships that are peculiar and common to its members" [[7], p. 110]. Instructors and students make choices (implicit or explicit), which, in collection, establish a microculture with specific norms and expectations of the participants [5,8,9]. These microcultures can be described by the everyday activities of the participants, their ways of talking and interacting with each other, and their selective use of tools during their ongoing activity [10,11]. In our classrooms, students come to understand these microcultures in parallel with developing an understanding of physics content.

For the purposes of this research study, we take learning to be the internalization of social norms and practices. This definition captures both how students come to know a physics topic (e.g., the scientifically accepted definition of Newton's Second Law) and how students may simultaneously come to know how science is done (e.g., what counts as justification for a scientific answer). The emphasis in this definition on the everyday activities of the classroom is helpful for understanding how students may be learning things in physics classrooms that physics instructors are not explicitly intending to teach, often referred to as the "hidden curriculum" [12]. In the context of physics learning, investigating classroom microcultures is important because the practices of the classroom are tightly coupled to the new understandings about physics, the nature of learning, and the nature of physics that students develop as part of the class [9,13].

For example, consider a student working on completing a ConcepTest about Newton's Second Law while discussing with his or her peers. In completing this task, the student may develop an understanding of Newton's Second Law, but simultaneously the student may develop an understanding about how discussion with others is useful for clarifying scientific ideas (see Fig. 1). In this way the classroom practice of talking with your neighbor during class can simulta-

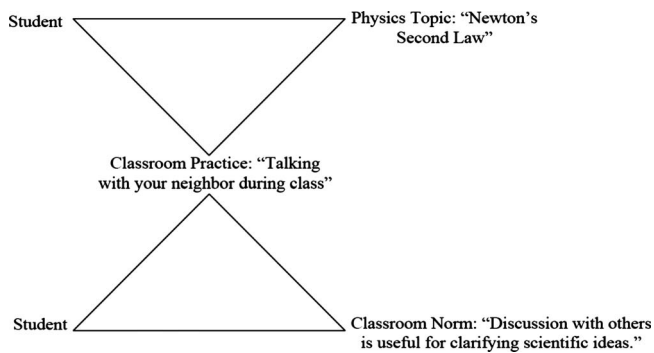


FIG. 1. Mediation of student learning by classroom practices.

neously support students' understanding of a physics topic and their understanding of classroom norms.

B. What are classroom norms and how are they constructed?

Norms are shared meanings or interpretations about the roles and rules of a social activity [5]. These shared meanings are largely implicit, but provide a degree of coherence and stability to local shared activity. Norms allow participants to coordinate their activity and share some common sense of "What is it that's going on here?" [14].

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

Norms are established through repeated engagement in social practices. Norms are socially negotiated and collectively agreed upon, although power and authority may not be equally distributed among participants. These norms carry with them implicit value sets of the culture (or microculture). Approaching classrooms as cultural systems (and describing their norms) helps to draw attention to the fact that teaching and learning "necessarily involves affording or constraining access to value-laden resources that affect the level and kinds of participation that individuals might achieve in a community" [7], p. 111]. For example, if communicating one's scientific ideas is a valued practice in a scientific discipline and students are not given the opportunity to practice communicating their ideas in a science course, then we have constrained students' access to this valued disciplinary resource.

Although describing patterns in social practices and classroom norms is useful, one must keep in mind that individuals

are always improvising and modifying these systems. As Lemke says, "People are not slaves to the activity structure of their community. We do not just 'follow the rules'—we use those rules as resources for playing the game according to our own strategies" [[15], p. 9]. The actions and interpretations of these actions are always highly contingent on the context.

C. Classroom norms in the context of Peer Instruction

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

We distinguish between two kinds of small-group discussions, one which consists of student-only small-group work and another which consists of the educator and student(s) working together in a small group. Whole-class discussion in the context of PI occurs largely during the introduction of the CQ, the first stage in the progression, and during the public discussion of the CQ solution, the third stage in the progression. These three modes of participation (whole-class discussion, student-only small-group discussion, and student-educator small-group discussion) were found to organize the vast majority of classroom interactions during Peer Instruction [16–19].

Within each of these modes of participation, we found multiple possible types of interaction that may occur. By specifying particular types of interactions that occur during these three modes of participation, we seek to codify the relations among participants, the corresponding roles of participants, and the normative expectations for appropriate behavior [20]. Each interaction consists of a series of actions or instructional moves such as the instructor leaving the stage. We refer to these instructional moves as part of clicker use since they are applicable to a wide variety of pedagogical uses of clickers (more broadly defined than PI). Each instructional move or classroom practice may have multiple possible meanings, however across the coordination of multiple practices over time particular meanings become more prevalent or preferred [21]. These prevalent meanings or interpretations make up the norms of the community.

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

mative feedback by fellow peers or the professor.

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

We define three dimensions of classroom culture that are of particular interest during the activity of Peer Instruction, emphasis on: faculty-student collaboration, student-student collaboration, and sense-making vs answer-making. Individual instructors (along with particular tools and students) establish norms (here we focus on three particular dimensions). These norms are reflected in the prevalence of particular practices and the values embedded in these practices. For example, one instructor may engage in practices that reflect a high value being placed on faculty-student collaboration and support the development of this norm in his or her class. While another instructor may engage in practices that reflect a low value being placed on faculty-student collaboration and which inhibit the development of this norm in his or her class. We note that a given practice may contribute to multiple classroom norms which results in norms that are overlapping and coconstituting. We have argued that classroom practices over time build up norms. Those norms are select instances of valuing a specific dimension of classroom culture (such as faculty-student collaboration). Sets of norms over time establish overall class culture.

We describe the common patterns of engagement of students and faculty in the classrooms of professors Yellow, Green, and Red. For full comparisons of the practices in these classes, please see the associated tables and figures in Ref. [2]. These data are summarized here. The descriptions of these classroom cultures and their constituent practices allow us to compare the norms of these classrooms along three dimensions. We then present student survey responses on questions designed to probe student thinking about the norms in their class along these dimensions. We show substantial correspondence between the researchers' interpreta-

tion of the PI classroom norms and students' perceptions of these classrooms. We see that particular collections of PI implementation practices may support students' perceptions of these activities as valuing faculty-student collaboration, student-student collaboration, and sense-making.

IV. REVIEW OF RESEARCH ON PI AND CLICKERS

Prior research literature relevant to understanding Peer Instruction, clicker use, and impacts on students falls into three categories: impacts of clicker use on students' content knowledge, effects of grading incentives on student discussion practices, and students' perceptions of the use of clickers. Reviewing the impacts of clicker use on students' content knowledge shows the complex nature of clicker implementation and its relation to student learning. The diversity of results related to student content learning suggests the importance of better understanding clicker implementation. This research also shows that the impact of clicker use has been fairly limited to measures of students' content knowledge neglecting other possible improvements in students' ability to communicate or generate scientific explanations. Research into the impacts of grading incentives on student discussion practices during PI demonstrates how professors' instructional practices can influence how students engage in the course. In this paper, we will investigate additional instructional practices that influence student engagement and perceptions. Researchers have devised a variety of constructs and associated assessment instruments to better understand students' own perspectives on their experiences learning physics. These instruments include: the Maryland Physics Expectations (MPEX) Survey [22], the Colorado Learning Attitudes about Science Survey (CLASS) [23], clicker-specific surveys, and typical midterm or end-of-term course evaluations. These instruments all attempt to build an understanding of the meaning that students' are making of physics instruction. Here we focus on reviewing clicker-specific surveys which are directly related to the measures of students' perceptions that will be presented later in this paper.

A. Impacts of clicker use on students' subject-specific content knowledge

As recent literature reviews of clicker use (broadly defined) in higher education across a variety of disciplines have stated [24–28], there is contradictory evidence concerning the impacts of clicker use on student learning across a variety of disciplines. While multiple studies have shown statistically significant improvements in students' course grades or examination scores through the use of clickers in their courses [29–32], others have shown no statistically significant improvement [33–40]. Studies show mixed results on the impact of clicker use on student learning. It is important to note that clicker use is very broadly defined in the literature reviewed in these articles and there is limited use of validated instruments to measure student learning.

However in the context of Peer Instruction, the specific way in which clickers are used is more specified and student

learning has been more systematically studied through the use of validated instruments. Research studies correlating effects of Peer Instruction with student learning show more consistent and promising implications. Crouch and Mazur demonstrated that the use of Peer Instruction in both their algebra-based and calculus-based physics courses at Harvard University resulted in improved Force Concept Inventory [41] performance, in addition to improvement on other measures [42]. Fagen *et al.* also showed that the conceptual learning gains documented at Harvard were found to be consistent with the learning gains documented across a range of institutions and courses [43]. It is important to note that in Fagen's study a specific criterion was applied to identify courses that implemented Peer Instruction [[44], p. 18].

The literature suggests a number of possible explanations for the conflicting research results regarding the impact of clicker use on student learning: (1) the rarity of valid and reliable measures of student learning across a variety of disciplines [26], (2) when comparing clicker and nonclicker courses, many additional pedagogical changes were made in the clicker courses that were not simultaneously changed in the nonclicker courses [27], and (3) the diversity of ways in which clickers were implemented in the experimental condition (clicker) classrooms [24]. The variation in student learning outcomes through the use of clickers is not well understood. In a review by Judson and Sawada, they claim however, that, "The only positive effects upon student academic achievement, related to the incorporation of electronic response systems in instruction, occurred when students communicated actively to help one another understand" [28]. This suggests that student discussion may be a critical element of PI. We use the term Peer Instruction (PI) loosely to describe what faculty at CU are doing with clickers (as part of the enacted curriculum), but when we are talking about particular actions associated with implementing PI we will often say 'clicker use' since these are decisions about or elements of clicker use that may be applicable more broadly.

B. Effects of grading incentives on student discussion practices

Recently several research articles have described the effects of grading incentives on student discussion practices during Peer Instruction [45–48]. We summarize the first two of these studies since they clearly link grading incentives to specific student discussion behaviors.

James studied how the assessment practices relating to CQs influenced the nature of conversations and degree of participation that occurred during Peer Instruction [45]. This study was conducted in two astronomy courses for first year nonscience majors being taught by two different instructors using three to five questions per lecture of their own design. One instructor (in the high-stakes course) had CQs count for 12.5% of the students' grade, where incorrect responses were awarded one-third the credit earned for the correct response. The other instructor (in the low-stakes course) had CQs count for 20% of the students' grade and incorrect responses earned as much credit as correct responses. CQs were discussed among pairs of students in both courses. Conversa-

tions between 12–14 pairs of students were recorded in three different classes of each course. These conversations were analyzed for "conversation bias," the difference between the fraction of all statements made by one partner and the fraction of all statements made by the other partner. If each person made the same number of statements, the conversation bias would be zero. The researchers found that in the high stakes classroom there was greater conversation bias and the students with more knowledge (indicated by higher course grade) tended to dominate the peer discussions. Students in the high stakes classroom were also more likely to vote in the same way as the other member of their discussion pair (only disagreed 8% of the time), whereas in the low-stakes classroom student pairs disagreed quite often (37% of the time). They conclude that when there is a grading incentive that strongly favors the correct responses to CQs, the question response statistics may exaggerate the degree of understanding that actually exists and confound the ability of the instructor to make accurate pedagogical decisions based on student response feedback [45].

In a follow-up study, James *et al.* followed the same instructor in the same environment and studied the impact of only changing grading incentives [46]. They found that with low-stakes grading the conversation bias was reduced and that the fraction of CQs where the discussion pairs disagreed in their responses went up [46]. These results confirmed that grading incentives had significant influence on these student discussion practices. This study suggests that course grading practices have clear effects on student-student collaboration. This study provides an example of how concrete instructional practices can influence how students' engage in PI activities associated with the course.

C. Students' perceptions of the use of clickers

In Judson and Sawada's recent survey of clicker research [28], they compare assessments of early use of clickers from the 1960s and 1970s when use tended to be informed by a behaviorist pedagogical orientation, to more recent assessments of clickers from the 1990s when use tended to be informed from more of a constructivist pedagogical orientation. Based on students' positive perceptions of clickers documented in multiple research studies [38–40,49–54] across a variety of implementations, Judson and Sawada conclude, "Students will favor the use of electronic response systems no matter the nature of the underlying pedagogy" [[28], p. 177].

The aim of more recent research studies has been to characterize students' experience and engagement with clickers along particular dimensions to better understand *why* students find clickers useful and enjoyable [55–61]. Some of these research studies have developed more sophisticated and robust instruments to measure students' perceptions rather than relying on single or few item survey instruments [60]. Currently, there are calls for additional research to specifically examine students' perceptions across varied pedagogical approaches and across varied student demographics [27]. A few recent research studies have begun to characterize whether students' perceptions of clickers vary based on

TABLE I. Course characteristics.

	Number of students enrolled in the course	Number of lecture sections	Average number of students attending a lecture section
Phys1 (Yellow)	566	2	212 \pm 5
Phys2 (Green)	339	1	227 \pm 5
Phys3 (Red)	156	1	132 \pm 2

characteristics of the students themselves [56,60]. We summarize two of these more recent detailed examinations of students' perceptions.

Graham *et al.* conducted a research study based on student survey responses aimed at answering the following research question: How does clicker use impact students who are least inclined to participate in class? [56]. This research study included 11 different instructors and a total of 688 students enrolled in nine different courses in the fields of chemistry, biology, physics, psychology, education, statistics, and marriage family and human development. Graham *et al.* identified three different ways of identifying students as "at-risk for low participation" based on students' self-reported orientation toward classroom learning in the survey: (1) students reluctant to share opinions in class, (2) students hesitant to ask questions in class, and (3) students who do not prefer courses where there is student participation. Reluctant students of type 1 and type 2 did not perceive clickers differently than their non-reluctant counterparts. Reluctant participants of type 3 were less likely than non-reluctant participants to view clickers as a helpful tool in the classroom. By also examining students' perceptions by course, the researchers note that differences in clicker pedagogical practices may have a larger effect on students' perceptions of its helpfulness than any of the psychological characteristics of the students [56].

MacGeorge *et al.* have developed a new instrument to measure students' perceptions of clickers across multiple dimensions called the Audience Response Technology Questionnaire (ART-Q) [60]. Using their newly developed instrument, they surveyed students in three large ($N > 200$) introductory courses (Communication, Forestry & Natural Resources, and Organizational Leadership & Supervision). Demographic characteristics (gender, ethnicity, and year in school) of students were not found to significantly correlate with student perceptions of clickers [60]. Similarly, in the forestry class, which had a variety of different majors from different schools on campus, there were no differences in students' perceptions of clickers by school or major. The researchers call for subsequent work to examine whether students' perceived benefit of clickers can be increased or lost if clickers are used in specific ways [60].

These wide-ranging lines of research into Peer Instruction all point to the need to investigate possible associations between PI implementation and impacts on students [27,62]. Building on prior research into student learning outcomes, student discussion behaviors, and students' perceptions, we examine three introductory physics courses which all include substantial student discussion during PI activities. We examine correlations between clicker implementation practices

and students' perceptions across three courses. We are less concerned about differences in the student body of these courses because of the lack of dependence of students' perceptions of clickers on student demographic characteristics that has been established in the literature. There are limited studies based on systematic classroom observations with two notable exceptions [51,63], and the research studies that have examined classroom practices have not compared a variety of implementations and associated impacts on students. This current work addresses this gap in the research. In this paper, we describe in detail how PI was implemented across three classrooms and show that students perceive norms of clicker use in these classrooms differently.

V. DESCRIPTION OF SETTING AND METHODS

A. Description of courses and instructors' background

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

B. Mixed-methodology research study design

In order to compare how Peer Instruction was implemented and to understand how participants made sense of

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

1. Ethnographic methods

Ethnographic research methods [64,65] 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 [66].

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

The observation rubric which was developed in the early stages of this research study was subsequently used to collect data in an additional 10 class periods which constituted at least 20% of the class periods for the course. Observations of these courses captured between 29–51 CQs for each course (depending on the frequency of CQs per class period). For other claims about classroom practices, all CQs were analyzed using data from the clicker software program such as percentage of students getting the question correct, etc. Unusual class periods, such as exam review days, or exam days, were excluded from our analysis [67].

Professors were asked to participate in the research study at the beginning of the semester and the observing researcher dropped into classes unannounced throughout the semester

and sat in different areas of the classroom among the students. In many instances the professor was unaware of the researcher's presence. On many occasions the researcher was also treated as a student by the surrounding students enrolled in the course (students would at times turn the researcher to discuss the CQ). These interactions as well as the low visibility of the observing researcher to the professor make us fairly confident that the presence of the research did not dramatically alter the environment being observed (although some influence is unavoidable).

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

The reliability of the in-person observations was checked through audio-recordings that were collected during the observations. These audio-recordings allowed the researcher to revisit CQ episodes. A subset of CQ episodes were transcribed to allow the researcher to reflect on and reconsider interpretations of CQ episodes after field notes were collected. Additionally, reliability studies were also conducted on the observation rubric with two other researchers [one from the Colorado physics education research (PER) group and one from an outside institution] who were trained on its use solely through reading the associated User's Guide (available in Ref. [2]). A reliability of 80% or greater was achieved on all items, averaging 96% agreement overall.

Following from an ethnographic tradition, the data sources for this study are both qualitative and quantitative, including: interviews with each professor (conducted at the end of the term), audio recordings of a subset of lecture periods, daily electronic records of CQs asked along with student responses, broad descriptive field notes, focused observational field notes through the use of a rubric [2], and student survey responses surrounding clicker use and classroom dynamics. For further discussion of the methods used for collecting and analyzing observational field notes, please see Ref. [2].

2. Survey methodology

During the last two weeks of the semester, students in each of the classrooms studied were asked to complete an optional anonymous online survey. The survey was announced on two different days of class as well as posted at the top of the class website for each course. Students were awarded a small amount of extra credit points for completing the survey (affecting less than 1% of any student's grade). The surveys were available online for approximately two weeks.

Survey questions were designed by the authors to target broad perceptions of the utility and enjoyment associated with PI and specific perceptions of a subset of classroom norms along three dimensions associated with the use of PI. Some of these survey questions were designed based on typical end-of-course evaluations probing students' broad per-

ceptions of the utility and enjoyment associated with PI. Additional survey questions were designed by the authors to target a subset of student thinking about faculty-student collaboration, student-student collaboration, and the emphasis on sense-making. Each question was worded to inquire about Peer Instruction or clicker use in a particular course specifically using student language found on prior long answer responses from prior semesters of survey data. The resulting survey included a mix of multiple-choice questions and open-ended short answer questions: approximately 25 Likert-scale items and about five open-response questions. For the complete survey used to collect data on students' perceptions of Peer Instruction, see Appendix C and associated supplementary materials online.

Since this was an optional survey, the percent of students enrolled in each course that completed the survey are 57% of Yellow's students ($N_{\text{respondents}}=323$), 45% of Green's students ($N_{\text{respondents}}=153$) and 58% of Red's students ($N_{\text{respondents}}=91$). These moderate response rates increase the possibility that survey respondents represent a nonrandom sample and this adds additional uncertainty to the results. Past analyses of student surveys administered at CU have shown that students who complete the surveys tend to receive slightly higher grades in the course on average than students who do not complete the surveys [68]. Results that are found may only well-represent the higher performing students and care must be taken in generalizing to the entire student population.

3. Methods for statistical analysis of student survey responses

A variety of statistical methods are used to identify course-by-course differences in student responses on Peer Instruction-specific survey questions. For survey questions that are categorical, but not rank ordered, a Chi-squared Test for Independence [[69], p. 204] is used to identify statistically significant variations across courses [70]. For survey questions which are categorical and rank ordered, a Kruskal-Wallis Test for k -independent samples [[69], p. 288] is used to identify statistically significant variations across courses [71]. A summary of the statistical test used for each question is given in Appendix A. When comparing the set of all courses, p -values from $0.10 > p > 0.05$ are taken to be marginally significant and p -values less than 0.05 are taken to be significant. On the questions which show significance or marginal significance, we compare semesters in a pairwise fashion (using either a Chi-squared Test for Independence or a Mann-Whitney U Test [[69], p. 272], depending on whether the answer options are rank-ordered or not). When determining the statistical significance of the pairwise comparisons, we will decrease the threshold for statistical significance depending on the number of pairwise semester comparisons that are being made for that particular question. The resultant threshold values for significance and marginal significance are summarized by question in Appendix A.

C. Study limitations

As mentioned above, there are multiple limitations to this study. With respect to the ethnographic data collection, the

possible effects due to the presence of the observing researcher cannot be fully understood, since video-recordings of the classes were not collected. We also note that the make-up of student majors enrolled in these courses is not the same (with the Phys3 course having a larger portion of engineering majors). Although this might be a factor, prior research has shown no significant correlation between students' perceptions of clicker use in a course and their declared major [60]. With respect to the survey methodology, the survey response rates pose a concern. Without all students' taking the survey, the survey respondents represent a nonrandom sample of the student population enrolled in the course. This could skew our results. Lastly, extensive validity and reliability studies have not been conducted on the survey used in this study. Unfortunately, another relevant instrument, the ART-Q survey instrument did not exist when this study was conducted [60]. However, the survey questions asked in this survey were designed using language found in students' long answer responses during prior terms. So although we cannot be certain that students' are interpreting these questions consistently, we believe that the questions are phrased using common student language. These concerns inject some uncertainty to our claims and suggest that research studies aimed at replicating these findings should be conducted.

VI. DATA AND RESULTS: COMPARING NORMS OF THREE INTRODUCTORY PHYSICS COURSES

A. Faculty-student collaboration

1. Defining faculty-student collaboration

Faculty-student collaboration is of particular interest because it is a prevalent method of scaffolding student learning and a common mode of feedback between faculty and students (formative assessment [72–74]). Understanding faculty-student collaboration requires answering the following questions: What types of faculty-student interaction are happening? How often are different types of faculty-student interaction occurring? What roles and associated responsibilities are available to which classroom participants (faculty or students)? We define “high” faculty-student collaboration to be a classroom where there are many types of faculty-student interaction that occur often during class and in which students and faculty take on and move among a diversity of roles in the class. In contrast, we define “low” faculty-student collaboration to be a classroom where there are few types of faculty-student interaction, which occur infrequently and in which students and faculty take on but a few stable roles in the class. These definitions describe the extremes along a particular dimension of classroom culture. Particular classroom norms are placed along a continuum representing a set of norms along this dimension. With these definitions and our observations of PI implementation practices [2], we can then characterize the classrooms of each of the professors that were observed.

In our prior work, we described observable aspects of faculty-student interactions that may occur during PI. Here we expand on this work, describing multiple types of

TABLE II. Types of Interactions

Small-group mode of participation

Type 1 Faculty and student in close proximity: When the instructor leaves the stage, the instructor is within earshot of the student conversations and the students can more easily get the instructor's attention while students are engaged in collectively solving the CQ. Possible instructor responsibilities: listening to students' talk, or scanning for students' bids for attention. Possible student responsibilities: listening to peer's descriptions of CQ solutions, criticizing or rebutting a peer's physics ideas, describing their physics ideas to their peers, working the physics problem on paper, listening or talking about out-of-class topics, or bidding for the professor's attention.

Type 2 Faculty responds to student questions: In this type of interaction a student raises his or her hand or calls out to the instructor as he or she passes by and the faculty member usually approaches the student to respond to the student's inquiry in a more private setting. Possible instructor responsibilities: listening to students' inquiry, or responding to the students' inquiry. Possible student responsibilities: getting the instructor's attention, presenting their question or inquiry, requesting clarification, requesting information, or requesting approval of their physical ideas.

Type 3 Faculty discusses with students: In this type of interaction, the professor approaches a student or group of students and instigates an interaction with an open-ended inquiring statement or question such as "what are you thinking about?" and engages in a discussion with students about the CQ. Possible instructor responsibilities: posing an open-ended question of students, listening to students' reasoning, asking additional follow-up questions or clarifying questions, and possibly challenging students' explanations. Possible student responsibilities: describing their physics ideas to the professor, responding to bids by the professor to elaborate or clarify their physical ideas, defend their physical ideas, modify their physical ideas, and articulate a physical argument.

Whole-Class mode of participation

Type 4 Faculty describes CQ problem: In this type of interaction, the professor presents a problem for the class to consider. Possible instructor responsibilities: designing the problem to be considered, describing the problem to the students, and linking the current problem to prior material presented in class. Possible student responsibilities: listening to professor's description of the problem and taking notes on material that is mentioned by the professor during the introduction of the problem.

Type 5 Faculty describes CQ solution: In this type of interaction, the professor presents a description of the CQ solution. Possible instructor responsibilities: completely describing the CQ solution. Possible student responsibilities: listening to the professors' explanation and taking notes on the professor's solution. (We note that it is quite possible that the students are engaged in other cognitive tasks; however, we focus only on the observable actions of students that are made public for feedback from either peers or the professor.)

Type 6 Faculty and student(s) describe CQ solution: In this type of interaction, the faculty and a student share responsibility for describing the CQ solution publicly, but the faculty member retains the majority of the responsibility for evaluating the CQ solution as well as designating speakers' turns. Possible instructor responsibilities: requesting student explanations, nominating a student for contributing an explanation, listening to the student's explanation, asking clarifying questions of the student, evaluating the student's explanation, and offering a revised expert explanation. Possible student responsibilities: listening to other students' explanations, offering their own explanations of the CQ solution, listening to the professor's explanation, and taking notes on the professor's solution.

Type 7 Faculty and student(s) jointly describe and evaluate the CQ solution: In this type of interaction the faculty member and students share responsibility for publicly describing and evaluating explanations of CQ solutions. A common rough indicator of this type of interaction is the inclusion of multiple students' perspectives in the public description of the CQ solution, since it is in these occasions that students participate to some degree in evaluating CQ explanations. Possible instructor responsibilities: requesting student explanations, nominating students for contributing explanations, listening to the students' explanation, asking clarifying questions of the students, requesting other students to comments on their peers' physics thinking, driving consensus in students' explanations, and offering a short summary of an expert explanation. Possible student responsibilities: listening to students' explanations, offering an explanation of one's CQ answer, commenting on their peers' physics thinking, debating or disagreeing with other students' reasoning, listening to the professor's explanation, and taking notes on the solution discussion.

faculty-student interaction (or modes of participation) and how these types of interactions limit the set of roles (or responsibilities) available to the participants (See Table II). In the prior section, we described two broad modes of participation for faculty-student interactions: (A) Student-educator small-group discussions in which only a small subset of students have access to this exchange and can therefore be considered semiprivate and (B) Whole-class discussions in which all students have access to this exchange and therefore occur in the public sphere. Within these two broad modes of participation, we detail types of faculty-student interaction and common, associated responsibilities of the faculty and students. The frequencies with which the various types of faculty-student interaction are employed, and the diversity of

the types used within a classroom setting, contribute to establishing a norm of valuing faculty-student collaboration.

We note that interaction Types 5–7 are mutually exclusive categories and as such each interaction was placed in one and only one of these categories. Any whole-class discussion of a given CQ solution falls into only one of these categories.

These different types of interactions are important because they constrain which responsibilities fall to the shoulders of different participants. They are also important in that they create different resources which are then available for use by the participants during ongoing classroom activities. For example, when a faculty member approaches a group of students discussing a physics concept, the faculty member has an opportunity to learn something new about students'

thinking that he or she was not previously aware of. This knowledge may simply inform how the instructor decides to proceed during the class period, or this talk can become a resource to be used in the immediate unfolding of the particular PI episode. The professor may privately ask the students if they would mind sharing their reasoning with the whole class. If the students agree, the professor now has an additional resource to use in leading a public discussion of the solution. We will return to discuss these emergent resources in the presentation of the final norm along the dimension of emphasis on sense-making vs answer-making.

2. Comparing three introductory physics courses based on implementation practices

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

We place the traditional lecture format at low faculty-student collaboration. This is because there is only one primary type of faculty-student interaction and all other types of faculty-student collaboration occur infrequently, if ever. This means that the roles and responsibilities of the educator and the students are very clearly defined and the responsibilities that students take on do not change often from class to class or over the course of the semester. The location of a traditional lecture format is shown leftmost in Fig. 2. We note that the continua presented in this paper are not based on a precise quantitative scaling, but represent a heuristic for the relative relations of the classrooms with respect to two extremes.

In order to describe the placement of Yellow, Green, and Red along this continuum, we code the frequency with which each type of faculty-student interaction occurred. A summary of this analysis can be found in Fig. 3.

In Yellow’s class, the professor rarely leaves the stage (only 12% of CQs Type 1), rarely answers student questions

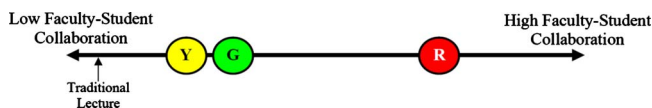


FIG. 2. (Color) A Continuum of a set of norms along the dimension of faculty-student collaboration.

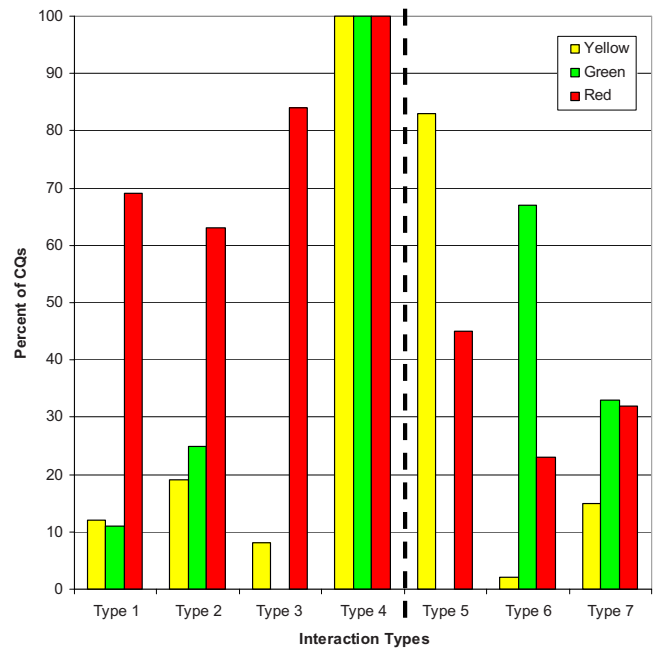


FIG. 3. (Color) Percent of clicker questions where the professor was observed to engage in each type of faculty-student interaction (Types of interactions are defined in Table I). The dashed line denotes that Types 5–7 at the rightmost side of graph are mutually exclusive categories.

(only 19% of CQs Type 2), rarely discusses with the students (only 8% of CQs Type 3), and rarely hears student explanations publicly (only 17% of CQs either of Type 6 or 7). Two percent of whole-class CQ explanations include the contribution of only one student, Type 6 as described in Table II. Fifteen percent of whole-class CQ explanations include the contributions of two or more students (Type 7). This means that the professor is usually conducting the description of the CQ solution entirely on his own (Type 5). Since faculty-student interactions Types 1–3 as described in Table II occur infrequently in Yellow’s class, we see that there are few opportunities for Yellow’s students to interact with the instructor within a small-group mode of participation. Similarly, since Type 5 is the highly prevalent mode of participation during the whole-class solution discussion, the students and faculty have fairly stable roles where students are rarely given responsibility over the solution description or the evaluation of the proposed solution description. These aspects of faculty practice provide evidence to support the placement of Yellow’s classroom at the lower end of the faculty-student collaboration continuum.

In Green’s class, the professor rarely leaves the stage (only 11% of CQs), rarely answers student questions (only 25% of CQs), never discusses with the students (0% of CQs), and always hears student explanations publicly (100% of CQs). Green usually hears from only one student during the whole-class public explanation of the CQ solution (67% of CQs). This student has usually offered a correct explanation. Green is usually quick to reveal correctness of student explanations. Since faculty-student interactions Types 1–3 occur infrequently in Green’s class, there are few opportunities for Green’s students to interact with the instructor within

a small-group mode of participation. Similarly, since Type 6 is the dominant mode of participation during the whole-class solution discussion, the students and faculty have fairly stable roles where students are often allowed some responsibility over the solution description, but little responsibility over the evaluation of the proposed solution. Since the majority of CQ solution descriptions occur between the faculty and a single student, students rarely have opportunities to comment on or disagree with the physical reasoning presented. These aspects of faculty practice provide evidence to support the placement of Green's classroom at the lower end of the faculty-student collaboration continuum.

In Red's class, the professor usually leaves the stage (69% of CQs), usually answers student questions (63% of CQs), usually discusses with the students (84% of CQs), and often hears student explanations publicly (55% of CQs). 23% of whole-class CQ explanations include the contribution of only one student. Thirty-two percent of whole-class CQ explanations include the contributions of two or more students. When Red requests student explanations, Red often hears from multiple students. Public debate and disagreement is supported and encouraged during these occasions. Red often withholds expert evaluation of answer correctness until consensus develops. Since faculty-student interactions Types 1–3 occur frequently in Red's class, we claim that there are many opportunities for Red's students to interact with the instructor within a small-group mode of participation. In Red's class, each of Types 5–7 occur fairly often. During the whole-class solution discussion, the students and faculty have fairly flexible roles where students have a varying degree of responsibility over the solution description and the evaluation of the proposed solution depending on question. These aspects of faculty practice provide evidence to support the placement of Red's classroom at the higher end of the faculty-student collaboration continuum.

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

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

To understand whether students perceived there to be relatively high or low value on faculty-student collaboration, we asked them four specific survey items that probed a subset of this classroom norm. For example, the first statement was: *It is awkward to ask my professor questions during class.* The students were given five answer options: *Strongly Disagree*, *Somewhat Disagree*, *Not Sure*, *Somewhat Agree*, and *Strongly Agree*. The distribution of student responses on this question is shown below in Fig. 4.

From this plot, we can see that students from Red's class are less likely than Green and Yellow's students to think that it is awkward to ask the professor questions. In Red's course we see that twice as many students are choosing disagree (somewhat disagree or strongly disagree) than are choosing

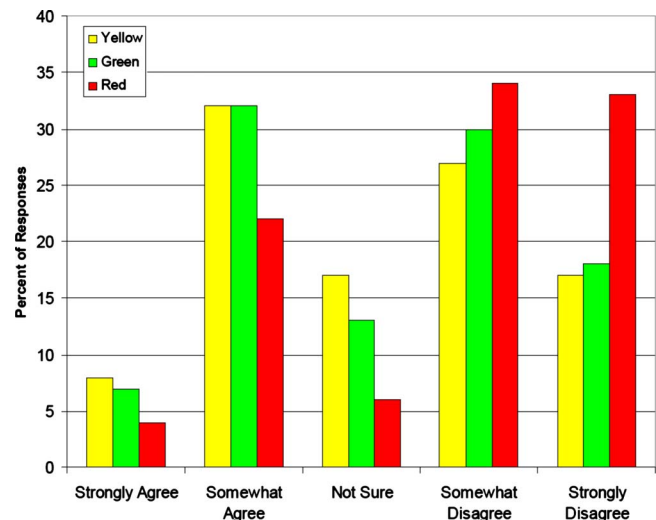


FIG. 4. (Color) Example Item—distribution of student responses to the statement: It is awkward to ask the professor questions in class.












agree (somewhat agree or strongly agree). Red's students tend to choose higher (more favorable) categories that Green and Yellow's on this question. We see in Table III that these differences are statistically significant).

We investigate whether this trend persists across an additional three survey questions. Question 16 asked, "How often do you raise your hand or ask questions in class?" (Never, About once a semester, About once a month, Nearly every week, and Nearly every class). Question 10 asked, "If my professor were to approach me in class during a clicker question, I would be comfortable discussing the content with my professor." (Strongly Disagree, Somewhat Disagree, Not Sure, Somewhat Agree, and Strongly Agree). Question 12 asked, "How often do you speak directly to the professor during class?" (Never, Once or twice a semester, Once every few weeks, Nearly every week, and Nearly every class). All of these questions have answer options that are categorical and rank ordered. Since there were statistically significant differences between these three groups via Kruskal-Wallis Test on all questions, the courses are compared pairwise using the Mann-Whitney U Test. The results of the pairwise comparisons are provided in Table III.

We see from Table III that Green and Red differ significantly on all four questions, with Red's students reporting more favorably on all questions [75]. We also see that Yellow and Red vary significantly on three of the four questions, with Red's students reporting more favorably on all questions. We do not find statistically significant differences in students' perceptions of faculty-student collaboration between Yellow and Green's courses. We conclude that Red's students perceive there to be a higher value placed on faculty-student collaboration than do both Yellow and Green's students. Yellow and Green's students perceive faculty-student collaboration similarly.

These four survey questions show that students from these different courses notice some of the same differences in classroom norms as the research observer. The correspondence between the observing researcher's interpretation of

TABLE III. (Color) Mann-Whitney U Test Results for Pair-wise Comparisons on Faculty-Student Collaboration Survey Questions. The arrows in the table indicate which group tends to yield higher (more favorable) responses for all p -values less than 0.1. Bold font (with asterik) indicate statistically significant results.

	Yellow-Red	Green-Red	Yellow-Green
Q11 Awkwardness asking professor questions	 $p < \mathbf{0.001}^*$	 $p = \mathbf{0.002}^*$	$p = 0.629$
Q16 Frequency asking the professor questions	 $p < \mathbf{0.001}^*$	 $p < \mathbf{0.001}^*$	 $p = 0.032$
Q10 Comfort discussing with professor	 $p = 0.033$	 $p = \mathbf{0.001}^*$	 $p = 0.034$
Q12 Frequency speaking with professor	 $p < \mathbf{0.001}^*$	 $p < \mathbf{0.001}^*$	 $p = 0.024$

the norm of faculty-student collaboration and students’ perceptions regarding a subset of this norm supports the claim that concrete classroom practices associated with the implementation of PI can impact the way that students’ perceive faculty-student collaboration.

B. Student-student collaboration

1. Defining student-student collaboration

Student-student collaboration is of particular interest because it is a highly prevalent method of promoting active engagement [76] and a common strategy for making students’ thinking visible to the educator and other students allowing for formative assessment [72–74]. Understanding student-student collaboration requires answering the following questions: what types of student-student interaction are happening? How often are different types of student-student interaction occurring? What roles and associated responsibilities are available to students? We have not studied student-student interactions in detail in this study (as discussed in the methods section). We examine student-student collaboration by looking for faculty practices that constrain or allow for different amounts and kinds of student-student collaboration. We focus on the faculty practices of grading incentives, fraction of CQs where peer collaboration was allowed, prevalence of introductory comments encouraging peer collaboration, fraction of class time spent in peer collaboration, and the frequency and type of opportunities for faculty to model discussion practices with students both in small-group and whole-class modes of participation.

We define “high” student-student collaboration to be a classroom where there are low-stakes grading practices, sig-

nificant opportunities for peer collaboration (i.e., peer collaboration usually allowed, relatively long CQ voting time provided, and relatively high percent of class time devoted to explicit peer collaboration), consistent explicit encouragement of peer collaboration (i.e., introductory comments supporting or encouraging talk among peers), and frequent opportunities for the instructor to model scientific discourse (i.e., prevalence of faculty answering questions and discussing with peers in small-group interactions, and Type 6&7 interactions during whole-class discussion). In contrast, we define “low” student-student collaboration to be a classroom where there are high stakes grading practices, few opportunities for peer collaboration (i.e., short length of CQ voting time, and low percent of class time devoted to explicit peer collaboration), little explicit encouragement of peer collaboration (i.e., few introductory comments supporting or encouraging talk among peers), and few opportunities for the instructor to model scientific discussion practices (i.e., few occasions of faculty answering questions and discussing with peers in small-group interactions, and Type 6&7 of whole-class discussion). These definitions describe the extremes along a particular dimension of classroom culture. Particular classroom norms are placed along a continuum representing a set of norms along this dimension. With these definitions and our observations of PI implementation practices [2], we can then characterize the classrooms of each of the professors that were observed.

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

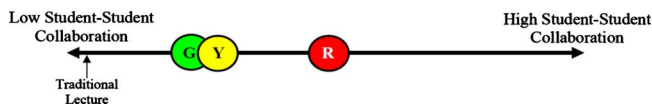


FIG. 5. (Color) A Continuum of a set of norms along the dimension of student-student collaboration.

discussion practices that arise around disagreement or dispute. In Type 7 interactions, there is potential to model these discussion practices surrounding disagreement and dispute.

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

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

2. Comparing three introductory physics courses based on implementation practices

Here, we examine variation in classroom practices and associated norms along the dimension of student-student collaboration. We begin by characterizing student-student collaboration in a traditional lecture format. In a traditional lecture format course, there are usually few to no opportunities for students to test out their thinking without being graded. There are also few to no opportunities for students to talk with other students during the lecture; therefore, if peer collaboration is explicitly encouraged at all, it is expected to occur outside of the regular lecture time. Instructors rarely have opportunities to model discussion practices in traditional lecture courses since the instructor usually does all or most of the talking. We indicate that traditional lecture format would be placed at low student-student collaboration (Fig. 5). We also note that on this continuum none of the instructors are at the high end of student-student collaboration because even in the class that is most encouraging of student-student collaboration, only approximately a third of

class time is spent engaged in student-centered activities.

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

Green’s class had moderate-stakes grading practices because there was some evaluative emphasis placed on correctness, but, as with Yellow, only for awarding extra credit. In Green’s class, peer collaboration was allowed for all CQs, students were given about three and a quarter minutes on average to complete a CQ with their peers (with 64% of CQs lasting more than 2 min), and approximately 17 percent of class time was explicitly devoted to peer collaboration. Green’s students and Yellow’s students have similar amounts of time to respond to CQs, however slightly more class time is spent on CQs in Yellow’s class. Based on these criteria significant opportunities for peer collaboration were available in Green’s class, as in Yellow’s class. Green usually read the CQ out loud elaborating on what the diagrams represented, but rarely made introductory comments explicitly encouraging peer collaboration. Green made fewer remarks encouraging peer collaboration as compared to Yellow. In Green’s class (as in Yellow’s) there were few opportunities for the instructor to model scientific discussion practices in the context of small-group interactions. Green rarely created opportunities to interact with students in a small-group format (left the stage 11% of the time, answered student questions 25% of the time, and never discussed with students). In Green’s class there were some opportunities for the instructor to model productive student discussion practices in the con-

text of whole-class discussions because student explanations were always heard. As mentioned in the prior section, Green usually only heard from a single student about his or her reasoning (67% Type 6). When Green did hear from multiple students, he was usually employing Type 7a of student-faculty interaction—that is, he was generally not mediating or supporting disagreement among students. To summarize, Green’s course had moderate-stakes grading practices, moderate opportunities for peer collaboration, but few explicit remarks encouraging peer collaboration and few opportunities for the instructor to model scientific discourse in a small-group format. There were however moderate opportunities for the instructor to model productive student discourse in the whole-class format. These aspects of faculty practice provide evidence to support the placement of Green’s classroom near the lower end along the continuum of student-student collaboration.

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

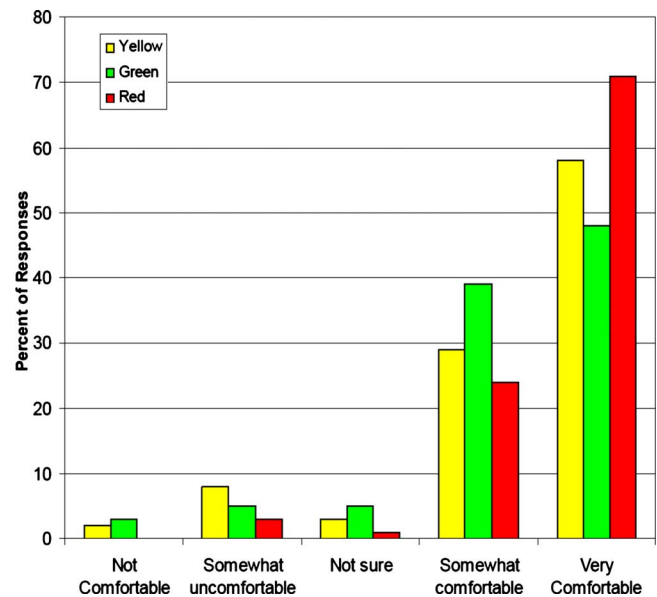


FIG. 6. (Color) Example question—distribution of student responses to Comfort level discussing physics with student peers (Q14)

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










To understand whether students’ perceived there to be relatively high or low value placed on student-student collaboration, we asked them four specific survey questions. For example, the first question was: *How comfortable do you feel discussing the course content with your peers during clicker questions? (Not comfortable, Somewhat comfortable, Not sure, Somewhat comfortable, and Very comfortable)*. The distribution of student responses on this question is shown in Fig. 6.

Figure 6 shows that most students from all three classes report being either somewhat comfortable or very comfortable discussing with their peers. However, Red’s students were more likely to report being very comfortable discussing with their peers. Between 15% and 20% more students in Red’s course chose “very comfortable” as compared to Yellow’s course and Green’s course.

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

Question 6 asked: “To what extent does your instructor usually encourage student-to-student discussion about clicker questions in class?” The students were given answer options that included: (1) does not allow discussion, (2) allows discussion, but does not encourage it, and a small fraction of students discuss, (3) allows discussion, but does not encourage it, and a large fraction of students discuss, (4) encourages discussion, and a small fraction of students discuss, (5) encourages discussion, and a large fraction of students discuss. No students chose the first answer option in any of the physics courses studied, so this answer option was eliminated in the analysis. Then the remaining answer options were col-

TABLE IV. (Color) Mann-Whitney U Test Results for Pairwise Comparisons on Student-Student Collaboration Survey Questions. The arrows in the table indicate which group tends to yield higher (more favorable) responses for all p -values less than 0.1. Bold font (with asterisk) indicate statistically significant results.

	Yellow-Red	Green-Red	Yellow-Green
Q6a Degree of encouragement for peer discussion	$p=0.384$	 $p=0.003^*$	 $p=0.001^*$
Q7 Usual discussion practices of individual students	 $p=0.043$	 $p=0.001^*$	 $p=0.077$
Q14 Comfort discussing with peers	 $p=0.009^*$	 $p<0.001^*$	$p=0.146$
Q15 Amount of speaking with peers	 $p<0.001^*$	 $p<0.001^*$	$p=0.291$

lapsed in two different ways for analysis. The first grouping, question 6a, was based on whether the instructor allowed discussion (options 2 and 3) or encouraged discussion (options 4 and 5). The second grouping, question 6b, was based on whether a small fraction of students discussed (options 2 and 4) or a large fraction of students discussed (3 and 5).

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

Question 15 asked: “On average, what fraction of class time do students speak either with each other or to the professor?” The students were given answer options that ranged from less than 5 min out of a 50-min class to more than 20 min out of a 50-min class (with answer options in between in increments of 5 min). All of these questions have answer options that are categorical and rank ordered. There were statistically significant differences between these three groups via Kruskal-Wallis Test on all questions, except question 6b. Thus, the level of student discussion perceived by students in each of these three courses was statistically indistinguishable (See Appendix A, Q6b). The courses are compared pairwise using the Mann-Whitney U Test on all questions, except for 6b. The results of the pairwise comparisons are provided in Table IV.

We see from Table IV that Green and Red differ significantly on all four questions, with Red’s students reporting more favorably on all questions. We also see that Yellow and Red vary significantly on two of the four questions, with Red’s students reporting more favorably on three of the four

questions. We find that Green and Yellow vary significantly on only one of the four questions, with Yellow’s students reporting more favorably on two of the four questions. We conclude that Red’s students perceive there to be a significantly higher value placed on student-student collaboration than do Green’s students. We conclude that Red’s students perceive there to be a somewhat higher value placed on student-student collaboration than do Yellow’s students. Overall, we do not see many significant differences between how Yellow’s students and Green’s students perceive student-student collaboration.

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

C. Emphasis on sense-making versus answer-making

1. Defining emphasis on sense-making versus answer-making

Researchers have distinguished two modes of student engagement in classroom science learning: answer-making and sense-making. Answer-making has been defined as a less productive framing of school science activities aiming to get students to know the scientifically accepted answers that scientists have developed to describe the natural world [77,78]. In this mode, students are usually trying to come to the explanation that they think the teacher wants to hear rather than coming to an explanation that makes sense to the student. Sense-making has been defined as a productive framing of school science activities in which the aim is to get students to build sensible and plausible models of the natural world that are intelligible to the students themselves [77,78]. Students as well as instructors play an active role in framing school science activities with one of these emphases. Researchers at the University of Maryland have done significant work characterizing evidence of students using a sense-making frame

versus an answer-making frame in small collaborative group problem-solving environments [78]. Here, we seek to focus on instructional moves, and to characterize classroom norms along the dimension of sense-making versus answer-making within Peer Instruction activities.

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

2. Comparing three introductory physics courses based on implementation practices

To begin with a familiar starting point, we describe the emphasis of sense-making versus answer-making in a traditional lecture format. As mentioned in the prior dimension of classroom culture, traditional lecture courses generally use high-stakes grading since there is usually few to no opportunities for students to test out their thinking without potentially affecting their grade. There are also few to no opportunities for students to share their physical reasoning with other students or with the instructor during the lecture; therefore there are few opportunities for student discussion of physical reasoning in either small group or whole-class discussion formats. Since the instructor usually does all or most of the talking in a traditional lecture, there is little opportunity for students to test out and receive feedback on their physical reasoning which provides few occasions for students and the instructor to negotiate meaning. For these reasons, a traditional lecture course would be categorized as at the emphasizing answer-making end of the continuum rather than sense-making end of the continuum (Fig. 7).

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

Green's class, like Yellow's class, had moderate-stakes grading practices. In Green's class, there was some explicit emphasis on sense-making or reasoning by the instructor. Green, as with Yellow, rarely reminded students to explain their reasoning behind choosing a particular answer when introducing a CQ. However, CQ answers and associated explanations were made available to students in Green's class. Similar to Yellow, there were many opportunities for students to discuss their physical reasoning in small-group formats since the CQ voting time was relatively long and the majority of CQs were conceptual. In Green's class, students also had frequent opportunities to discuss their physical reasoning in whole-class discussions. Student explanations were always heard during the public discussion of the CQ solution (100% of the time). Green, as with Yellow, rarely interacted with students in small-group formats. However, Green always heard a student explanation in the whole-class format. Green therefore had moderate exposure to students' reasoning and a few opportunities to encourage student reasoning in face-to-face interactions. In Green's class, incorrect answers to the CQs and associated flaws in reasoning were rarely discussed during the public solution description (only 25% of the time). There was a moderate amount of opportunities for the instructor and the students to negotiate meaning of physical ideas. These aspects of classroom practice provide evidence to support the placement of Green's classroom near the "answer-making" end of this continuum (Fig. 7).

Red's class had fairly low-stakes grading since evaluative emphasis was rarely placed on correctness as described previously. In Red's class, there was often explicit emphasis on sense-making or reasoning by the instructor. Red often reminded students to explain their reasoning behind choosing a particular answer when introducing a CQ, and often told students that they were going to be asked to offer their expla-

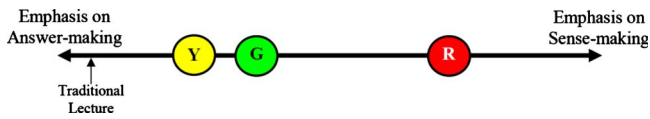


FIG. 7. (Color) A Continuum of a set of norms along the dimension of answer-making versus sense-making.

nations to the rest of the class. Additionally, CQ answers and associated explanations were made available to students and integrated into the lecture notes for the class. There were many opportunities for students to discuss their physical reasoning in small-group formats since the CQ voting time was relatively long and the majority of CQs were conceptual. In Red's class, students also had frequent opportunities to discuss their physical reasoning in whole-class discussions. Student explanations were usually heard during the public discussion of the CQ solution (55% of the time). Red often interacted with students in small-group formats (either by answering questions or discussing with students) and he usually heard student explanations in the whole-class format. Therefore, Red had high amounts of exposure to students' reasoning and opportunities to encourage student reasoning in face-to-face interactions. In Red's class, incorrect answers to the CQs and associated flaws in reasoning were often discussed during the public solution description (58% of the time). There were many opportunities for the instructor and the students to negotiate meaning of physical ideas. Red's classroom is placed toward the "sense-making" end of this continuum.

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

To understand whether students perceived the emphasis of CQs in the course to focus on sense-making versus on answer-making, we asked them two specific survey questions. The first question was: *In class, how important is it for you to articulate your reasoning either to your peers or during whole-class discussion? (Not important at all, Not very important, Somewhat Important, Important, and Very Important)*. The distribution of student responses on this question is shown in Fig. 8.

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

We investigate whether this trend persists across a second survey question: *Knowing the right answers is the only important part of the clicker questions (Strongly Disagree, Somewhat Disagree, Not Sure, Somewhat Agree, and Strongly Agree)*. Both of these questions have answer options that are categorical and rank-ordered. Since there were statistically significant differences between these three groups

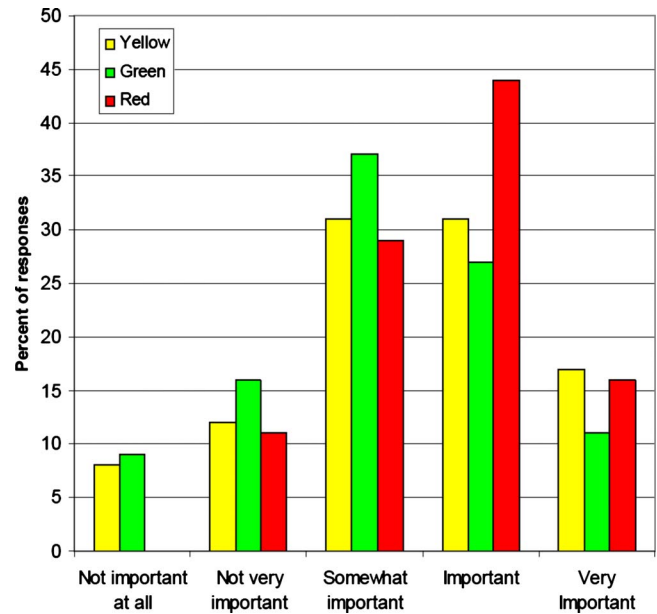


FIG. 8. (Color) Example question—distribution of student responses to the importance of articulating their reasoning.

via Kruskal-Wallis Test on both questions, the courses are compared pairwise using the Mann-Whitney U Test as described before. The results of the pairwise comparisons are provided in Table V.






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

These two survey questions show that students from these different courses notice some of the same differences in classroom norms as the research observer. The researcher perceived Green to emphasize sense-making more than Yellow, however these differences were small. It is unclear if the targeted survey questions asked of students were sufficient to fully characterize students' perceptions of sense-making in these classes or if the changes detected by the researcher were smaller than the students detected. Overall, the correspondence between the observing researcher's interpretation sense-making in the classroom culture and students' perceptions regarding this dimension of classroom culture supports the claim that concrete classroom practices associated with the implementation of PI can impact the way that students perceive the emphasis on sense-making.

VII. DISCUSSION AND CONCLUSION

We have defined continua representing a set of norms along three dimensions, emphasis on: faculty-student col-

TABLE V. (Color) Mann-Whitney U Test Results for Pairwise Comparisons on Sense-making Survey Questions. The arrows in the table indicate which group tends to yield higher (more favorable) responses for all p -values less than 0.1. Bold font (with asterik) indicate statistically significant results.

	Yellow-Red	Green-Red	Yellow-Green
Q17 Articulate reasoning	 $p=0.067$	 $p<0.001^*$	 $p=0.045$
Q13 Right answer clicker	 $p=0.002^*$	 $p=0.001^*$	$p=0.392$

laboration, student-student collaboration, and sense-making vs answer-making. Through observations and analysis of classroom practices, we note the norms associated with three different PI classrooms and place these classrooms along three continua of classroom culture. Students enrolled in these courses perceive differences in classroom norms. We claim that concrete actions in the classroom lead to differences in classroom norms that are reflected in students' survey responses.

The curious reader may question whether there is a simpler explanation for the differences in students' perceptions. Red is the only physics education researcher and therefore may have a deeper understanding of the underpinning principles that make PI effective. Participation in PER may correlate with the construction of particular classroom norms, but we are interested in characterizing how these norms are constructed in the classroom. Defaulting to personal characteristics of Red does not help us to describe *how* Red's knowledge is employed in the classroom. We seek a description of a plausible mechanism for how students develop different understandings about physics, the nature of learning, and the nature of physics that students develop as part of the classroom culture and associated social practices. We seek to describe the social practices (concrete interactions between students and instructors) in each of these classrooms in hopes of better understanding how students come to different understanding of what PI is about and what is valued in their classes. We have drawn attention to concrete actions in the classroom that may support students in developing a sense that faculty-student collaboration, student-student collaboration, and sense-making are important in learning physics during Peer Instruction.

Students' prior experiences in physics courses and particularly physics courses that utilize PI will certainly influence the meaning that students' make of PI in a given course. Enculturation of students over multiple semesters of using PI will lead to different expectations and possibly different perceptions of classroom norms. In some sense, that is what we argue within the span of a course. If however, the variation in student perceptions were solely a matter of prior experience with Peer Instruction, we would expect to see a more dramatic shift in students' perceptions between Phys1 (Yellow) and Phys2 (Green). More explicitly, in comparisons of two different Phys2 courses (Green vs Purple) in Appendix B, it is clear that positive student perceptions are not simply a matter of how many semesters students have been exposed to

clicker use. Students' perceptions vary dramatically within the same Phys 2 course offered by different instructors. For more details about Purple's PI implementation practices [79], please see Ref. [2]. Since Purple's students had taken the prior course, Phys1, with Yellow, we see that students' perceptions of classroom norms surrounding PI can shift dramatically over the course of a single semester.

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

This work suggests many avenues for future research. There may be a wide variety of benefits to students associated with particular ways of using clickers and PI. Researchers could begin to assess other possible benefits to students of particular implementations of PI such as changes in how students verbally communicate their scientific ideas, or changes in how students explain their reasoning associated with physics problem solutions. Attempts to replicate the findings of this paper across other classrooms would also be

useful. Further investigations of PI implementation would be worthwhile to address the following questions: how does PI implementation (practices and norms) vary within and across different institutions? At institutions where faculty closely collaborate in their teaching or spend a significant amount of time observing each other's teaching, is there more consistency in PI implementation? Additional contrasting cases of PI implementation collected from a broader range of institutions would help in fleshing out a more complete understanding of clicker implementation and its associated impacts on students.

Prior research has called for a more detailed examination of classroom practices surrounding the use of clickers and associated impacts on students [62]. Professors' PI practices do vary and PI is perceived differently by students in these classrooms. We define collections of PI practices which are associated with different classroom norms. In this paper, differences in classroom norms documented through classroom observations are associated with differences in how PI is perceived by students. We specify collections of classroom practices that appear to support students' sense that PI is about faculty-student collaboration, student-student collabo-

ration, and sense-making. Classroom norms may be specific to the personal preferences and learning goals of individual instructors, but the concrete practices articulated in this paper may help instructors enact the goals (norms) they seek to encourage in their classrooms during Peer Instruction.

ACKNOWLEDGMENTS

We are grateful for the contributions of many faculty members in the University of Colorado physics department for opening their classrooms to researchers and their willingness to examine their classroom practices. We would also like to thank the Physics Education Research Group at Colorado for their thoughtful comments on earlier versions of this work. We are particularly grateful for the useful discussions with Lauren Kost, Jessica Watkins, and Elias Quinn. This work has been supported by the National Science Foundation (Grants No. 0410744 and No. 0448176), the AAPT/AIP/APS (Colorado PhysTEC program), the Center for the Integration of Research, Teaching, and Learning—CITRL (NSF: 0717768), and the University of Colorado at Boulder.

APPENDIX A: KRUSKAL-WALLIS p -VALUES FOR SET OF THREE SEMESTERS OF DATA

Question number	p -value	Pairwise significance thresholds
Q2	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q3	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q4a	$p = 0.002$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q4b	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q5a	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q5b	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q6a	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q6b	$p = 0.034$	Not significant
Q7	$p = 0.006$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q8a	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q8d	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q10	$p = 0.003$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q11	$p = 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q12	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q13	$p = 0.002$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q14	$p = 0.002$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q15	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q16	$p < 0.001$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$
Q17	$p = 0.005$	Marginally Significant= $0.033 > p > 0.003$; Significant= $p < 0.003$

APPENDIX B: SEMESTER-BY-SEMESTER PAIRWISE COMPARISONS VIA MANN-WHITNEY TEST

		Associated p-values for Instructor/Course Pair-wise comparisons			
		Yellow vs. Red	Green vs. Red	Yellow vs. Green	Green vs. Purple
		Phys1 vs. Phys3	Phys2 vs. Phys3	Phys1 vs. Phys2	Phys2 vs. Phys2
Overall Course Perceptions	Q2	p<0.001	p<0.001	p=0.083	p<0.001
	Q3	p<0.001	p<0.001	p=0.014	p<0.001
Broad Perceptions of Clickers	Q4b	p=0.944	p<0.001	p<0.001	p<0.001
	Q5b	p=0.978	p=0.002	p<0.001	p<0.001
	Q8a	p<0.001	p<0.001	p=0.629	p<0.001
	Q8d	p=0.001	p=0.986	p<0.001	p=0.093
Emphasis on Sense-making	Q13	p=0.002	p<0.001	p=0.392	p<0.001
	Q17	p=0.067	p=0.001	p=0.045	p<0.001
Faculty-Student Collaboration	Q10	p=0.033	p=0.001	p=0.034	p<0.001
	Q11	p<0.001	p=0.002	p=0.629	p=0.167
	Q12	p<0.001	p<0.001	p=0.024	p=0.002
	Q16	p<0.001	p<0.001	p=0.032	p=0.376
Student-Student Collaboration	Q6a	p=0.384	p=0.003	p=0.001	NA
	Q6b	NS	NS	NS	NA
	Q7	p=0.043	p=0.001	p=0.077	NA
	Q14	p=0.009	p<0.001	p=0.146	p=0.029
	Q15	p<0.001	p<0.001	p=0.291	p=0.834

NS: No statistically significant difference found via Kruskal-Wallis Test

NA: Not Available (Data for one of the semesters, not available for comparison)

Bold typeface: p-values below the threshold for marginal statistical significance

Shading: Color of the shading denotes the professor whose students responded more favorably (shading shown when $p < 0.1$).

APPENDIX C: SUPPLEMENTARY AUXILIARY MATERIAL

See separate auxiliary material for complete survey used to collect data on students' perceptions of Peer Instruction.

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

[2] C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020101 (2009).

[3] C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur, Peer Instruction: Engaging Students One-on-One, All At Once, in *Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney, published by (American Association of Physics Teachers, College Park, MD, 2007).

[4] D. L. Ball and D. K. Cohen, Reform by the Book: What is—or Might Be—the Role of Curriculum Materials in Teacher Learning and Instructional Reform?, *Educ. Res.* **25**, 6 (1996).

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

[6] K. Gutierrez and B. Rogoff, Cultural ways of learning: Individual traits or repertoires of practice, *Educ. Res.* **32**, 19 (2003).

[7] K. Kumpulainen and P. Renshaw, Cultures of learning, *Int. J. Educ. Res.* **46**, 109 (2007).

[8] P. Cobb, M. Stephan, K. McClain, and K. Gravemeijer, Participating in classroom practices, *J. Learn. Sci.* **10**, 113 (2001).

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

[10] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation* (Cambridge University Press, Cambridge, MA, 1991).

[11] E. Wenger, *Communities of Practice: Learning, Meaning and Identity* (Cambridge University Press, Cambridge, MA, 1998).

[12] E. F. Redish, *Teaching Physics with Physics Suite* (John Wiley & Sons, Hoboken, NJ, 2003).

[13] L. Vygotsky, *Mind in Society: The Development of Higher*

- Psychological Processes* (Harvard University Press, Cambridge, MA, 1978).
- [14] R. Scherr and D. Hammer, Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics, *Cogn. Instruct.* **27**, 147 (2009).
- [15] J. L. Lemke, *Talking Science: Language, Learning, and Values* (Ablex Publishing Corporation, Norwood, NJ, 1990).
- [16] We note that these modes of participation have been referred to by other education researchers as participant structures (see Philips, 1983; Herrenkohl and Guerra, 1998) or participation frameworks (see Goffman, 1981).
- [17] S. Philips, *The Invisible Culture: Communication in the Classroom and Community on the Warm Springs Indian Reservation* (Longman, White Plains, NY, 1983); S. Philips, *Participant structures and communicative competence: Warm Springs children in community and classroom* edited by C. B. Cazden, V. P. John, and D. Hymes, *Functions of Language in the Classroom* (Teachers College Press, New York, NY, 1972), pp. 370–394.
- [18] L. R. Herrenkohl and M. R. Guerra, Participant structures, scientific discourse, and student engagement in fourth grade, *Cogn. Instruct.* **16**, 431 (1998).
- [19] E. Goffman, *Forms of Talk* (University of Pennsylvania Press, Philadelphia, PA, 1981).
- [20] I. Horn, in *A Study of Teaching: Multiple Lenses, Multiple Views, JRME Monograph #14*, edited by D. Ball and A. Schoenfeld, Journal for Research in Mathematics Education Monograph Series (Reston, VA, 2008).
- [21] J. V. Wertsch and C. Toma, in *Constructivism in Education*, edited by L. P. Steffe and J. Gale (Erlbaum, Hillsdale, NJ, 1995), pp. 159–174.
- [22] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**(3), 212 (1998).
- [23] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, 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).
- [24] J. R. MacArthur and L. L. Jones, A review of literature reports of clickers applicable to college chemistry classrooms, *Chem. Educ. Res. Pract.* **9**, 187 (2008).
- [25] V. Simpson and M. Oliver, Electronic voting systems for lectures then and now: A comparison of research and practice, *Australasian Journal of Educational Technology* **23**, 187 (2007).
- [26] J. E. Caldwell, Clickers in the large classroom: Current Research and Best-Practice Tips, *CBE Life Sci. Educ.* **6**(1), 9 (2007).
- [27] C. Fies and J. Marshall, Classroom response systems: A review of the literature, *J. Sci. Educ. Technol.* **15**, 101 (2006).
- [28] E. Judson and D. Sawada, Learning from past and present: Electronic response systems in college lecture halls, *J. Comput. Math. Sci. Teach.* **21**, 167 (2002).
- [29] J. Poulis, C. Massen, E. Roberts, and M. Gilbert, Physics lecturing with audience paced feedback, *Am. J. Phys.* **66**, 439 (1998).
- [30] R. H. Hall, H. L. Collier, M. L. Thomas, and M. G. Hilgers, *Proceedings of the Eleventh Americas Conference on Information Systems*, August 11–14, 2005, Omaha, NE. (Curran Associates Inc., Red Hook, NY, 2005).
- [31] D. W. Bullock, V. P. LaBella, T. Clingan, Z. Ding, G. Stewart, and P. M. Thibado, Enhancing the student-instructor interaction frequency, *Phys. Teach.* **40**, 535 (2002).
- [32] 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).
- [33] C. Paschal, Formative assessment in physiology using a wireless classroom communication system, *Adv. Physiol. Educ.* **26**, 299 (2002).
- [34] D. Bunce, J. VandenPlas, and K. Havanki, Comparing the effectiveness on student achievement of a student response system versus online WebCT quizzes, *J. Chem. Educ.* **83**, 488 (2006).
- [35] G. Kennedy and Q. Cutts, The association between students' use of an electronic voting system and their learning outcomes, *J. Comput. Assist. Learn.* **21**, 260 (2005).
- [36] W. C. Bessler, Doctoral dissertation, Ball State University, 1969.
- [37] W. C. Bessler and J. J. Nisbet, The use of an electronic response system in teaching biology, *Sci. Educ.* **55**, 275 (1971).
- [38] J. D. Brown, An evaluation of the Spitz student response system in teaching a course in logical and mathematical concepts, *J. Exp. Educ.* **40**, 12 (1972).
- [39] J. J. Bapst, Doctoral dissertation, University of Washington, 1971.
- [40] J. Casanova, An instructional experiment in organic chemistry. The use of a student response system, *J. Chem. Educ.* **48**, 453 (1971).
- [41] D. Hestenes, M. Wells, and G. Swackhammer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- [42] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
- [43] A. P. Fagen, C. H. Crouch, and E. Mazur, Peer Instruction: Results from a range of classrooms, *Phys. Teach.* **40**, 206 (2002).
- [44] A. P. Fagen, Doctoral dissertation, Harvard University, 2003.
- [45] M. C. James, The effect of grading incentive on student discourse in peer instruction, *Am. J. Phys.* **74**, 689 (2006).
- [46] 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**, 37 (2008).
- [47] P. M. Len, Different reward structures to motivate student interaction with electronic response systems in astronomy, *Astron. Educ. Rev.* **5**(2), 5 (2006).
- [48] S. D. Willoughby and E. Gustafson, Technology talks: Clickers and grading incentive in the large lecture hall, *Am. J. Phys.* **77**, 180 (2009).
- [49] A. L. Abrahamson, *International Workshop, New trends in Physics Teaching*, Puebla, Mexico, <http://www.bedu.com/Publications/PueblaFinal2.html> (unpublished).
- [50] N. Cue, *Proceedings of the First Quality in Teaching and Learning Conference*, Hong Kong International Trade and Exhibition Center (HITEC), Hong Kong SAR, China, December 10–12, 1998, <http://celt.ust.hk/ideas/prs/pdf/Nelsoncue.pdf> (unpublished).
- [51] R. J. Dufresne, W. Gerace, W. Leonard, J. Mestre, and L. Wenk, Classtalk: A classroom communication system for ac-

- tive learning, *J. Comput. High. Educ.* **7**, 3 (1996).
- [52] J. A. Shapiro, Electronic student response found feasible in large science lecture hall, *J. Coll. Sci. Teach.* **26**, 408 (1997).
- [53] D. P. Garg, Conference on Computers in the Undergraduate Curricula, Fort Worth, TX. (ERIC Document Reproduction Service No. ED111355), <http://www.eric.ed.gov/PDFS/ED111355.pdf> (unpublished).
- [54] R. Littauer, Instructional Implications of a low-cost electronic system, *Educational Technology* **12**, 69 (1972).
- [55] J. Barnett, Implementation of personal response units in very large lecture classes: Student perceptions, *Australasian Journal of Educational Technology* **22**, 474 (2006).
- [56] C. R. Graham, T. R. Tripp, L. Seawright, and G. L. Joeckel, Empowering or compelling reluctant participators using audience response systems, *Active Learn. High. Educ.* **8**, 233 (2007).
- [57] A. Hoekstra, Vibrant student voices: exploring the effects of the use of clickers in large college courses, *Learn. Media Technol.* **33**, 329 (2008).
- [58] R. Kaleta and T. Joosten, Student response systems: A University of Wisconsin system study of clickers, *EDUCAUSE Center for Applied Research Research Bulletin* **10**, 2 (2007).
- [59] C. Keller, N. Finkelstein, K. Perkins, S. Pollock, C. Turpen, and M. Dubson, "Research-based practices for effective clicker use." Proceedings of the 2007 Physics Education Research Conference, *AIP Conf. Proc.* **951**, 128 (2007).
- [60] E. L. MacGeorge, S. R. Homan, J. B. Dunning, D. Elmore, G. D. Bodie, E. Evans, S. Khichadia, S. Lichti, B. Feng, and B. Geddes, Student Evaluation of audience response technology in large lecture classes, *Educ. Technol. Res. Dev.* **56**, 125 (2008).
- [61] A. R. Trees and M. H. Jackson, The learning environment in clicker classrooms: Student processes of learning and involvement in large university-level courses using student response systems, *Learn. Media Technol.* **32**, 21 (2007).
- [62] W. R. Penuel, L. Abrahamson, and J. Roschelle, in *Audience Response Systems in Higher Education: Applications and Cases*, edited by D. A. Banks (Information Science Publishing, Hershey, PA, 2006).
- [63] S. Davis, Observations in classrooms using a network of hand-held devices, *J. Comput. Assist. Learn.* **19**, 298 (2003).
- [64] J. Maxwell, *Qualitative Research Design* (Sage Publications, London, 2005).
- [65] F. Erickson, in *Handbook of Research on Teaching*, 3rd ed., edited by M. C. Wittrock (Macmillan, New York, 1986) pp. 119–161.
- [66] B. Rogoff, *The Cultural Nature of Human Development* (Oxford University Press, New York, NY, 2003), pp. 3–36.
- [67] It is interesting to note that this methodological choice stems from the fact that many classrooms impose a dual-culture on students where exam days have a different set of norms than other course meetings. Often assessment practices may not align with the values reflected in other classroom activities.
- [68] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009).
- [69] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. (John Wiley & Sons, New York, 1999).
- [70] A Chi-squared Test is used to determine if we are justified in rejecting the null hypothesis which states that the row variables (instructor) and the column variables (possible answer options) are unrelated.
- [71] A Kruskal-Wallis Test is used to determine if we are justified in rejecting the null hypothesis which states all of the populations are identical, versus the alternative that some of the populations tend to furnish greater observed values than other populations. This is analogous to an ANOVA for continuous data.
- [72] P. Black and D. Wiliam, Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan* **80**, 139 (1998).
- [73] D. Wiliam and P. Black, Meanings and Consequences: A basis for distinguishing formative and summative functions of assessment, *Br. Educ. Res. J.* **22**, 537 (1996).
- [74] J. Hattie and H. Timperley, The power of feedback, *Rev. Educ. Res.* **77**, 81 (2007).
- [75] We note that Red has slightly higher end-of-term course evaluation ratings than Yellow and Green. However, in our larger set of student survey responses across six different courses, higher instructor ratings do not uniformly correlate with more positive student responses on PI survey questions. For this reason, we do not believe that "liking" Professor Red accounts for differences in students' responses to PI survey questions.
- [76] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn, Expanded Edition* (National Academy Press, Washington, DC, 2000).
- [77] P. S. Hutchison, Doctoral dissertation, University of Maryland, 2008.
- [78] R. Scherr, A. Elby, and T. Bing, *American Association of Physics Teachers (AAPT) National Winter Meeting*, *Rev. Educ. Res.*(2009).
- [79] In Ref. [1], Purple's PI practices are described from a physics course for nonscience majors, but observations of Purple in his Phys 2 course were found to be similar.