

Sustaining educational reforms in introductory physics

Steven J. Pollock and Noah D. Finkelstein

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

(Received 19 February 2008; published 3 June 2008)

While it is well known which curricular practices can improve student performance on measures of conceptual understanding, the sustaining of these practices and the role of faculty members in implementing these practices are less well understood. We present a study of the hand-off of *Tutorials in Introductory Physics* [McDermott and Schaffer (Prentice-Hall, Upper Saddle River, NJ, 2002)] from initial adopters to other instructors at the University of Colorado, including traditional faculty not involved in physics educational research. The study examines the impact of implementation of tutorials on student conceptual learning across ten first-semester, and seven second-semester courses, for 15 faculty members over 13 semesters, and includes roughly 5000 students. It is possible to demonstrate consistently high, and statistically indistinguishable, student learning gains for different faculty members; however, such results are not the norm and appear to rely on a variety of factors. Student performance varies by faculty background—faculty involved in, or informed by physics education research, consistently post higher student learning gains than less-informed faculty. Student performance in these courses also varies by curricula used—all semesters in which the research-based *Tutorials* and learning assistants are used have higher student learning gains than those semesters that rely on nonresearch-based materials and do not employ learning assistants.

DOI: [10.1103/PhysRevSTPER.4.010110](https://doi.org/10.1103/PhysRevSTPER.4.010110)

PACS number(s): 01.40.Fk

I. INTRODUCTION

It is increasingly well understood how to promote student learning in large-enrollment, introductory physics courses. At least by measures of performance on conceptual surveys, a wide variety of pedagogical approaches based on the findings of physics education research (PER) have been demonstrated to dramatically improve student performance.¹⁻³ Two of the great remaining educational challenges are how to scale reforms beyond the location where they were developed originally and whether these reforms can be sustained. There is significant evidence that it is possible to replicate the successes of curriculum developers⁴⁻⁶; however, success is far from guaranteed.^{7,8} It appears that how these practices are enacted is critical. Notably while interactive engagement (IE) courses outperform nearly all non-IE classes,⁹ the variation in normalized learning gains among IE courses is dramatic. This variation of success among implementations points to some of the difficulties of sustaining educational change.^{10,11} What happens when proven strategies are handed off to faculty not involved in the design or initial implementation of educational reforms?

This paper explores the role of sustaining reforms that have been demonstrated to be successful both by the original creators and by the initial implementers at a second institution. We address the following research questions:

(i) Is it possible to hand off these educational reforms from one faculty member to another and sustain the level of student achievement realized initially?

(ii) Does the instructor's background matter (PER vs non-PER)?

(iii) Do the particular curricula matter or is it entirely a matter of faculty practices?

(iv) Can we demonstrate the development of faculty members as they implement these new pedagogical practices?

II. BACKGROUND, ENVIRONMENT, AND DATA SOURCES

In prior work,⁶ we described our implementation of *Tutorials in Introductory Physics*¹² and demonstrated that it was possible to replicate measurements of conceptual shifts in students that were remarkably similar to those of the curricular authors, researchers at University of Washington (UW) Physics Education Group. Figure 1 illustrates this similarity. Data from UW are shown for post-instruction results without the use of tutorials (white) and following their introduction (grey).¹³ Results from first implementation of tutorials at the University of Colorado (CU) at Boulder are shown in black. Each category represents conceptual questions from the UW¹³ used to demonstrate the effectiveness of their tutorial materials.

To understand how we achieved such similar results, our prior analysis⁶ examined various nested levels, or frames of

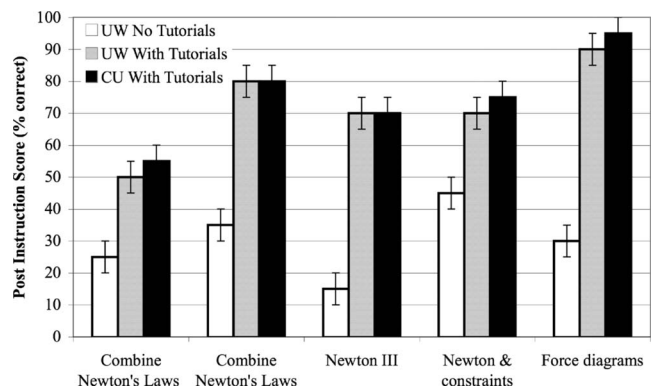


FIG. 1. Post-test results on four published (Ref. 13) conceptual questions asked on midterm exams. Results from UW and CU are shown. “No-Tutorial” scores are shown for courses at UW only. Following UW reporting practices, our results are rounded to the nearest 5%.

context,¹⁴ ranging from the specific tasks that students engaged in, to the surrounding departmental and institutional structures. The two critical frames were the *situation* and the *class culture*. Situations refer to the surrounding structure of students working at tutorials, including the rooms, teaching assistants (TAs) and materials, and social setting during the 50 min recitation sessions. The class culture is defined as the environment in which these situations are set—the norms of behavior for the course, the purposes of the tutorials, grading policies, etc. While the specific tasks, the tutorial workbook problems, may be straightforward to hand off from one faculty member to another, the situation and class culture levels are harder to prescribe and thus potentially subject to greater variation. As the leading role for these courses is handed from one faculty member to another, so too might we anticipate shifts in the enacted practices that constitute situations and class culture. Here, we begin to explore these issues in studies of the sustained implementation of tutorials and related course reforms, including *Peer Instruction*¹⁵ in our introductory course sequence.

A. Course description

The studies at the CU occurred in the calculus-based, large-scale ($N=300$ – 600) introductory physics sequence. The student population is a mix of engineering majors (50%), natural science majors (20%), and a variety of others, including undeclared majors. The course is 75% male, and over half the students are freshmen. Data were collected in ten first-semester courses (Fall 2001, then Fall 2003–Fall 2007) and seven second-semester courses (Fall 2004–Fall 2007). All instructors in this study since Fall 2003 used concept tests¹⁵ during three 50 min lectures/week in a 200-to 300-student classroom, online homework systems,^{16,17} one 50 min week/recitation section, and a staffed help room, which was voluntary for students. The introductory laboratory experience is decoupled from this course and typically taken concurrently with the second semester lecture course. All second-semester courses, and many of the first-semester courses, implemented the University of Washington's Tutorials¹² during the weekly recitation sections. The remaining first-semester courses ran more traditional recitation sections, as will be discussed below. The recitation sections were staffed by one departmentally assigned TA, and in the case of tutorials, an additional undergraduate assistant, a learning assistant (LA) (will be discussed below), was added. Two faculty members are generally assigned to teach these large courses; one is in charge of lectures, and the other is in charge of tutorials and the rest of the course administration. The faculty member lecturing is historically the lead of the two-person team, setting course pace, goals, and overall workload. The “secondary” faculty member is typically in charge of the course management, TAs, logistics, and the recitation sections, where tutorials were implemented. As such, the lead faculty is dominant in establishing the class culture and the secondary instructor establishes and frames the situations for tutorials. The CU physics faculty are cycled through these classes; during the course of this study, 11 different faculty members participated in the first-semester

course and 8 faculty members participated in the second-semester course. The faculty in this study are categorized by their familiarity with physics education research, the underpinning principles of learning theory, and the practices built from them. The PER faculty members are defined as members of the PER group, have conducted research in student learning, and published and taught classes in the field. “Informed” faculty are those who are familiar with the field of physics education research, some of its findings, and have engaged in workshops led by PER faculty, participated in departmental brownbag discussions, and/or engaged in readings from the field. These faculty are observed to enact some of the principles from the field, such as those described by Redish.¹ Traditional faculty are the typical university faculty members who receive little or no preparation in teaching, may well report knowing about the field,^{10,11} have not participated in departmental activities designed to inform faculty about PER, and are not observably enacting the principles of the field. Faculty are denoted by “P” (PER researchers), “I” (informed), and “T” (traditional).¹⁸ The initial implementations of tutorials in both the first- and second-semester courses were led by PER research faculty. Following, there were only two other cases (both in the second-semester course) where another PER-faculty member directed the course.

We attempted to implement a fairly high-fidelity replication of the University of Washington tutorials,¹² but local departmental and institutional constraints inevitably resulted in a wide variety of differences. Our recitation sections are all held on one day of the week. We created three temporary spaces (three-walled bays) in the basement laboratory areas, where we fit seven tables of four students each. Sections run in parallel, resulting in a somewhat noisy and cramped environment. Sections are staffed by a graduate TA and one or the two undergraduate LAs. The LAs are students who had previously performed well in this class, expressed interest in teaching, and became part of a broader program to support their development as teachers.¹⁹ TAs and LAs are prepared in a weekly meeting several days prior to the tutorial where they look over student responses to online pretests to generate discussion of incoming student ideas, and are then guided through the upcoming tutorial itself, working in small groups just as students do. When secondary faculty were not members of the PER group, the PER faculty worked fairly closely with these secondary instructors for the first week or two to try to explain the methodology and goals behind the tutorials. We have created a written faculty guide which provides weekly tips and suggestions, along with pretest responses from previous terms. After the first few weeks, faculty are left on their own to run these preparatory meetings as they see fit.

Data collection

Each term, we administer preinstruction and postinstruction assessments of student conceptual understanding. The force and motion concept evaluation (FMCE)²⁰ was used in the mechanics course, generally issued the first and one of the last weeks of the term in recitation. Students are in-

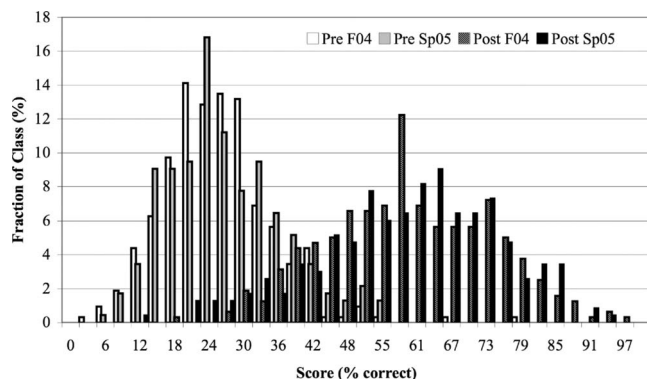


FIG. 2. BEMA results pre-(left) and post-(right), shown as % correct. The data show two semesters of results in different shades. The term-to-term reproducibility of these histograms is striking. The two semesters were taught by different faculty, both PER researchers.

formed that the exam is diagnostic and does not impact their grades, but we encourage them to try their best. We obtain matched prescores–postscores for roughly 60% of students enrolled in the class each term. We use the brief electricity and magnetism assessment (BEMA)²¹ instrument for second-semester physics, with typically 75% pre-post matched return rates. Notably, these instruments provide only a coarse evaluation of course achievements and only represent one dimension of student achievement. Nonetheless, these evaluations allow us to compare, in aggregate, the impact of different course implementations by different faculty on the overall performance of students enrolled in these courses.

STUDIES OF SUSTAINING TUTORIALS

Initial hand-off and intrainstitutional replication

The initial implementation of the tutorials in each of the first- and second-semester calculus-based physics courses was conducted by the lead author (S.J.P.), denoted as P1. The tutorials were implemented in the first-semester sequence (mechanics) for the first time in Fall 2003 and implemented for the first time in the second-semester sequence (electricity and magnetism) in Fall 2004. The following semester, Spring 2005, another faculty member of the PER group and nationally recognized educator taught the second-semester course. This is the first instance of hand-off of the tutorial curriculum at CU and arguably conducted under extremely favorable circumstances. The second instructor, P2, had assisted with the initial implementation of the tutorials, is well versed in physics education research, and read many of the UW papers. He understood the underpinning pedagogy and purpose of the tutorials. When P2 led the course, Sp05, an informed faculty member, I4, served in the secondary role. The course structures themselves were nearly identical from Fa04 to Sp05, but there was some variation in pacing, particular lecture notes, some concept tests, and text (Knight²² vs Halliday *et al.*²³). The tutorials were implemented in nearly identical manner, with teaching and learning assistants, weekly preparatory sessions, and covering the same tutorial topics.

Figure 2 shows the distributions of pretest and post-test

scores on the BEMA for both the initial implementation (dark) and the secondary instructor (light). The post-test score averages were the same, 59%, and the normalized learning gains indistinguishable $44\% \pm 1\%$ (standard error on the mean) for the initial implementation and $43\% \pm 1\%$ for the secondary implementation.

Standardizing curricular choice

While the lead author was implementing tutorials in the second-semester course, the first-semester course was assigned to non-PER faculty. At that time, the tutorials were new at CU and not an institutionalized activity. The faculty members were not convinced of the differential effectiveness of tutorials over traditional recitations, and there were no existing mechanisms to pay for extra costs of implementing this enhanced educational experience. As a result, faculty members were left on their own to select which activities would run in the recitation sections. As a matter of standard practice, the rest of the course activities, lectures with interactive concept tests, online homework system, and recitation sections of some sort were continued and considered institutional practices at that point. The lead author (S.J.P.) taught the first-semester course with tutorials in Fall 2003 and again in Spring 2004. The following semester, the new instructor opted to use the same pedagogical structure, small groups, and different curricular materials, selecting the workbook that came with the text.²² A notable difference was that the institutional support for additional recitation leaders, in the form of learning assistants, did not exist. In recitation sections, the teacher to student ratio was roughly half what it was the prior semester, and the weekly preparatory sessions for recitation were more *ad hoc*, and certainly did not follow the UW approach.^{6,12} The semester following, Spring 2005, the course instructor opted to revert back to more traditional recitation sections, where TAs were often working at the board in front of the students reviewing homework problems or worksheets from the text or of their own design. Again, the recitations were led by single instructors rather than two individuals, and the weekly preparatory sessions were limited and less structured. Table I lists the lead and/or secondary roles, the type of recitation experience, number of students, post-test scores, and the normalized learning gains for each of the first-semester courses studied. The two most significant variations among these semesters are the backgrounds of the instructors and recitation experiences for students.²⁴

Table I and Fig. 3 summarize the overall measures of student conceptual learning gains in first-semester courses. These allow for comparisons of tutorial with traditional recitation sections. Figure 3 is a reproduced plot from Hake⁹ (with permission), a histogram of student normalized learning gains for different environments. The light data represent traditional courses and the dark self-described the “interactive engagement” environments. Overlaying this plot are arrows denoting the span of learning gains for tutorial and traditional recitation experiences offered at CU, as detailed in Table I.^{26,27} Notably, all of the CU courses, whether traditional or tutorial recitations, are interactive environments us-

TABLE I. Summary of FMCE normalized gains over eight semesters of data collection.

| Semester | Faculty background lead and/or secondary | Recitation | N (matched) | Average post-test score ^a | Normalized gain $\langle g \rangle^b$ |
|----------|--|-------------|---------------|--------------------------------------|---------------------------------------|
| F01 | T0 (alone) ^c | Traditional | 265 | 52 | 0.25 |
| F03 | P1 (alone) | Tutorials | 400 | 81 (FCI data) | 0.63 |
| S04 | P1 (alone) | Tutorials | 335 | 74 | 0.64 |
| F04 | I1/T1 | Workbooks | 302 | 69 | 0.54 |
| S05 | T1/T2 | Traditional | 213 | 58 | 0.42 |
| F05 | T2/T3 | Traditional | 293 | 58 | 0.39 |
| S06 | I2/T4 | Tutorials | 278 | 60 | 0.45 |
| F06 | I3/I1 | Tutorials | 331 | 67 | 0.51 |
| S07 | I4/T4 | Tutorials | 363 | 62 | 0.46 |
| F07 | I2/I5 | Tutorials | 336 | 69 | 0.54 |

^aThe average post-test score is a percentage, scored using Thornton’s recommended FMCE rubric (Ref. 3), for those students with matched pre-post data. (The standard error of the mean is between 1% and 2% for all terms). Only F03 used the force concept inventory (FCI) (Ref. 25) exam pre and post, all other terms are FMCE. (Note the similar gains for F03 and S04 on these two exams.)

^bNormalized gain in the last column is computed as the gain of the average prescores and postscores for matched students. (The standard error of average gains is roughly ± 0.02 for all terms.)

^cThe F01 course included traditional recitations and peer instruction with colored cards in lecture.

ing peer instruction and personalized, online homework systems. While CU courses span nearly the entire range of learning gains documented for IE courses elsewhere,⁹ we note that at CU, all courses with tutorial experiences lead to learning gains higher than all classes that have traditional recitation experiences. Furthermore, there is sizable variation of success among these implementations, suggesting a potential effect of faculty in these environments.

From Spring 2006 to Fall 2007 (and ongoing) in the first-semester course and from Fall 2004 to Fall 2007 (and ongoing) in the second-semester course, tutorials have been implemented consistently in the recitation section, including the increased teacher: student ratio by using LAs. While there are many conditions that have led to the institutionalization of tutorials, from our multiyear history of involvement, we hypothesize that critical features that have sup-

ported adoption of these materials may include the following.

(i) *Open discussions with faculty* about course goals and the impact of tutorials.

(ii) *Increased support for the recitation sections.* The number of learning assistants allocated to the department increased, initially through grant funding (APS, AIP, and AAPT PhysTEC funding²⁸ coupled to an NSF STEM-TP grant²⁹). Subsequently, the increased support came from the department and dean, and now the university level support. Roughly \$50,000 per year supports the learning assistant program in the physics department.³⁰ Furthermore, the section size was decreased from 32 to 28 students, following requests from the faculty and department chair to the dean, based on the data presented above.

(iii) *The use of consistent and validated measures of student learning.* By using the FMCE and BEMA surveys of conceptual mastery, collecting and presenting data such as those included in this paper, various stakeholders (faculty, department chair, dean, and provost) were convinced that these programs were valuable for promoting student learning.

The detailed mechanism of why and how these programs is institutionally supported is left for another study.³¹

Toward sustainability: Intra-institutional replication to nonphysics education researcher faculty

After the second-semester (electricity and magnetism) introductory course was successfully implemented and handed off to a second PER faculty member (Spring 2005), it was handed off to a more traditional but informed faculty member. This faculty member had been serving in the secondary role in the course during the Spring 2005 semester by running the weekly LA-TA tutorial preparation sessions. In the Fall 2005, this instructor, I4, assumed lead role in the course.

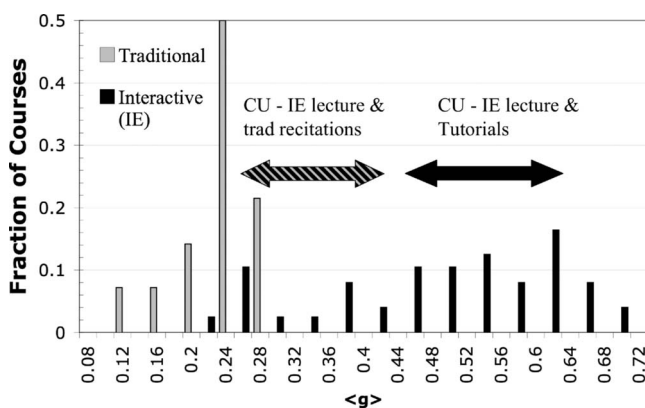


FIG. 3. Plot of normalized learning gains. Data from Hake (Ref. 9) show FCI gains (background bars.) Data for CU are FMCE gains, showing the span of CU interactive engagement (IE) lectures using tutorials (right arrow) and IE lectures with traditional recitations (left arrow).

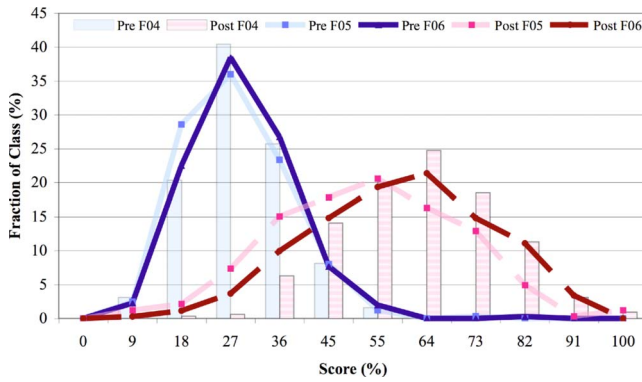


FIG. 4. (Color) Pre- (on the left) and post- (right) BEMA scores for second semester into physics. The bars (shown lightly, in background) are coarsely binned data taken from Fig. 2, representing student performance in classes taught by PER faculty. The line plots are predata (solid) and postdata (dashed) for instructor I4 during two different implementations. The lighter lines are the first implementation: the darker lines are his second time teaching this course.

Instructor I4 was provided all the course materials used by the PER instructor, P2, who led the class the prior term. These materials included lecture notes, concept tests, exams, homework, and tutorial preparation materials. Observations and interviews with this faculty member documented that he primarily used these materials without modification.³¹

Figure 4 plots the prescores and postscores of student performance on the BEMA in this implementation; the Fall 05 results are the lighter colored lines. (Post-test results are shown with dashed lines.) The bar plots (shown in the background) are the same data shown in Fig. 2, more coarsely binned for comparison. In Fall 2006, the same professor, I4, taught this course once again (dark lines). This time, the professor had the assistance of a PER faculty member, P1, serving in the secondary role. The normalized learning gains are 0.33 ± 0.01 (in Fall 2005) and 0.40 ± 0.01 (in Fall 2006).

These should be compared to the 0.44 ± 0.01 gains of the PER instructors previously.

Table II summarizes student performance in the second-semester implementation of tutorials and the associated BEMA post-test scores and normalized learning gains. (Pre-test scores are not shown as they are extremely consistent from term to term, $26\% \pm 1\%$.) Data shown include results for hand-off from one PER faculty member to another (those data shown in Fig. 2), and to informed and traditional faculty members. Notably, there is a high level of fidelity where similar results are seen for the same faculty (Fa04 and Fa07) over a span of several years. As seen in Table I, the most significant variations among semesters are associated with the backgrounds of the instructors.

DISCUSSION

The data from our studies span ten first-semester and seven second-semester courses, roughly 5000 students and 13 total semesters of implementing the tutorials in recitation sessions. From these data, we may begin to draw some conclusions about our abilities to sustain these effective educational practices, the role of instructors, curricular approaches, and some suggestions about faculty development.

We note first that it is possible to hand off these curricular reforms. The initial replication of the UW success was previously documented,⁶ and the subsequent hand-off of this course transformation to another faculty member (Fig. 2) shows remarkable fidelity. The aggregate results of student performance are statistically indistinguishable for the Spring 2005 and Fall 2004 semesters.³³ Notably, this is not a *de facto* result, and other faculty implementing these curricula do not necessarily show similar student learning gains, as seen in Tables I and II. Furthermore, these are aggregate results and we see significant variation within individual questions on the BEMA exam and on individual common exam questions for the two PER-instructor implementations. These

TABLE II. Summary of BEMA scores over seven semesters of data collection. All semesters shown used tutorials in recitation. The standard error of the mean for this table is very close to those in Table I, 1%–2% on post-test, and roughly ± 0.02 for normalized gains.

| Semester | Faculty background lead and/or secondary ^a | N (matched) | Average post-test | |
|----------|--|------------------|-------------------|--|
| | | | score (%) | Normalized gain $\langle g \rangle$ |
| F04 | P1/P2 | 319 | 59 | 0.44 |
| S05 | P2/I4 | 232 | 59 | 0.43 |
| F05 | I4 (Alone) | 325 | 50 | 0.33 |
| S06 | I6/T5 | 196 | 53 | 0.37 |
| F06 | I4/P1 | 351 | 56 | 0.40 |
| S07 | I7/I3 | 261 | 53 | 0.37 ^c |
| F07 | P1/T6 | 161 ^b | 61 | 0.47 |

^aCoding scheme is continued from Table I, with common faculty given the same code, so, e.g., in S05, the secondary faculty, labeled I4 here, is the same person as the lead faculty (I4) in Table I, Spring 2007.

^bIn F07, half the class took the conceptual survey of electricity and magnetism (CSEM) (Ref. 32) test, with similar results.

^cNo pretest was given in S07, but the average pretest has been $26\% \pm 1\%$ every other term, so we used this to estimate the normalized gain for that term.

variations reveal fine-grained differences in content emphasis and student variation in the two semesters. We note that while there was variation in individual faculty practice at a fine-grained level, the implementation of the tutorials was remarkably similar at the level of the situation and the course culture levels. Each of the two instructors, and their respective students, emphasized the importance of sense making, reasoning, and discussion in tutorial sections. The tasks (specific choice of which tutorial chapter to cover) and the situations were the same—students working in groups at the tables knew what the activity was about. The course culture emphasized making sense of the given content and the role of conceptual understanding, what researchers at University of Maryland refer to as framing.³⁴ Which specific messages are sent and how they are sent is subtle and left for future detailed analysis.^{31,35}

While it is possible to hand off curricular reforms with success, it is not guaranteed. There is some suggestion that this has to do with the background of the faculty member, and whether or not these individuals are well versed in teaching practice that is informed by physics education research. In order to implement a course practice, one must attend to specific details of the local environment, institutional specifics that tend to vary with time, and make a myriad of decisions that are not prescribed or even documented— instructors must adapt their approaches and associated curricular practices.¹¹ For faculty members well versed in the field of physics education research, and familiar with the development of the specific innovations, in this case tutorials, this adaptation can happen in a manner that is informed and aligned with the curricular goals. These faculty members can maintain a course culture and create situations supportive of the established pedagogical goals. Here and elsewhere,³⁶ we observe that PER innovations maintain their fidelity when handed off to PER faculty. Notably from the data, Tables I and II, we observe that instances when PER faculty are involved in instruction, in either the lead or secondary role, students post the highest learning gains. When PER-informed faculty, “I’s,” are involved in instruction, students post higher learning gains than when only traditional faculty are involved. An example is seen in Table I by comparing Spring 2006 and Fall 2007 implementations. The same lead instructor, I2, partnering with a traditional (Sp06) or informed (Fa07) secondary instructor yields significantly different results.³⁷ In all cases, we observe the positive impact of the curricula and the interactive nature of the courses, but not the same degree of effectiveness. While students are engaged in the same tasks of working through the tutorials, the meaning of these tasks may be different—the course culture and situations that students engage in shift, in subtle but substantial ways, resulting in differential performance on these conceptual surveys.

At the same time, there is significant indication that the specific curricular practices matter. Of course, the most compelling study would be to control for faculty member and vary the curricula. However, we have not had opportunity to observe the same faculty member teaching in both a traditional and a tutorial course.³⁸ Nonetheless, we do observe that all instances of courses that include traditional recitation sections lead to student normalized learning gains that are

lower than any course that implements tutorials, as seen in Fig. 3. One may argue that these results may arise in part from having a greater number of instructors in the room or that the training and preparatory sessions for the recitation leaders are different. We agree. That is, we believe that part of these innovative curricular practices is an increased teacher: student ratio and structured preparation sections. We also believe the materials themselves, which embody both content, emphasizing concepts and student reasoning, and social practice, group work, and explicit justification, are important. There is a possible counterpoint that the semester (Fa04, first semester course) using Knight’s workbook materials post significant learning gains, comparable to those seen with the University of Washington’s tutorials. We find this result to be intriguing and suspect it to be due to two main factors. First, it is worth noting that the social practices (of students working in groups justifying reasoning) and the content of the workbooks appear to be remarkably similar to the tutorials. In these environments, students were encouraged and observed to work in groups.³⁹ Ultimately, the situations that students engaged in that semester (Fa04) were similar to those of the other semesters that implemented tutorials. Second, the course instructor, while not a member of the PER faculty, had taken it upon himself to attempt an implementation of scale-up.⁵ In addition, to reading about those particular curricular innovations, he visited North Carolina State University and learned about the structure, practice, and philosophy of the reforms. While he was ultimately not successful because of institutional constraints (not finding a suitable room), we believe this faculty member holds a greater understanding of the underpinning philosophy and approach to the tutorials as a result of this attempt. We suspect the course culture that he created was in greater alignment with PER goals than any of the other non-PER faculty.

Finally, there is some evidence that engaging the implementation of these reforms can promote learning—for the faculty. While not certain, the data in Fig. 4 suggest that professor I4 developed expertise at teaching, at least as measured by shifts in student performance on conceptual surveys. It may be the case that I4’s beliefs about his role, the goals for the students and ultimately the course culture shifted as a result of teaching in his first attempt. These are corroborated by interviews with I4 over the course of his teaching in these two implementations.³¹ A caveat to this interpretation is that I4 had additional resources in the form of a PER-versed secondary instructor, P1, in the second implementation—perhaps the greater instructional resources in the second term led to greater learning for the students. Nonetheless, we suspect that this approach of modeling practices may well be effective and sustainable. An instructor new to these practices begins by serving in the secondary role with a well-versed faculty member leading the course (e.g., I4 in Spring 2005), then assumes a lead role for the course (Fall 2005), and refines his or her approach to teaching in these reformed manners (Fall 2006). Such an approach is a modification of the coteaching approach, advocated and studied by Henderson *et al.*⁴⁰

CONCLUSIONS

While it is well known which curricular practices can

improve student performance on measures of conceptual understanding, the sustaining of these practices and the role of faculty members in implementing these practices are less well understood. We have examined the implementation of *Tutorials in Introductory Physics* at the University of Colorado over a four-year span, covering 13 different implementations of tutorials by 15 faculty. We observe that it is possible to demonstrate strong and consistent learning gains for different faculty members; however, such results are not the norm and rely on a variety of other factors. We observe that both faculty background and the particular curricula used in

recitation sections are significant contributors to the variation in student performance from semester to semester.

ACKNOWLEDGMENTS

This work was conducted with the support of the National Science Foundation (CCLI, STEM-TP, CAREER, and CIRT), and the AAPT/AIP/APS PhysTEC. We are particularly grateful to UW PEG, Valerie Otero, Chandra Turpen, the CU Faculty participating and opting to implement tutorials, and the Physics Education Research and Discipline-Based Education Research groups at CU.

-
- ¹E. R. Redish, *Teaching Physics with the Physics Suite* (Wiley, New York, 2003).
- ²L. C. McDermott and E. R. Redish, Resource letter: PER-1: Physics education research, *Am. J. Phys.* **67**, 755 (1999).
- ³K. Cummings, J. Marx, R. Thornton, and D. Kuhl, Evaluating innovation in studio physics, *Am. J. Phys.* **67**, S38 (1999).
- ⁴A. P. Fagen, C. H. Crouch, and E. Mazur, Peer instruction: Results from a range of classrooms, *Phys. Teach.* **40**, 206 (2002).
- ⁵J. D. H. Gaffney, E. Richards, M. B. Kustus, L. Ding, and R. Beichner, Scaling up educational reform, *J. Coll. Sci. Teach.* **37** (2008).
- ⁶N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- ⁷M. S. Sabella, in *Proceedings of the 2002 Physics Education Research Conference*, edited by S. Franklin, K. Cummings, and J. Marx (PERC, New York, 2002), p. 79.
- ⁸S. 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).
- ⁹R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- ¹⁰C. Henderson and M. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020102 (2007).
- ¹¹C. Henderson and M. Dancy, Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations, *Am. J. Phys.* **76**, 79 (2008).
- ¹²L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- ¹³L. C. McDermott, P. S. Schaffer, and M. D. Somers, Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine, *Am. J. Phys.* **62**, 46 (1994).
- ¹⁴N. D. Finkelstein, Learning physics in context: A study of student learning about electricity and magnetism, *Int. J. Sci. Educ.* **27**, 1187 (2005).
- ¹⁵E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- ¹⁶Mastering Physics (<http://www.masteringphysics.com/>).
- ¹⁷CAPA (<http://www.lon-capa.org/>).
- ¹⁸The two authors independently coded faculty as "P," "I," or "T" based on personal experience, classroom observations, and discussions with these faculty. An independent researcher, who is studying faculty change, coded a subset of the faculty in this study with which she was familiar through interviews and observation. All three researchers' coding of faculty were the same, with one exception that was resolved through discussion.
- ¹⁹V. Otero, N. Finkelstein, S. J. Pollock, and R. McCray, Who Is Responsible for Preparing Science Teachers?, *Science* **28**, 445 (2006).
- ²⁰R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula, *Am. J. Phys.* **66**, 338 (1998).
- ²¹L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
- ²²R. Knight, *Physics for Scientists and Engineers: A Strategic Approach* (Pearson, San Francisco, 2004).
- ²³D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics*, 6th ed. (Wiley, New York, 2004).
- ²⁴All faculty implemented peer instruction and an online homework system and had similar weightings of course components with respect to time spent and fraction of grade assigned. However, the *manners* in which these practices were implemented may, in fact, have varied considerably. An ongoing study examines the question of how faculty implement practices, but such an examination occurs at a finer-grain size than the present study, which focuses on curricular choices, and faculty background at a broader scale.
- ²⁵D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- ²⁶Fall 2004 is not included in Fig. 3 as is a hybrid environment in our opinion (discussed in the text). CU data in Fig. 3 (shown as arrows) report learning gains on the FMCE, not the FCI. However, scores on these two instruments have been shown (Ref. 27) to closely correlate (correlation coefficient is $r=0.8$ in a study by the FMCE author), and we confirm this by comparing our CU gains for F03 and Sp04, otherwise nearly identical courses in all respects.
- ²⁷K. Cummings, J. Marx, R. Thornton, and D. Kuhl, Comparing item responses on the FMCE and FCI, *Am. Astron. Soc. Photo-*

- Bull. **209**, 248 (2006).
- ²⁸N. Finkelstein, S. Pollock, and V. Otero, Colorado PhysTEC, American Physical Society, Physics Teacher Education Coalition, 2004–2008 <http://stem.colorado.edu>.
- ²⁹V. Otero, S. Pollock, and M. Klymkowsky, LA-TEST: Learning assistant model of teacher education in science and technology <http://stem.colorado.edu>.
- ³⁰This financial support is part of a broader effort. The CU Learning Assistant Program receives roughly \$150,000 per year from the university to support six departments' use of learning assistants
- ³¹C. Turpen, Studies of Faculty and Institutional Change in Introductory Physics (unpublished).
- ³²D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
- ³³It is worth noting that the prescores and standard deviation suggest both that the instrument is reliable and the student body is consistent.
- ³⁴D. Hammer, A. Elby, R. Scherr, and E. F. Redish, in *Transfer of Learning: Research and Perspectives*, edited by J. Mestre (Information Age, Greenwich, CT, 2005), p. 89.
- ³⁵C. Turpen and N. Finkelstein, Understanding physics faculty's use of peer instruction, *Proceedings of the 2007 Physics Education Research Conference* (AIP, Melville, NY, 2007), p. 951.
- ³⁶S. B. McKagan, K. K. Perkins, and C. E. Wieman, Why we should teach the Bohr model and how to teach it effectively, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010103 (2008).
- ³⁷It may also be argued that I2 learns in the process. In fact, we suspect so and become yet more informed (and support the claim that faculty can learn). However, this is a seasoned faculty member who appears to be engaging in similar approaches to framing and enacting the class practices.
- ³⁸We take this as a sign that once faculty members teach with tutorials, they are convinced of the effectiveness of this approach or at least of the inertia of the system working to sustain these innovations.
- ³⁹While it is possible to run tutorial-style activities with a single instructor, it is not recommended. See, R. E. Scherr, An Implementation of Physics by Inquiry in a Large-Enrollment Class, *Phys. Teach.* **41**, 113 (2003).
- ⁴⁰C. Henderson, A. Beach, and M. Famiano, Diffusion of educational innovations via co-teaching, *Proceedings of the 2006 Physics Education Research Conference* (AIP, Melville, NY, 2007), Vol. 883.