

Transforming the junior level: Outcomes from instruction and research in E&M

Stephanie V. Chasteen,¹ Steven J. Pollock,¹ Rachel E. Pepper,² and Katherine K. Perkins¹

¹Science Education Initiative, Physics Department, University of Colorado at Boulder, UCB 390, Boulder, Colorado 80301, USA

²Department of Integrative Biology and Department of Civil and Environmental Engineering,

University of California, Berkeley, California 94720, USA

(Received 28 December 2011; published 29 August 2012)

Over the course of four years, we have researched and transformed a key course in the career of an undergraduate physics major—junior-level electricity and magnetism. With the aim of educating our majors based on a more complete understanding of the cognitive and conceptual challenges of upper-division courses, we used principles of active engagement and learning theory to develop course materials and conceptual assessments. Our research results from student and faculty interviews and observations also informed our approach. We present several measures of the outcomes of this work at the University of Colorado at Boulder and external institutions. Students in the transformed courses achieved higher learning gains compared to those in the traditionally taught courses, particularly in the areas of conceptual understanding and ability to articulate their reasoning about a problem. The course transformations appear to close a gender gap, improving female students' scores on conceptual and traditional assessments so that they are more similar to those of male students. Students enthusiastically support the transformations, and indicate that several course elements provide useful scaffolding in conceptual understanding, as well as physicists' "habits of mind" such as problem-solving approaches and work habits. Despite these positive outcomes, student conceptual learning gains do not fully meet faculty expectations, suggesting that it is valuable to further investigate how the content and skills indicative of "thinking like a physicist" can be most usefully taught at the upper division.

DOI: 10.1103/PhysRevSTPER.8.020107

PACS numbers: 01.40.Di, 01.40.Fk, 01.40.G-, 01.40.gb

I. INTRODUCTION

In studying upper-division courses (primarily classical mechanics, thermodynamics, and quantum mechanics) [1,2], researchers have started to document a variety of student conceptual difficulties, allowing researchers and instructors to consider effective pedagogical strategies. While physics education researchers have examined the teaching of electricity and magnetism (E&M) at the introductory level in some detail [3,4], research on upper-division E&M is still fairly limited [4–6] and documentation of experimental course approaches are rare [7]. However, preliminary experimental evidence suggests that upper-division students are not achieving the level of skill that is expected of them. Some work has suggested that upper-division cognitive skills are more challenging than instructors may realize [5].

For example, Singh [4] has found that students enrolled in a junior-level E&M course do not perform better than introductory students on a conceptual E&M survey on Gauss's law, and do not improve their understanding of these introductory concepts over the course of junior E&M. Similarly, a longitudinal study by Pollock [2] has

determined that juniors' scores on the Basic Electricity and Magnetism Assessment (BEMA) do not improve from the freshman to the junior year, or over the course of the first semester of upper-division E&M (E&MI). Bing and Redish [8], studying upper-division students in a variety of courses, found that these students often do not use mathematics efficiently as a tool to solve problems, "getting stuck" using a limited set of tools. These results suggest that, at the least, junior-level courses are not fixing introductory student deficiencies and are not helping students progress on these introductory-level topics.

What about graduate student preparation? An early paper by Breitenberger [9] found that graduate students are often deficient in mathematical techniques at which faculty expect full proficiency, and do not see mathematics as a tool for constructing understanding. A recent study on quantum mechanics by Zhu and Singh [10] found that undergraduate and graduate students shared many conceptual difficulties. Thus, undergraduate courses may not be preparing students for graduate level study in physics.

The work reported here provides further confirmation of the need for new instructional approaches at the upper level: Students in traditional lecture courses do not achieve the level of material mastery expected by our faculty. These research results are consistent with the broad consensus of our faculty at the University of Colorado at Boulder (CU) that junior-level students were not learning what faculty wanted them to learn. In informal "brown

Published by the American Physical Society under the terms of the [Creative Commons Attribution 3.0 License](#). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

bag” discussions before the beginning of the current project, faculty indicated discontent with student progress in these upper-division courses. An additional drive to investigate upper-division student learning came from faculty members using interactive engagement techniques in lower-division courses, who felt the lack of such resources at the junior level. Many faculty were also already using classroom response systems (“clickers”) at the upper division [11], with positive student response, and we recognized that there was a need for research both (a) to determine whether these techniques were still appropriate at the junior level of instruction or needed modification for this student audience and (b) to develop research-based materials for use at the upper level.

Together, these motivations provided the impetus for the current project—to determine consensus learning goals for junior E&MI, research student learning and difficulties at this level, develop instructional materials, and test the impact of these materials. This is not a fully developed curriculum based on carefully documented student difficulties, such as the extensively researched Tutorials in Introductory Physics [12]. This is a case study of a first-pass effort.

Some background information on the physics department at CU will provide context for these efforts. The department has used interactive engagement techniques at the lower division for many years, including the incorporation of the Washington Tutorials [13] and peer instruction [14] using personal response systems (clickers). These interactive course approaches have become normative for CU faculty, not just those involved in physics education research (PER)—over half the faculty members have used clickers in a course, many in upper-division courses [11]. Thus, the department could be seen as receptive, on average, to interactive and research-based teaching techniques. However, a few faculty also expressed some reservations about applying such approaches to the junior level, citing concerns about “coddling” these upper-division students, or worries about spending more time on conceptual understanding at the expense of important computational skills.

Faculty prioritized the transformation of the first semester of our two-semester junior-level sequence in electromagnetism (referred to here as PHYS301). This course covers electro- and magnetostatics in vacuum and in matter (Chaps. 1–6 of the text by Griffiths [15]) and is typically taken in the fall of the junior year. Prerequisites for PHYS301 are the three-semester introductory physics sequence and the combined mechanics and math methods course. Typically 30–50 students enroll in a given semester of PHYS301.

We followed the course transformation model [16,17] developed by the Science Education Initiative (SEI) [18] and used in other courses at CU and our sister institution at the University of British Columbia. Our earlier publications [6] report on specific student difficulties with

the content in PHYS301. The current publication outlines the outcomes from this case-study project:

- (1) What do faculty think students should achieve in this course (the learning goals)?
- (2) What instructional methods were developed to address these learning goals?
- (3) How well did these new materials help students achieve these learning goals?

While upper-division E&M typically focuses on formalism and theory, by derivation of formulas and presentation of common worked problems, our aim was to also provide a more substantial focus on phenomena and concepts. We adapted pedagogical techniques used successfully at the lower division, such as peer instruction and tutorials. The final course package included (a) faculty-developed learning goals, (b) documented student difficulties to help drive instruction, (c) clicker questions, (d) optional out-of-class conceptual tutorials, (e) optional out-of-class group homework help sessions, (f) modified homework problems, and (g) a conceptual assessment to measure student learning. Thus, the transformed course results in increased time on task for students engaging in these optional activities. The transformed course has been taught five times at the University of Colorado at Boulder [19], and our materials have been used in courses in at least three external institutions.

This paper is intended to document the process of our course transformation, as well as to provide evidence for the utility of alternative instructional approaches at the junior level. We found that students enjoyed the transformed courses, and the changes had positive effects on student learning. However, there is still much to explore as teachers and researchers in this area, as a substantial gap remains between our goals for our physics students and what they are able to do at this point in their career. In this paper we characterize this gap and suggest fruitful areas for future research. A companion publication (Ref. [20]) provides a briefer overview of the project and its outcomes, aimed at instructors.

II. OBJECTIVES OF THE TRANSFORMATION

Our course transformation efforts were driven, primarily, by the desire to better prepare our upper-division physics students for the sophisticated mathematical, analytical, and conceptual skills required of a physicist. We first needed to operationally define what “thinking like a physicist” meant in the context of a junior-level physics student. What do our faculty think that this course is about? Here we describe the process and outcomes of our determination of course objectives.

A. Time line of transformation

To help provide a context for the course transformation activities, following is a brief time line. We do not identify the particular semesters that any particular course was taught for privacy reasons.

Year one

Semester 1. Observe traditional course. Develop consensus learning goals. Begin development of conceptual assessment [the Colorado Upper-Division Electrostatics (CUE) Assessment].

Semester 2. Material development. First iteration of transformed course. Alumni survey. Continue development of CUE (continues through year four).

Year two

Continue material development, and iterate existing materials. Second and third iterations of transformed course.

Year three

Fourth and fifth iterations of transformed course.

Year four

Traditionally taught course [21].

B. Faculty learning goals

To identify goals of the course transformation, we met biweekly with a faculty working group (typically 8–10 faculty out of a core group of 15) during the semester prior to the transformation. This ensured that the project team was guided by consensus faculty values, rather than the development team creating materials that were disconnected from the opinions of the department. This faculty working group discussed learning goals for the course and assisted in generating questions and providing feedback on a conceptual diagnostic for the course. Details on the process of defining learning goals are described elsewhere [22].

The broad course-scale learning goals for the course are abbreviated as follows (the complete document is included in the Appendix and is available online [23]):

- (1) Math-physics connection
- (2) Visualization of problems
- (3) Organized knowledge
- (4) Communication
- (5) Problem-solving techniques
- (6) Problem-solving strategies
- (7) Expectation and checking of solution
- (8) Intellectual maturity
- (9) Coherency of Maxwell's equations
- (10) Build on earlier material

These learning goals developed for the course have proven to be a valuable document for faculty at CU and elsewhere [22] wishing to define what upper-division courses in general should be accomplishing. Operationalization of these goals is provided in Sec. IID, as well as in the Appendix.

C. Alumni survey

With these learning goals in hand, we surveyed our alumni [24] to determine how well the course in the past had met these objectives, and how these graduates—with the benefit of hindsight—saw PHYS301 within the context of their undergraduate career. A total of 369 physics, engineering physics, and astrophysics majors, graduating

in the years from 1997 to 2009 were sent an electronic survey via Email. They were entered into a raffle for a \$20 gift certificate for their participation. A total of 139 surveys (38%) were returned. Respondents were primarily physics majors (76%), followed by engineering physics (16%) and astrophysics (15%; respondents could have more than one major), most having graduated in the years from 2003 to 2009. They were asked about their current career paths and how the upper-division physics courses did or did not meet their needs within their current positions, as well as being asked for general comments regarding those courses.

Alumni reported that the most useful part of PHYS301 was the problem-solving techniques whereas the math and physics content of the course were of varying utility. This matches the faculty perception that this course is central to students' development as expert problem solvers. We see key differences between those alumni who did not attend graduate school in any subject (33% of sample) and those who did pursue graduate studies, such that the course content is perceived as less useful or retained by the students who did not continue to graduate school. Considering that those who do not continue to graduate studies represent a third of the population of the course, our survey suggests that the current format of the course was not fully serving a large portion of our audience. Results of the alumni attitude survey are included in the Appendix.

Overall, what sticks in alumnis' minds is the challenge and rigor of the course. Alumni specifically expressed their appreciation of challenging homework problems, a rigorous course with high expectations, and instructors who respect students and devote personal attention to them. We asked students if they felt that they had “matured as a physicist or as a student” (learning goal no. 8) during the course, and 73% of respondents indicated that they had, and that the course is a milestone in their identity as a physics student: “I learned to sit back and examine a problem before diving into the math,” “I became more mathematically sophisticated,” “This was the first really challenging physics course I took and it gave me a better understanding of what physics was really all about.”

In this, and other upper-division courses, however, they often expressed a desire for (a) more real-world examples or the addition of labs and demos, (b) more connections between the physics and the math, and (c) fewer derivations and a focus on concepts. Several of these themes were echoed by students on a survey after a traditionally taught semester of PHYS301. We explicitly attempted to address these complaints in the transformed course, by adding these interactive, conceptual elements to the existing outline of the traditionally taught course (rather than creating an entirely new course approach).

D. Central themes

From the comments of the alumni, and the learning goals developed by faculty, we identify three central ideas that form the fabric of the course:

- (1) Mathematical sophistication
- (2) Problem-solving expertise
- (3) Developing as a physicist

These themes have some precedence in the physics education research literature, which contains bodies of work on the use of mathematics in physics [25], problem solving [26], and metacognition and expertlike approaches to problem solving and learning [27]. These categorizations formed much of the underlying focus in our course transformation. We provide working definitions of these themes, below, in order to provide a more complete answer to the question of what PHYS301 is about.

1. Mathematical sophistication

- (a) *Translating physics to mathematics* (e.g., in setting up a problem, including visualization of the problem and its parameters; learning goals no. 1 and 2)
- (b) *Conceptual and physical understanding of mathematics* (e.g., in making sense of equations and problem solutions; learning goal no. 1).

The mathematical tools in PHYS301 (primarily differential equations and vector calculus) allow students to build upon their conceptual understanding from introductory physics (learning goals no. 1 and no. 10). This course represents one of the first opportunities for students to apply these newly acquired mathematical tools to physical situations: “This was one of the first times that material I had previously learned came back in a much more mathematically rigorous way, and I saw how you can add depth to a problem,” said one graduate. Many alumni mentioned particular techniques (e.g., Laplace’s equation) in their comments about the course, indicated that the course “expanded my mathematical toolkit,” and one specifically claimed that, “I became more mathematically sophisticated.” Our observations indicate that students typically struggle to apply these abstract mathematical tools to physical systems (i.e., analyzing a problem to choose the appropriate tool or limits of integration), and to make physical meaning of their answer [6].

2. Problem-solving expertise

- (a) *Use of expert problem-solving tools* (e.g., approximations and limits; learning goal no. 5)
- (b) *Expert problem analysis* (e.g., recognizing key parameters, and combining knowledge from multiple sources to devise and execute a strategy; learning goals no. 3, no. 6, and no. 9).

An expert is facile both in the use of the tools of the trade and knowing when to apply them. Thus, not only should students be able to use mathematical techniques, but they need to be able to use them in problem-solving situations.

Problem-solving skills were rated as one of the most useful outcomes of the course by alumni, and one of the defining characteristics of the course.

3. Developing as a physicist

- (a) *Independence and discipline* (e.g., taking responsibility for learning, and working hard; learning goal no. 8). *Metacognition* (e.g., reflecting on one’s own learning and learning how to learn; learning goal no. 8). *Expertlike approaches to problems* (e.g., planning an approach, having the confidence to see a problem through to the end, and checking one’s work: learning goals no. 6 and no. 7). *Ability to articulate one’s reasoning* (e.g., explaining one’s approach and method of solving a problem: learning goal no. 4).

This has been the most difficult category to define, particularly in a measurable way. One aspect of intellectual maturity is the ability to “step up to the plate” and manage the increased rigor of the course. Alumni indicated that this course required more effort and self-discipline than previous courses, due primarily to the longer and more complicated problem-solving homework sets: “This class is a step up,” said one, and it shows you that “hard work can pay off” remarked another. Faculty in the working group indicated that they expect this increased responsibility on the part of students. Students noticed this (“instructors hold your hand a little less each time”) and expressed particular appreciation for instructors who were tough, but respected them. These ideas form the basis of the *independence and discipline* aspect of maturity.

Physicists, as well as being experts in the content area of physics, are also expert learners. Students must also be able to reflect on their own learning in order to “learn how to learn.” Many alumni indicated that it was in this course that they first recognized the value of working with students in a group to solve problems. Our full learning goals (see the Appendix) indicate that intellectual maturity is evidenced by “asking sophisticated, specific questions; being able to articulate where in a problem they experienced difficulty; and then take action to move beyond that difficulty.” Alumni and faculty agree that this is an important aspect of the course—the *metacognition* aspect of maturity.

Lastly, the problems in PHYS301 are no longer simple problems that can be solved through application of a formula in a few lines. Alumni and faculty both indicated that learning to do these long, multistep problems is a key aspect of the course (learning goal no. 6): “This was the class where I learned to interpret and understand complicated, multipage physics problems,” and “E&M, more than any other undergraduate physics class, teaches you to sit down with a physics problem that can take pages to

solve and not wince. After that, you are not really intimidated by big hairy problems," explained alumni. The ability to solve such problems involves the confidence and perseverance to see the problem through to the end, which is related to students' technical skill and independence, as well as to a separate category that we call *expert-approaches to problems*. These include reflective techniques such as planning a solution and checking one's work [28].

By the end of this process of faculty discussion and alumni surveys, our team had a much better grasp of the role of this course in the undergraduate curriculum and the key skills that students should acquire. This drove the development of the course materials described below.

III. THE TRANSFORMED COURSE

A. Preparation for the transformation

In order to create materials that properly targeted student difficulties with the learning goals, we investigated student learning prior to and during the first semester of the transformations. First, one of the authors (S. V. C.), a postdoctoral science teaching fellow, reviewed relevant research. She also observed or facilitated several semesters of lecture, help sessions, and tutorials, including a traditionally taught course prior to the transformation. In addition to these observational measures and accompanying field notes, students were interviewed using think-aloud protocols as they solved problems, answered conceptual questions, or sketched concept maps connecting central ideas in the course: A total of 20 students were interviewed one or more times during the three-year lifetime of the project, for a total of 50 interviews. A set of 6 students of varying ability were interviewed multiple times during the first iteration of the transformed course. These field notes and interviews provided valuable insight into key areas of difficulty, driving material development.

The first-pass course materials were developed primarily by two authors (S. V. C. and S. J. P.) with assistance from two undergraduate students focused on teaching and learning [29]. Materials were modified and improved by later instructors and another coauthor (R. E. P.). Interviews were conducted with each instructor during and after each iteration of the transformed course, often resulting in revision of course materials for ease of use and providing insight into the process of teaching the transformed course. All course materials are freely available and archived online [23].

B. Elements of the transformed course

Our transformed course was designed to implicitly or explicitly address the core themes—mathematical sophistication, problem-solving expertise, and developing as a physicist—as well as to target commonly observed student difficulties. Our pedagogical approach was also informed

by our previous experience with interactive engagement in large-lecture introductory courses, and incorporated a variety of interactive and research-based techniques and pedagogical materials. Unlike other models that have switched completely to small group work [7] our approach was largely traditional, with lectures interspersed with interactive techniques. Below we list the items that made up the transformed course approach.

Student difficulties.—A detailed list of observed student difficulties with conceptual content and mathematical techniques was compiled and organized by chapter and topic. This was provided to instructors within the course archive and intended to guide instruction.

Explicit learning goals.—The learning goals developed with the faculty working group were provided to future instructors, who were encouraged to share them with students as well as to use them to guide instruction.

Lectures were fairly traditional, with an emphasis on blackboard work, but included a variety of interactive elements, including clickers, spontaneous questioning, simulations, and student work on small whiteboards.

Clicker questions with peer instruction.—About 2–4 clicker questions were used in each 50-minute period. Clicker questions helped students make meaning of mathematics, expanded upon material that had just been covered, or led into the next topic, serving to scaffold student understanding and reduce cognitive load.

Homework assignments were redesigned to provide an intellectual challenge required to meet the course goals, and explicitly required students to connect abstract problems to real-world situations or physical contexts, articulate what they expected the answer to be, make sense of their answer, and draw on common physicists' tools such as approximations, expansions, and estimations. We also scaffolded their metacognitive strategies by explicitly asking students to think about whether the answer made sense or to check their work (and grading on their ability to do so). Many of these goals were achieved with minimal effort by adding a sense-making component to more-traditional problems. Typical sense-making components used were: sketch, plot, what is the limit, approximate, estimate, describe your mathematical solution in words, interpret, what are the units, is this consistent with . . . , check your answer, explain your reasoning, and use two methods to solve and compare.

Students could attend *optional help sessions* (typically replacing standard office hours), offered twice a week for 1–2 hours, in which they worked on the homework assignments in groups with an instructor present to facilitate; in some semesters students were provided with a large whiteboard as a public space for discussion. Working with other students not only helped with homework, but also assisted with students' metacognitive strategies and communication skills as they could watch other students (and the instructor) model problem-solving strategies and were required to articulate their reasoning to one another.

A series of *optional weekly tutorials* was developed and refined over two years, with the later addition of tutorial pretests. These were designed to reinforce topics presented in lecture, expand on these topics, and prepare students for the upcoming homework. Student attendance at tutorials was optional ($\sim 40\%$ of the class attended on average), and students worked in groups of 3–5 to complete a conceptually focused worksheet on the material. Tutorials were facilitated by the instructor, an undergraduate learning assistant, and, in some semesters, a science teaching fellow. Tutorials were often designed to provide students with a deeper conceptual understanding of material related to the homework, and/or to scaffold their development as expert problem solvers by breaking problems up into smaller pieces, or showing multiple ways to attack a problem. As with the help sessions, tutorials were also geared to help students with metacognitive strategies and communication skills, as well as to allow the instructor to model effective strategies.

More detail on the course components and their development is found in prior work [6,16]. The fidelity of the course implementation varied by instructor, and there is some evidence that the more faithfully instructors attended to common student ideas and used interactive techniques, the higher student satisfaction with the course and the higher learning gains on the conceptual assessment (CUE) [19].

IV. ASSESSMENT

A. Methods

In order to investigate effects on student learning, we compared five iterations of the transformed PHYS301 which used physics education research (PER) techniques (courses PER-A through PER-E) and two courses taught using more standard (STND) lecture-based techniques (courses STND-A and STND-B). The sequence of these courses has been randomized for privacy—i.e., PER-A was *not* necessarily the first semester of the transformation. The first two semesters of the new course were taught by PER faculty individually or in a team with a non-PER faculty, and subsequent semesters were taught by non-PER faculty individually. The traditional lectures (STND-A and STND-B) were taught with a standard lecture format—either by the instructor’s choice or because our materials had not been developed. STND-B used clickers irregularly, primarily for review rather than for peer instruction; otherwise, these lecture classes did not use clickers or techniques other than standard lecture and we do not consider them transformed.

We used a variety of data measures to compare courses (summarized in the Appendix). These seven courses are similar in terms of a variety of measures (i.e., differences are statistically insignificant as determined by the chi-square test), including students’ grade-point average (GPA) in prior courses, course demographics, course rating, and attendance. Students also completed the Basic Electricity and Magnetism Assessment (BEMA) [30]—measuring conceptual understanding of introductory-level

topics in E&M—at the end of their freshman physics sequence, and (in some cases) at the end of PHYS301. Freshman BEMA scores were compared to the average across courses (60%) by t-test. Courses were overall similar on average freshman BEMA scores. Course enrollment varies significantly from a low of 22 (PER-C) to a high of 56 (PER-D), due to certain courses being offered in the “off” semester.

Some notable differences between courses (significant and insignificant) are listed below.

PER-A has higher lecture attendance and lower percentage of physics majors ($p < 0.05$) and an accompanying higher percentage of engineering physics majors.

PER-C has higher incoming BEMA scores (69%; $p < 0.05$) and higher lecture attendance.

PER-D has fewer females.

PER-E has lower incoming BEMA score (55%; $p < 0.05$) and fewer females.

STND-B has lower incoming BEMA scores (55%; $p < 0.05$), lower incoming cumulative GPA (2.9 compared to 3.1; $p < 0.05$), and fewer females.

Thus, overall, we conclude that students are similar across CU courses except that students in PER-E and STND-B may be less prepared and those in PER-C may be better prepared. Such demographic data were not available for courses outside of CU.

A conceptual diagnostic—the CUE [31,32]—was developed to assess students’ conceptual understanding before and after the course and provide an independent measure of student learning. In order to assess students’ ability on more traditional measures, five exam problems taken primarily from a standard lecture-based course were given in common across two or three semesters of the course and graded on a common rubric. Student attitudes were assessed through opinion surveys at the end of PHYS301.

B. Conceptual assessment—CUE

While some of the skills identified in the course learning goals are tested in traditional exams (such as setting up and solving integrals), many of these techniques and their conceptual underpinnings are not. To address this gap and assess the relative success of these transformations, as well as to document student difficulties, we developed a post-test for the course—the CUE. This assessment was developed in large part with the assistance and collaboration of the faculty working group. The CUE is a 17-question open-ended conceptual test (15 electrostatics and 2 magnetostatics questions) testing students’ ability to choose a problem-solving method and defend that choice, sketch electric field patterns, graph electric field strength and potentials, and explain the physics and mathematics concepts which underlie steps in common problems. The CUE has been found to be a valid and reliable measure, and inter-rater reliability using the grading rubric is high. More

TABLE I. Courses included in the CUE results. Fall 2009 acceptance rate for each university obtained from Ref. [35] with the exceptions of Non-CU-STND1, Non-CU-STND2, and Non-CU-STND6—a Canadian institution; that admission rate represents the number of students admitted to their first-choice program (compared to the number of applicants to first-choice programs), offering a measure of selectivity. “No. physics degrees” represents the number of undergraduate degrees awarded by the department, per year. The remaining course data were obtained directly from instructors. “Semesters E&M” indicates whether the course in question represents a single semester encompassing both electrostatics and electrodynamics (“One”) or whether it was the first semester of a two-semester E&M sequence (“Two”). “Student N ” represents the number of students who took the CUE and did not drop the course. “Regularly” is defined as several times a week for clickers, and at least once a week for tutorials. “Some” clicker use is defined as using clickers, but less than once per week.

Course	Acceptance rate; no. physics degrees/yr	Institution type	Semesters E&M	Student N	Used clickers regularly?	Used tutorials regularly?	Times taught previous	Notes
CU courses (STND-A &B, PER-A-E)	84%; 50+	University	Two	41, 41, 48, 37, 22, 56, 46	Y (PER only)	Y (PER only)	2; 0; 0; 2; 0; 0; 0	Full details in Appendix
Non-CU courses								
Non-CU-PER1	57%; 1–5	Private 4-year college.	One	5	Y	Y	0	
Non-CU-PER2	31%; 6–20. Ranked top 20 liberal arts colleges.	Private 4-year college	Two	12	Some	Y	0	
Non-CU-PER3	49%; 6–20	Public university	Two	14	Y	Y	0	
Non-CU-STND1	38%; 50+	Public university	One	48	N	N	3	Engineering physics course
Non-CU-STND2	38%; 50+	Public university	One	35	N	N	1	
Non-CU-STND3	71%; 6–20	Public university	Two	6	Y	N	9	Used colored cards for clickers
Non-CU-STND4	70%; 6–20	Private 4-year college	One	18	Some	N	11	Used out-of-class group work a few times
Non-CU-STND5	89%; 1–5	Public university	One	5	Some	N	6	Used CU clicker questions to prompt discussion
Non-CU-STND6	38%; 50+	Public university	One	138	Some	N	2	Used clicker questions and some group work

details about the CUE and its development are given in [33,34].

The CUE was administered to upper-division E&M students for the seven semesters of the study at the University of Colorado at Boulder. Our points of comparison for the effectiveness of course transformations were the two courses taught with a “standard” lecture-based instruction (STND-A and STND-B). However, because we wished for a more robust comparison, and because we have evidence that STND-A was a less effective course than the norm for that instructor, we also administered the CUE at other institutions. The CUE was also given in nine courses at seven outside institutions (six traditionally taught and three using our materials).

Thus, in total we present data for 488 students in order to assess the impact of research-based course transformations on student learning. Courses are summarized in Table I in order to aid in data interpretation. A course is labeled “PER-based” or “PER” if it used at least two course elements (generally, clickers and tutorials) developed by CU.

CUE results for all courses are shown in Fig. 1. The exam was given at the end of the semester, in class, as part

of the course. This minimized selection effects and resulted in high response rates (75–100%), except for STND-A (63%) due to overall low student attendance at lecture. Because the CUE was modified from semester to semester, a total of three different versions were given across institutions. The post-test reported here is made of a subset of the complete CUE given in common across all exams (88 out of 118 points). A future publication [32] will include scores on the complete exam in its current form. A pretest was developed consisting of a portion of the CUE questions which a student entering PHYS301 could reasonably be expected to answer given a good understanding of introductory E&M. The pretest scores of students entering PHYS301 are similar to those of freshmen who have just completed introductory E&M, about 30%.

The CUE post-test scores of students in courses using PER-based instructional techniques are statistically significantly higher than all the courses using a standard lecture format (STND) at CU and elsewhere. Taking each student as a data point, the average CUE score is higher in PER courses ($57 \pm 1.3\%$) than in STND courses ($44 \pm 1.6\%$ $p < 0.001$). Taking each course as a data point, the same

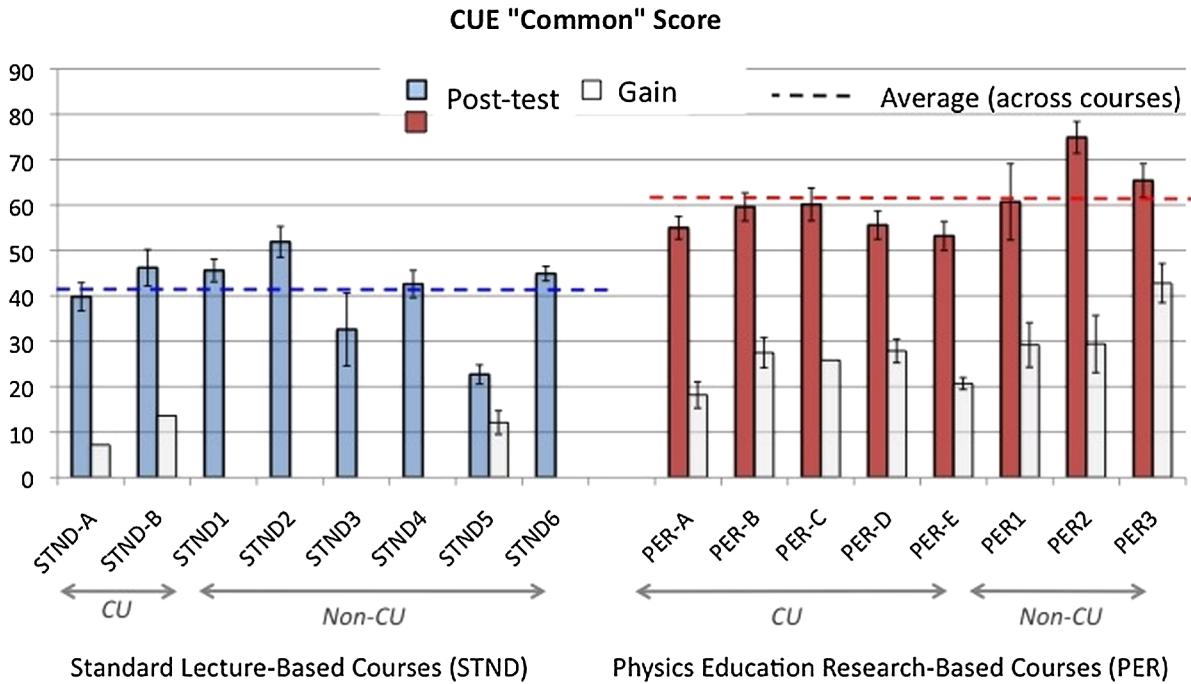


FIG. 1 (color online). CUE scores across institutions for $N = 488$ students. “Post-test” represents course average score (% correct) for the subset of CUE questions given in common across all exams (88 out of 118 possible points). “Gain” represents the course average for the difference between the pretest and the subset of the post-test which matches the pretest. Because of the lack of pretests for PER-C and STND-A and STND-B, pretest scores are estimated based on the stable pretest scores (33%) for other semesters of PHYS301 at CU. Error bars represent 1 standard error of the mean.

result holds ($61 \pm 4\%$ PER versus $40 \pm 4\%$ STND, $p < 0.001$). If the CUE were a graded exam, this would be comparable to a gain of two letter grades.

These results also held across students of different performance levels. We compared CUE results for students at CU with high or low grades in the course. While the “A” students in STND and PER courses did not differ on formative or summative measures of incoming GPA, BEMA scores, or exam scores, the “A” students in four PER-based courses (PER-A through D) had significantly higher CUE scores ($70 \pm 1.6\%$, $N = 58$) than did the “A” students in STND-A ($48 \pm 4.8\%$, $N = 12$, $p < 0.001$). The same was true for the lower-performing students: While students receiving C’s and D’s in STND and PER courses did not differ on incoming measures of preparation, those in the PER courses performed better on the CUE ($45 \pm 1.7\%$, $N = 54$) than did those in the STND-A course ($30 \pm 4.9\%$, $N = 7$, $p < 0.01$). This is not due to a predominance of higher course grades in the PER courses; average course grade in STND and PER courses is not significantly different. So, the course transformations appear to have positively affected both the top and the bottom performing portions of the student body.

Overall, these results suggest that the interactive techniques were consistently successful, over five semesters at CU and at three external institutions, in improving students’ facility with the concepts and problem-solving methods of junior E&M. As the CUE was developed based

on the faculty-developed learning goals, this suggests that we achieved some measure of success in our aim of supporting the cognitive skills of developing physicists. Examination of the demographics of individual courses at CU (see the Appendix) and elsewhere (Table I) shows that these results cannot be easily explained by factors related to the students or instructors, such as incoming GPA, number of students enrolled, type of institution, incoming score on the introductory conceptual assessment (BEMA), or instructor experience. Indeed, some of the highest scores on the CUE occur in classes where the instructor had no prior experience teaching the course. The robustness of these results over time, across instructors, and across institutions also suggests that the course transformation effects can be sustained from instructor to instructor and across institutions [19].

Gains are useful because they allow us to compare students with different backgrounds. Learning gains from pretest to post-test are calculated by taking the difference between student scores on the pretest and the same seven question subset of the post-test. Thus, it is a matched, absolute learning gain on a subset of the exam, and not equivalent to the total post-test minus the pretest score. We note that pretest scores for incoming juniors are very similar to the scores for freshmen who have just completed the introductory course. We chose to use the absolute gain, rather than the normalized gain $\langle g \rangle$ because pretest scores are quite similar across time and narrowly distributed.

Some courses do not have a pretest score, because either the pretest had not yet been developed or the instructor chose not to administer it. For these courses, the pretest score is estimated based on the stable pretest score across other courses. We suspect that preexposure to the pretest does not affect performance on the post-test. Only some students recall the pretest: In a limited data set of five students who took the CUE in validation interviews, three out of the five recalled having seen the questions previously on the post-test. Student scores on the pretest are universally and consistently low, and answers to the pretest are not discussed with the students. PER courses with or without the pretest show similar post-test scores, and one course (STND5) shows low CUE scores even with the use of the pretest.

Gains on the CUE are similar across PER courses (about $30 \pm 2.5\%$), representing a rough doubling of the score from pretest to post-test. By comparison, the estimated gain for STND-A and STND-B were only 7% and 14%. The learning gains on the CUE are significantly correlated with measures of student preparation: BEMA scores post-PHYS102 and students' incoming GPAs ($p < 0.01$, Pearson correlation coefficient r values range from 0.2 to 0.4). Additionally, gains for the "A" students (in PER-A through PER-D, $29 \pm 2\%$) are higher than those for the "C" and "D" students ($15 \pm 2\%$; $p < 0.001$). Thus, it appears that the course transformations are most helpful for the students who are better prepared. This is perhaps not surprising—though the course transformations help both the bottom and top tier of students, it is easier for our techniques to reach those who are better prepared to learn.

C. Traditional assessments

The CUE directly targets some of the conceptual skills identified by faculty on the learning goals. However, the CUE does not assess student facility with many of the essential elements of the physicists' toolkit, such as integrations, approximations, and vector calculus (learning goal no. 5). Traditional exams in upper-division courses are more suited to assessing this goal, as they typically assess students on their ability to solve long, mathematical, abstract problems.

Thus, we gave several exam questions in common between a standard course (STND-A) and two semesters of the transformed course (PER-C and PER-D). These questions were graded by a single grader blind to the conditions of the study on a common, detailed rubric, but scores were not validated through inter-rater reliability. Thus, these scores are shown for general comparative purposes only.

Five questions were given in common to students in the three courses (with the exception of Q4, not given to STND-A, and Q5, not given to PER-D). All questions were taken from exams in a traditionally taught course (STND-A), except for Q1, which was written by PER instructors. The wording below does not represent the

complete wording given to students, but is paraphrased for brevity.

Q1: Gauss's law. Imagine an evenly charged cube, surrounded by an imaginary cubical surface. Is Gauss's law true in this situation? Is it useful?

Q2: Dielectric cylinder [figure not shown]. Given a dielectric cylinder with a frozen-in polarization of $\vec{P} = k\hat{z}$, calculate (a) the bound charge and (b) the potential outside the cylinder, on-axis. (c) How would you *expect* the potential to drop off with r (and thus z) for large distances? Show that your answer from (b) gives that z dependence at large distances.

Q3: Separation of variables. A nonconducting spherical shell has a frozen-in potential, $V(r = R) = 3\cos^2(\theta) + 1$. Find the potential everywhere.

Q4: Direct integration of disk [figure not shown]. Given a flat uniform disk with surface charge $\sigma = \sigma_0 \cos(\phi/4)$, calculate (a) the total charge on the disk and (b) the voltage on the z axis.

Q5: B of a cylindrical wire [figure not shown]. A steady current I flows up an infinitely long cylindrical wire of radius R . The volume current density is given by $J = ks^2$. Determine the constant k and calculate B inside and outside the wire.

The results for all questions followed the same general trend (see Fig. 2). Student scores in PER-C were improved over those in STND-A, while those in PER-D matched the level of STND-A for most questions. We do have evidence that students in PER-C were better prepared than those in other courses (see Sec. IV A). It is thus difficult to determine whether students in the transformed course are more successful at traditional calculations than those in lecture-based courses. However, we can say that at least the course transformations did *not* have a statistically significant *negative* effect on the development of students' skills in these types of calculations. This helps to answer a concern voiced by some faculty about whether students in the PER-based courses would show poorer calculational skills due to the increased focus on conceptual understanding and sense making, and additional time spent on clicker questions at the expense of worked-example problems.

Upon closer examination of these results, however, we see evidence that the course transformation *did* effectively address two key learning goals that it was designed to meet—namely, students' ability to (a) articulate their reasoning about an answer and (b) identify the proper method for solving a particular problem. While all exam problems included a calculational or problem-solving aspect, some also required students to provide reasoning about their answer (learning goal no. 4) or use of limits (learning goal no. 7). The exam scoring rubric distinguished elements of each problem as "calculational," "reasoning," and "articulating the expected limiting behavior" by one of the authors (S. V. C.) familiar with the problems and the

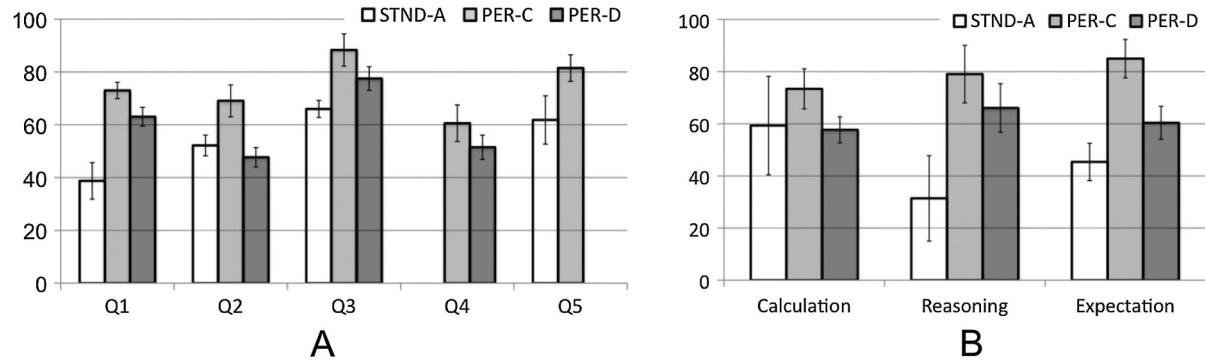


FIG. 2. Traditional exam results. (A) Traditional exam questions: Average percent score on five traditional exam questions across three semesters. Error bars represent 1 standard error of the mean. (B) Traditional exam strategies: Portions of questions grouped by problem-solving strategy: calculation and correctness (i.e., Q1 correctness; Q2 calculation of bound charge V and limiting behavior; Q3 calculation of V), reasoning (Q2 explanations, Q3 reasoning), and expectation of limiting behavior (Q2 articulation of limiting behavior). SE for categories in (B) is obtained by adding the SEs for the individual questions in that category, in quadrature.

learning goals. Figure 2(b) shows that while students in all courses performed similarly on the calculational aspect of problems, the students in PER-based courses outperformed those in STND-A in their articulation of their reasoning and description of limiting behavior.

Thus, while the course transformations did not improve students' ability to perform raw calculations, they did assist with some of the sense-making aspects of problem solving.

D. Student attitudes

In addition to the learning outcomes reported above, we investigated the impact of the courses on student attitudes and satisfaction. Students were overall positive about the transformed PER courses. Here we report a variety of measures of student satisfaction. In Sec. IV E we will describe student feedback on specific elements of the course transformation (e.g., clickers or tutorials).

Students engaged more fully in the PER courses in some ways. End-of-term course evaluations ask students how many hours they spend on a course per week, including class time. Students in the PER courses consistently reported spending more time on the course per week than those in the STND courses in the study ("10–12 hours" versus "7–9 hours"). That is, with regard to the distribution of reported times, a larger fraction of students in the PER courses reported spending more time on the course than their STND course peers. The difference in reported times is significant; the average time reported by PER students is more than 1 standard deviation higher than the average reported across the department as a whole. These results also hold for time spent on homework: In end-of-term attitude surveys, students in PER courses consistently report spending more hours per week on homework (8 ± 1) than in STND-A (4). Using attendance on the day that course evaluations were administered, we also find that attendance improved slightly in the PER courses

(75%) compared to attendance in the three previous years of traditional courses (69%). Restricting our scope to the seven courses in the study, attendance on the day that the CUE was administered was higher in PER courses (82%) than in STND courses (59%). However, an analysis of attendance rates (using contingency tables) on these two lecture dates did not detect any significant association between attendance and whether the course was transformed. The trends reported appear to favor the PER courses, but additional data are needed to detect any significant differences. Together, these results suggest that extra out-of-class sessions, additional reported time on homework, and possibly improved attendance resulted in greater time on task for students in the transformed course.

We also compared end-of-term student evaluation of faculty or course from the university (Faculty Course Questionnaire, or CU-FCQ) for the five semesters of PER-based courses with those from the previous 15 semesters of standard lecture-based iterations of PHYS301. Overall, the PER and STND-A courses were rated similarly, with slight increases in the rating of the PER courses in terms of instructor effectiveness (5.4 vs 5.0 out of 6) and challenge (5.4 vs 5.2 out of 6). Because of lack of access to the statistics underlying these university measures, significance was not tested.

Because the CU-FCQ course evaluations provide only a broad picture of the impact of the course, we developed an end-of-term survey about students' experiences in the course: What they liked, what they did not like, and what they found effective for their learning. Students were generally positive on measures of course satisfaction, such as interest in the subject, ability to do mathematical calculations, and a perception of a connection between various course elements. Five questions given across all five PER-based semesters are reported below.

- (1) I think E&M is an interesting subject (Agree: $84.9 \pm 5.8\%$) [36]

- (2) My intuition is useful in solving PHYS301-level E&M problems (Agree: $67.3 \pm 8.6\%$)
- (3) I am generally able to do the mathematical calculations required in PHYS301 (Agree: $77.8 \pm 7.5\%$)
- (4) More than half the time, even when I successfully calculate the answer to a problem in PHYS301, I don't understand the physics of the answer (Agree: $40.2. \pm 6.4\%$)
- (5) Connections between mathematical techniques and important physical concepts were made regularly in lectures and/or homework (Agree: $84.5 \pm 6.8\%$)

While these results are overall positive, they also indicate that there is more to be done: Question no. 2 indicates that students do not see their intuition as useful as we would have hoped, and question no. 4 shows that many students are still struggling to connect meaning to the results of calculations.

PER-A stands out as a notable exception. Student attitudes in this course were statistically significantly more negative than those in other courses on several questions (Q2: 59.8%; Q4: 68.3%; Q5: 77.2%), and students saw the course elements (such as lecture and clicker questions) as less well connected. Commentary from students and instructor suggested that (a) students did not completely buy-in to the course approach and (b) the instructor may not have been fully attentive to student difficulties at the junior level. The implications for sustainability of course innovations are discussed in more detail in another publication [19].

E. Impacts of different course elements

It would be particularly helpful to know which aspects of the transformation must be faithfully adhered to in order to achieve maximum student learning (for minimal instructor time and effort). However, it is very challenging to extract the effects of any single element of a multipronged course transformation. In this section, we provide available evidence as to the particular effects of clicker questions, homework, homework help sessions, and tutorials. Though the effects of most course elements are difficult to ascertain, we are able to demonstrate that attendance at tutorials has a positive effect on students' conceptual understanding of the material. Many course elements have the additional effect of providing instructors insight into student difficulties, allowing them to appropriately adjust instruction.

1. Student attitudes about elements of the course

We find that student attitudes are positive about all course elements. Out of all the course elements, students indicated that lecture, clicker questions, and tutorials were useful for their learning (based on both quantitative ratings and open-ended questions), with minimal variation across courses. Student enjoyment of the course elements follows the same general pattern. Whiteboards were rated lowest, and were spontaneously discontinued by faculty in later implementations. Full results from quantitative ratings are

shown in the Appendix. Again, PER-A stands out as an outlier among the courses, especially in terms of low student enjoyment and perception of the utility of tutorials.

Clickers and tutorials were rated as most useful for learning compared to other aspects of the course in most courses. In open-ended questions, students across all semesters most often mentioned clickers and tutorials (and less often homework assignments and homework help sessions) as positive aspects of the course, and indicated that they would be disappointed if these were removed. A typical response: "I especially liked the tutorials, clicker questions, and help sessions—I wish every physics class was run like this one."

2. Tutorials

Students rated the tutorials highly in terms of the overall experience, effectiveness in encouraging interest in E&M, the intellectual challenge, amount learned, and overall enjoyment. Many students gave spontaneous positive comments about the tutorials on the end-of-semester surveys: "I really liked the Friday tutorials. They were (generally) fun, interesting, and a good jump start to keep me excited over the weekend. Also, I learned a lot." Attendance was fairly high for an optional Friday activity; about 40% of the class, on average. As with the clicker questions, tutorials also offer a valuable opportunity for instructors to interact with students and gain insight into their difficulties, as well as a positive experience: "It's also fun for us," said one instructor in interviews. "I enjoy the tutorials a lot. It's really a great time."

In addition to student and instructor enjoyment, we wanted to know whether the tutorials were truly beneficial to student learning. We find that tutorial attendance covaries with a variety of performance measures, such that students who attended more tutorials had exam scores that were higher than the median for that particular course ($r = 0.27$), higher course grades ($r = 0.36$), and higher CUE scores ($r = 0.26$, $p < 0.01$, $N = 158\text{--}198$). Similarly, compared to students who attend fewer than four tutorials, students attending four or more of the tutorials ($N = 100$, 49% of sample) have higher CUE post-test scores (60.7 versus $50.8 \pm 2\%$). Tutorial attenders also have higher exam z-scores [37] (0.27 ± 0.1 versus 0.23 ± 0.1) and course grades (3.2 versus 2.4 ± 0.1) compared to the course as a whole. This correlation could suggest that tutorials have a positive benefit for students.

These relationships hold true (and stronger) for students with weaker backgrounds: Among students with prerequisite course grade averages less than 3.0, those who attended more tutorials had higher course grades ($r = 0.49$) and CUE scores ($r = 0.39$, $p < 0.01$, $N = 56\text{--}79$), though not exam z-scores. These results suggest that perhaps tutorials are more beneficial for these lower-performing students.

However, because tutorial attendance is optional, these results are complicated by self-selection effects. We find that those students who attended 4 or more of the tutorials have stronger backgrounds compared to those who attended fewer tutorials, with significantly higher GPAs in prerequisite courses (3.2 versus 3.0 ± 0.06), though these tutorial-attending students did *not* score better on conceptual precourse assessments (BEMA scores and CUE pretest). Thus, many of the outcome variables (such as CUE and course grades) could be related to this background preparation rather than the tutorials. This concern is further justified by the difference in lecture attendance between tutorial attenders and nonattenders (87 versus $74 \pm 3\%$) suggesting that the highly motivated students are drawn to the tutorials. Thus, we performed a series of multiple regressions to investigate the effect of tutorials on student exam and CUE scores, when such background variables are taken into account. The details of this regression are described in detail in the Appendix and in a prior publication [38].

We find that student scores on the CUE are well predicted by their GPA in their prior physics courses or (for those students for whom we have BEMA scores after introductory physics) the BEMA. In either case, the addition of tutorial attendance as a predictor significantly improves the model. For those students with BEMA scores, we can account for 46% of the total variance in CUE score through students' BEMA scores and tutorial attendance. The effect of tutorials on CUE scores is roughly one-third that of either physics GPA or BEMA. This indicates that tutorial attendance does provide some improvement in performance on conceptual assessments, even when background performance is taken into account.

The same is not true for the traditional exams, however, BEMA scores and tutorial attendance did not enter into the model as significant predictors for the difference of a student exam score above the course mean. The PHYS GPA variable alone predicts 46% of the variance in student exam z-scores for the student population as a whole. Thus, it appears that conceptual reasoning (as measured by BEMA performance or the experience gained in tutorials) does not strongly affect students' ability on these calculation-focused assessments.

3. Clicker questions and lecture

Did lecture time help student learning? Lecture attendance (as gauged by the presence or absence of a student response to clicker questions in a particular lecture) is moderately correlated with post-test scores on the CUE ($r = 0.199$, $p < 0.05$, $N = 161$), while lecture attendance is strongly correlated with traditional measures such as course grade ($r = 0.35$, $p < 0.001$, $N = 201$) and average course exam score ($r = 0.29$, $p < 0.001$, $N = 200$). However, these correlations may be due in part to a self-selection effect: In a linear regression model (see *Tutorials*,

above), lecture attendance was not a significant predictor of student scores on the CUE or course exams when grades in prior physics courses were taken into account. Because of the low variability in lecture attendance, however, it is difficult to directly determine the effect of this variable on student performance.

Does the full impact of the course transformations lie in the impact of the tutorials on the minority of students who attend them? We find that it does not. Despite the fact that those students in PER courses who did not attend any tutorials ($N = 42$) had lower CUE scores than tutorial-attending students, we find that they still achieved higher CUE scores than did the students in STND-A and STND-B courses ($N = 49$; 52% vs 43%, $p < 0.05$). The strength of this result is bolstered by the fact that the tutorial non-attenders represent lower-performing students in the class: They score lower than tutorial attenders on a variety of measures of motivation and aptitude (see *Tutorials*, above), such as overall course grade (68% vs 81%). Since the primary experience that these tutorial nonattenders had in common is attendance at lecture, this indirect evidence suggests that the addition of clickers and peer instruction is at least partly responsible for the increased CUE scores in transformed courses.

Student survey data also suggest that students find the clicker questions extremely valuable, as documented in other courses [39]. Students are very positive about the use of clickers in this course, as reported elsewhere [11]. In all PER courses, they agree that the addition of clickers, as opposed to pure lecture, is very useful for their learning, and overall recommend the use of clickers in upper-division courses. Write-in responses often mirrored this positivity: "I would not want the clicker questions to change because they give me a chance to talk over any questions I have with my classmates and work through the problem," said one student.

One important benefit of clickers is the window that they provide into student thinking, allowing the instructor to be more closely connected to student difficulties and to target their instruction accordingly. "I know whether the students understood what I just said, for instance, or whether I screwed up," explained one instructor, regarding the use of clickers.

Our lecture was fairly traditional, though broken up by clicker questions. Student comments suggest that lecture was not *ineffective* as implemented but perhaps could be made even more interactive. Several students requested that lecture include demonstrations, simulations, or connection to real-world phenomena and applications—comments also made by alumni regarding traditionally taught lectures. They also requested more worked-example problems—a comment also made by students in the STND-A class. Students in the less popular PER-A semester were the only ones who complained that lecture was too "slow," did not prepare them for the homework, or that they wanted "more math" in lecture.

4. Homework and homework help sessions

Out of all course elements, homework assignments and help sessions were mentioned the least frequently by students when giving spontaneous commentary about the course. However, given that homework is typically considered rather onerous, the student praise for this aspect of the course was satisfying. Students indicated that the challenge of the homework was beneficial to their learning: “(The homework assignments) were challenging enough that you would feel like you actually learned/accomplished something once you finished a problem” writes one student; “Homeworks ... were challenging, but very helpful to my learning,” writes another. This positive attitude towards homework appears to be an aspect of upper-division courses rather than of the transformed course, as department alumni (who had taken traditionally taught courses) often mentioned the challenging homework as being a rewarding aspect of their experience in the course. Several students in the transformed courses also mentioned that they appreciated the emphasis on conceptual understanding and/or an understanding of the physical world. The creation of homework assignments for which the solutions were not freely available on the Web (unlike those for the traditional textbooks [15]) also encouraged the development of problem-solving skills by removing this tempting crutch. The most common complaint was that the homework assignments were too long or hard: “I do not think the homework need be quite so bone-crushing I think I was more worried about getting answers down than actually concentrating on the physics.”

Traditional office hours were replaced by homework help sessions, with the instructor circulating among students working in groups. Attendance at these sessions was only recorded for PER-A and PER-C: On average, most students (86%) attended at least one help session, but their attendance varied quite widely, with an average attendance of 40% of the sessions. In a multiple regression analysis (not shown), we find that attendance at homework help sessions is a significant predictor of homework score in these two courses.

Students who attended at least four help sessions tended to be a bit higher performing than those who did not attend: These students had higher attendance at PHYS301 lectures and cumulative GPA prior to PHYS301, and there are nonsignificant trends such that these students had higher exam scores, CUE scores, and lecture attendance in PHYS 301, as well as incoming GPA. Thus, homework help sessions are likely attended by the more diligent or motivated students in the course.

It is difficult to gauge any particular effect of these homework sessions, but many students indicated that they would be disappointed if these sessions were no longer offered. Student comments suggest that they find it valuable to have this additional opportunity to interact with the instructor, and found this more useful than

one-on-one office hours or an instructor or teacher’s assistant presenting material to the group as a whole. Of particular interest to us was the number of student comments regarding how these sessions (and the difficulty of the homework in general) spurred both (a) communication and collaboration between students (learning goal no. 4) and (b) students’ own sense of what it meant to do physics. “(The homework help sessions) honed my skill of learning to ask the right question,” said one student. The instructor’s ability to guide students without solving the problem for them was also appreciated by some students. Thus, the homework help sessions are important opportunities for students to learn to “talk physics” with one another and to receive instructor guidance and feedback as they learn to become expert problem solvers (learning goals no. 1, 2, 5, 6, and 7). As with the clicker questions and tutorials, help sessions also offer instructors an opportunity to get feedback on common student difficulties, allowing them to tailor instruction accordingly. Instructor comments from interviews reinforce this: “On the whole it’s let me have really more communication with the class, to have a better sense of what’s going on with them than I would have otherwise,” said one instructor.

Thus, our data suggest that all course elements—clicker questions, tutorials, homework, and homework help sessions—are seen as valuable by students. Statistically speaking, we are only able to discern a significant impact of tutorials on student learning outcomes. In surveys, students show particular preferences for clicker questions and tutorials but describe qualitative benefits of all course elements.

F. Gender differences

Did males and females experience different outcomes of the course transformations? At the introductory level, there is a well-documented gender gap in course performance. Female students tend to score more poorly both on conceptual measures such as the BEMA [40] and on traditional exams [41]. There is mixed evidence as to whether interactive engagement in introductory courses reduces the gender gap [33] or not [34,40,42]. This gender gap at the introductory level at CU was shown to be related primarily to the female students’ incoming preparation; females receive lower scores on a conceptual exam at the beginning of the introductory course [41]. Because the students in upper-division courses are among the strongest, most committed, and best prepared students from the lower-division courses, we did not expect to see strong gender differences in student performance. However, we do find a gender gap on many of the measures in this course, which appears to be ameliorated by the course transformations. We find evidence that female students enter PHYS301 less well prepared than male students and receive lower grades in PHYS301, and that the PER-based courses may have helped to close this gender gap, primarily by improving

scores on exams. The full description of this analysis can be found in the Appendix.

It is hard to determine how broadly we may apply these results. Looking across an historical sample of upper-division physics and mathematics courses at CU, female students receive similar or higher course grades than men, typically by about 0.2 grade points, unlike the courses in our study. However, in investigating results from an external institution (STND-6; see Table I), we find a similar trend to that observed in PHYS301.

Thus, the course transformations appear to have a positive effect in closing the gender gap in performance on both conceptual measures targeted to PHYS301 and more traditional measures of assessment. These findings may be limited to the content area of PHYS301, which appears to pose more difficulty for female students than do other courses. These results are of particular note since they provide insight into the experiences of a population of physics majors very different from the introductory students studied previously. Also, due to the documented loss of female physics majors at the end of an undergraduate degree [43], information on how to improve the performance of female physics majors is of interest.

V. DISCUSSION AND CONCLUSIONS

We have transformed an upper-division physics course using faculty-consensus learning goals to direct the course content, and interactive techniques from lower-division courses to guide pedagogical approaches. We have researched student learning and difficulties, developed instructional materials, and used a newly developed conceptual assessment to test the impact of this approach.

We have found that our traditional courses appear to do a relatively good job of teaching calculational skills. In our first pass at transforming this course, we attempted to, additionally, address the concepts, process, and habits of mind of the upper-division course in a more fully supported way. Examining our course goals broadly categorized (mathematical sophistication, problem-solving expertise, and developing as a physicist), we have presented evidence of overall improvement of student scores on the conceptual CUE exam, reasoning and problem-solving approaches on traditional course exams, and student and instructor attitudes. Students report spending more time on the course and are enthusiastic about the various elements of the course transformations. Attendance at tutorials is associated with improved CUE scores, and those students who do not attend tutorials perform better on the CUE than students in traditionally taught courses, indicating that course elements such as clickers and peer instruction also have tangible benefits for student learning.

Students are not the only participants positively affected by the course. Instructors also enjoyed teaching the new course; this sort of buy-in reduces a significant barrier to dissemination and sustainability of course transformations.

Instructors also mention several aspects of the course as a valuable opportunity to gain insight into student thinking. Our sense is that this is where much of the utility of the course transformations lie—clickers, tutorials, and homework help sessions all provide additional communication between instructors and students, allowing instructors to appropriately adjust instruction.

The course transformations also help to change the classroom culture and norms. One instructor (teaching a similarly transformed quantum mechanics course [44]) indicated that clickers helped to frame the class as an interactive environment, essentially breaking the ice so that it was easier to generate conversation in the class. As in other course reforms [45], student buy-in is also key; in PER-A, where there was evidence of student resistance, learning gains and student attitudes were less favorable.

We have been successful by many measures—but more work remains. The course transformations did *not* hurt students' ability to perform mathematical calculations (e.g., separation of variables, direct integration), but neither did they particularly improve this skill, as measured by traditional exam problems. Student scores on the CUE—while higher than those in the standard lecture-based courses—still average only 60%. While tutorial attendance helped performance on the CUE, traditional exams remain unaffected—despite our intentions to help students with skills that ought to be tested on exams (e.g., to make meaning of the mathematics, break down problems into parts, and to clarify basic concepts). Students were positive about the lecture parts of the course, but their attitudes are not as positive as we would have expected on certain measures, and many complaints mirror those in traditional classes. While we have taken several steps towards creating a more effective learning environment, we may not have gone far enough in changing the classroom environment and creating instructional materials to reach our goals.

One limitation of our current materials may be a failure to adequately target instruction to students' difficulties at this level. In order to improve students' mathematical sophistication and problem-solving techniques, as outlined in the learning goals, we must know the specifics of where students struggle. We are limited in our ability to do this, as our knowledge of particular student difficulties at this level—and development of instructional techniques to target them—is also still in its infancy. Based on interviews, it appears that instructors who used our preliminary documentation of student difficulties when planning their instruction had courses with higher CUE gains than those who did not. However, we do not provide explicit guidelines for instructors in how to effectively use this information, and instructional materials are still only loosely based on these rough descriptions of what students struggle with. We find that the majority of students often answer our clicker questions correctly (which does not happen with our introductory-level questions based on well-researched

difficulties)—which tells us that these questions are not adequately targeting student difficulties and there is room to provide additional learning opportunities for students through clicker questions. There is need for further research into student difficulties and how to address them.

One potential area for future work is to make stronger use of the existing research on problem solving [28,46,47]. A “coaching, modeling and fading” [48] approach could be used, for example, where strategies are modeled for students, their attempts to implement are supported, and then students are given an opportunity to proceed on their own. Instructor guidelines for homework help sessions, focused aspects of the tutorials, and classroom activities and techniques could be developed to provide more explicit and research-based instruction in problem solving. Or, perhaps more importantly, development of richer, authentic problems would better prepare students for the kind of scientific thought that we expect of them. As such, it could be worthwhile to encourage students to reflect on the process and goals of problem solving [49]—why are we solving this problem and what does it mean?

Another area for future development could be to facilitate students’ conceptual organization of the course. During interviews, even the best students often did not connect key concepts (such as voltage and separation of variables) on a concept map in a manner approaching that of an expert—an observation also recorded among introductory students [50]. If students struggle to recognize what quantities the various analytical methods are solving for, how they are related to one another, what premises or assumptions they are based on, and where they fit into the big picture, they will likely experience overload when engaging in the more complex tasks involved in solving problems.

While our materials could conceivably be tweaked as described above, the fact that our gains are more modest than hoped could be interpreted as (a) an indication that the CUE itself is flawed or (b) evidence for the need for a different kind of approach. Considering the latter, perhaps applying the model of course reform that has proved effective at the introductory level—clicker questions, homework, and tutorials aimed at student engagement and conceptual understanding—only takes us so far. Similarly, interactive engagement techniques at the introductory level result in higher levels of learning gains with more extensive reforms (e.g., clickers, tutorials, and learning assistants) than modest transformations (e.g., clickers alone) [51]. Similarly, it may be that our techniques are not the “right level of radical” for this population. These students are more motivated, skilled, and mature than introductory students. Clicker questions and tutorials provide a high level of guidance—asking students to choose between predetermined answer choices or to follow a set of guided procedures. Rather, we may better achieve our learning goals if students begin to generate their own answers to open-ended questions or to generate the

questions themselves. The innovative techniques in the paradigms courses at Oregon State [52] are based on this philosophy: “Facilitating activities in these more advanced courses does not obviously follow from techniques that are successful in introductory courses,” they write on their Web site. In a paradigms course students are engaged in problem solving in class and active communication, sense making, and reflection upon the meaning of their results [53]. In one activity, for example, each group of students calculates the magnetic field due to different current distributions in a cylindrical shell, presents their results to the class using small whiteboards, and compare and contrast the different cases in terms of units, total current, and symmetry arguments. Students may also work in groups to solve a problem in chunks with the instructor’s mentorship, work on laboratory activities or computational techniques, or individually work out a calculation or open-ended question and share their answer with small whiteboards. This intensive environment results in the creation of a learning community that works together to gain expertise in the processes of physics. While the efficacy of paradigms remains untested, the model is intriguing in suggesting the “next step” in our own reforms.

Of course, we are limited in our ability to directly encourage particular instructor practices in the classroom. Instead, what we can provide are implementable “chunks” of course materials—namely, tutorials, clicker questions, and homework—which instructors choose *à la carte* to best suit their purposes and instructional goals. But pedagogical implementation of those materials depends on the individual instructor as well as institutional constraints. So, while our developers may create and implement diverse course techniques, radical pedagogical change is likely to be difficult to impart broadly [54]. At the very least, making use of shorter in-class activities (beyond clickers, but less intensive than tutorials) which target specific difficulties or skills, such as those in paradigms, could be a reasonable focus of future activity.

While we set out to design instructional materials to address the learning goals and to target student difficulties with course content, we hypothesize that many of the benefits of the course improvements are also related to cultural changes in the course, where students were encouraged to participate through a variety of methods. This course approach opened the door to create a more student-centered classroom culture. Instructors were given opportunities to see how students were thinking, dialogue and conceptual reasoning were valued, and students participated in active practice and application of physics, where they received support from peers and experts. While research may point the way towards further improvements in this area, this student-centered approach appears to be the first step in the goal of supporting the next generation of physics students as they work towards more sophisticated levels of scholarship and thought.

ACKNOWLEDGMENTS

We acknowledge the generous contributions of faculty at CU: Dr. Dubson and Dr. Kinney provided modifications and improvements to the course materials, Dr. Beale facilitated much of our work with the faculty working group and served as a valuable faculty advisor, and Dr. Carl Wieman provided thoughtful mentorship, guidance, and recommendations through the early years of the project. D. Tarshis assisted with initial development of many tutorials, and W. Handley helped draft early versions of clicker questions. We acknowledge our colleagues in the Physics Education Research group at CU for their helpful feedback over the lifetime of this project. We also appreciate the work of the faculty working group,

which included Dr. Beale, Dr. Betterton, Dr. DeAlwis, Dr. DeGrand, Dr. DeWolfe, Dr. Dubson, Dr. Finkelstein, Dr. Ford, Dr. Hasenfratz, Dr. Munsat, Dr. Parker, Dr. Perkins, Dr. Pollock, Dr. Rogers, and Dr. Zimmerman. This work is funded by The CU Science Education Initiative and NSF-CCLI Grant No. 0737118. R.E.P. is currently funded by the Miller Institute for Basic Research in Science at the University of California Berkeley.

SUPPLEMENTAL MATERIAL

See separate supplemental material for learning goals, survey, information on courses, student attitudes, regression, and gender.

[1] See, for example, B. S. Ambrose, Investigating student understanding in intermediate mechanics: Identifying the need for a tutorial approach to instruction, *Am. J. Phys.* **72**, 453 (2004), and the first six references therein.

[2] S. J. Pollock, Longitudinal study of student conceptual understanding in electricity and magnetism, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020110 (2009).

[3] For example, L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, *Am. J. Phys.* **67**, 755 (1999); Ruth Chabay and Bruce Sherwood, Restructuring the introductory electricity and magnetism course, *Am. J. Phys.* **74**, 329 (2006).

[4] C. Singh, Student understanding of symmetry and Gauss's law of electricity, *Am. J. Phys.* **74**, 923 (2006).

[5] C. A. Manogue, K. Brown, T. Dray, and B. Edwards, Why is Ampère's law so hard? A look at middle-division physics, *Am. J. Phys.* **74**, 344 (2006).

[6] R. Pepper, S. V. Chasteen, S. J. Pollock, and K. K. Perkins, Observations on student difficulties with mathematics in upper-division electricity and magnetism, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010111 (2012); R. E. Pepper, S. V. Chasteen, S. J. Pollock, and K. K. Perkins, Our best juniors still struggle with Gauss's law: Characterizing their difficulties, *AIP Conf. Proc.* **1289**, 245 (2010); C. Wallace and S. V. Chasteen, Upper-division students' difficulties with Ampère's law, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020115 (2010); S. V. Chasteen and S. J. Pollock, A research-based approach to assessing student learning issues in upper-division electricity & magnetism, *AIP Conf. Proc.* **1179**, 7 (2009).

[7] C. A. Manogue, P. J. Siemens, J. Tate, K. Brown, and A. J. Wolfer, Paradigms in physics: A new upper-division curriculum, *Am. J. Phys.* **69**, 978 (2001); B. R. Patton, Group learning-based approach to the graduate electrodynamics course: Jackson by inquiry, APS Forum on Education, August, 25, 1996 [<http://www.aps.org/units/fed/newsletters/aug96/index.html>]; S. V. Chasteen and S. J. Pollock, Transforming upper-division electricity and magnetism, *AIP Conf. Proc.* **1064**, 91 (2008).

[8] T. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020108 (2009).

[9] E. Breitenberger, The mathematical knowledge of physics graduates: Primary data and conclusions, *Am. J. Phys.* **60**, 318 (1992).

[10] G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, *Am. J. Phys.* **80**, 252 (2012).

[11] K. K. Perkins and C. Turpen, Student perspectives on using clickers in upper-division physics courses, *AIP Conf. Proc.* **1179**, 225 (2009); S. J. Pollock, S. V. Chasteen, M. Dubson, and K. K. Perkins, The use of concept tests and peer instruction in upper-division physics, *AIP Conf. Proc.* **1289**, 261 (2010).

[12] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall College Division, Upper Saddle River, NJ, 2001).

[13] L. C. McDermott, P. S. Schaffer, and the Physics Education Group at the University of Washington, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).

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

[15] D. J. Griffiths, *Introduction to Electrodynamics* (Prentice-Hall, Upper Saddle River, New Jersey, 1999), 3rd ed.

[16] S. V. Chasteen, K. K. Perkins, P. Beale, S. J. Pollock, and C. E. Wieman, A thoughtful approach to instruction: Course transformations for the rest of us, *J. Coll. Sci. Teach.* **40**, 70 (2011) [http://www.nsta.org/publications/browse_journals.aspx?action=issue&id=10.2505/3/jest11_040_04].

[17] These course transformation efforts were facilitated by the grant-funded employment of one of the authors (S. V. C.)—a postdoctoral Science Teaching Fellow supported by internal and external grant funding through the Science Education Initiative.

[18] The Science Education Initiative is a five-year \$5 million project to improve student learning by approaching the teaching of science as a science through systematic course

transformation in multiple departments at CU-Boulder, <http://colorado.edu/sei>.

[19] S. V. Chasteen, R. E. Pepper, S. J. Pollock, and K. K. Perkins, But does it last? Sustaining a research-based curriculum in upper-division electricity & magnetism, *AIP Conf. Proc.* **1413**, 139 (2012).

[20] S. V. Chasteen, S. J. Pollock, R. E. Pepper, and K. K. Perkins, Thinking like a physicist: A multi-semester case study of junior-level electricity and magnetism, *Am. J. Phys.* (to be published).

[21] One semester in year four did not elect to participate in the study.

[22] R. E. Pepper, S. V. Chasteen, S. J. Pollock, and K. K. Perkins, Facilitating faculty conversations: Development of consensus learning goals, *AIP Conf. Proc.* **1413**, 291 (2012).

[23] The full set of course materials, including learning goals, clicker questions, and assessments, is available at http://www.colorado.edu/sei/departments/physics_3310.htm and on Compadre at <http://www.compadre.org/psrc/items/detail.cfm?ID=7891>.

[24] The alumni survey is available online at <http://www.colorado.edu/sei/documents/phys-alumni-survey.pdf>.

[25] For example, J. Tumarino, Ph.D. dissertation, University of Maryland, College Park, 2004; B. Sherin, How students understand physics equations, *Cogn. Instr.* **19**, 479 (2001); see also Ref. [8].

[26] For example, J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980); D. Maloney, in *Handbook of Research on Science Teaching and Learning*, edited by D. Gabel (MacMillan, New York, 1994), pp. 327–354; F. Reif and J. I. Heller, Knowledge structures and problem solving in physics, *Educ. Psychol.* **17**, 102 (1982).

[27] For example, A. Elby, Helping students learn how to learn physics, *Am. J. Phys.* **69**, S54 (2001); D. Hammer and A. Elby, in *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, edited by B. K. Hofer and P. R. Pintrich (Lawrence Erlbaum, Mahwah, NJ, 2002), pp. 169–190; A. H. Schoenfeld, in *Cognitive Process Instruction*, edited by J. Lochhead and J. J. Clement (Franklin Institute Press, Philadelphia, PA, 1978), pp. 315–338; *Mathematical Problem Solving* (Academic, New York, 1985); in *Handbook of Research in Mathematics Teaching and Learning*, edited by Douglas A. Grouws (MacMillan, New York, 1992), pp. 334–369.

[28] See G. Polya, *How to Solve It* (Princeton University Press, Princeton, NJ, 1945); A. H. Schoenfeld, Teaching problem solving skills, *Am. Math. Mon.* **87**, 794 (1980).

[29] <http://nsfnoyce.org/>.

[30] L. Ding, R. Chabay, and B. Sherwood, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).

[31] S. V. Chasteen and S. J. Pollock, Tapping into juniors' understanding of E&M: The Colorado Upper-Division Electrostatics (CUE) Diagnostic, *AIP Conf. Proc.* **1179**, 109 (2009).

[32] S. V. Chasteen, R. E. Pepper, M. D. Caballero, S. J. Pollock, and K. K. Perkins, Colorado Upper-Division Electrostatics (CUE) Diagnostic: A conceptual assessment for the junior level, *Phys. Rev. ST Phys. Educ. Res.* (to be published).

[33] M. Lorenzo, C. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, *Am. J. Phys.* **74**, 118 (2006).

[34] J. Docktor and K. Heller, Gender differences in both Force Concept Inventory and introductory physics performance, *AIP Conf. Proc.* **1064**, 15 (2008).

[35] <http://usnews.com/college>.

[36] Percent agree is calculated by a weighted average, such that “strongly agree” is weighted as “4” and “strongly disagree” as “0,” normalized to 100, and then averaging over all student responses in that course. Thus, a higher number indicates stronger student agreement.

[37] Because course exam scores will vary by instructor, we compare student exam scores based on the standard deviations from the course mean (the *z* score) rather than differences on the exam score itself.

[38] S. V. Chasteen, Teasing out the effect of tutorials via multiple regression, *AIP Conf. Proc.* **1413**, 143 (2012).

[39] K. K. Perkins and C. Turpen, Student perspectives on the use of clickers in upper-division physics courses, *AIP Conf. Proc.* **1179**, 225 (2009).

[40] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010107 (2007).

[41] 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).

[42] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamela, Toward equity through participation in Modeling Instruction in introductory university physics, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010106 (2010).

[43] <http://www.nsf.gov/statistics/women/>.

[44] S. Goldhaber, S. Pollock, M. Dubson, P. Beale, and K. Perkins, Transforming upper-division quantum mechanics: Learning goals and assessment, *AIP Conf. Proc.* **1179**, 145 (2009).

[45] For example, L. Breslow, Wrestling with pedagogical change: The TEAL initiative at MIT, *Change* **42**, 23 (2010).

[46] A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, *Am. J. Phys.* **59**, 891 (1991).

[47] L. Hsu, E. Brewe, T. M. Foster, and K. A. Harper, Resource letter RPS-1: Research in problem solving, *Am. J. Phys.* **72**, 1147 (2004).

[48] A. Collins, J. S. Brown, and S. E. Newman, *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1989).

[49] A. H. Schoenfeld, in *Cognitive Science and Mathematics Education*, edited by A. H. Schoenfeld (Lawrence Erlbaum Associates, Hillsdale, NJ, 1987).

[50] E. Bagno and B. Eylon, From problem solving to a knowledge structure: An example from the domain of electromagnetism, *Am. J. Phys.* **65**, 726 (1997).

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

[52] C. A. Manogue, P. J. Siemens, J. Tate, K. Brown, and A. J. Wolfer, Paradigms in physics: A new upper-division curriculum, *Am. J. Phys.* **69**, 978 (2001).

[53] Based on materials online at <http://www.physics.oregonstate.edu/portfolioswiki>, a co-author's experience observing and teaching paradigms (S. J. P.), as well as selected publications previously listed, and E. van Zee and C. Manogue, Documenting and interpreting ways to engage students in “thinking like a physicist,” *AIP Conf. Proc.* **1289**, 61 (2010).

[54] M. Dancy and C. Henderson, Pedagogical practices and instructional change of physics faculty, *Am. J. Phys.* **78**, 1056 (2010); C. Henderson and M. Dancy, Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations, *Am. J. Phys.* **76**, 179 (2008).