Lessons from a large-scale assessment: Results from conceptual inventories

Beth Thacker, Hani Dulli, Dave Pattillo, and Keith West
Physics Department, Texas Tech University, Lubbock, Texas 79409, USA
(Received 2 January 2014; published 14 July 2014)

We report conceptual inventory results of a large-scale assessment project at a large university. We studied the introduction of materials and instructional methods informed by physics education research (PER) (physics education research-informed materials) into a department where most instruction has previously been traditional and a significant number of faculty are hesitant, ambivalent, or even resistant to the introduction of such reforms. Data were collected in all of the sections of both the large algebra- and calculus-based introductory courses for a number of years employing commonly used conceptual inventories. Results from a small PER-informed, inquiry-based, laboratory-based class are also reported. Results suggest that when PER-informed materials are introduced in the labs and recitations, independent of the lecture style, there is an increase in students' conceptual inventory gains. There is also an increase in the results on conceptual inventories if PER-informed instruction is used in the lecture. The highest conceptual inventory gains were achieved by the combination of PER-informed lectures and laboratories in large class settings and by the hands-on, laboratory-based, inquiry-based course taught in a small class setting.

DOI: 10.1103/PhysRevSTPER.10.020104 PACS numbers: 01.40.Fk, 01.40.G-

I. INTRODUCTION

While the introduction and adoption of physics education research-informed (PER-informed) materials [1] and teaching techniques by both departments and individual faculty are becoming increasingly common, there are still barriers to reform [2–5], and changes can be met with significant faculty resistance [6]. There are still institutions where traditional lecture instruction is the norm and any implementation of research-based materials is done by instructors on an individual basis, without departmental concurrence and often without departmental support. Texas Tech University (TTU) is one such university. It is a large research university where the instruction in physics is primarily traditional and there has not been a unified approach to the teaching of the introductory physics courses. There is, however, a small subset of instructors with an interest in reform. In 2007, an undergraduate committee was formed that decided on the introduction of PER-informed materials into the laboratories and the implementation of recitation sections that would include the use of PER-informed materials. With the support of the Department Chair at the time, even though not all faculty were unified or in agreement, PER-informed materials were introduced into the laboratories and the newly formed recitation sections. We had a situation, then, where changes were implemented in the laboratories and recitation sections, but the lecture instruction remained unchanged. We wanted to assess the impact of introducing PER reforms only in the labs and recitations and also the present state of instruction in the department.

The changes began in Spring 2008 and in Spring 2009 we applied for and were awarded (in Fall 2009) a National Institutes of Health (NIH) Challenge Grant [7] to support a large-scale assessment of the introductory courses and the changes being made in the laboratories and recitation sections. Because of the nature of the changes, our assessment needed to span all of the introductory courses, both calculus based and algebra based. We set out to assess our courses using existing assessment instruments and locally written free-response pretests and post-tests administered in the laboratories and recitation sections. It was a large scale assessment project of size and extent not usually carried out in large physics departments, an assessment of students' understanding and skills in all of the introductory courses based on commonly used research-based assessment instruments.

The main research-based assessment instruments in common use in the introductory courses are multiple-choice conceptual inventories. There are very few, if any [8] valid and reliable comprehensive assessment instruments, research based or not, in general use that are designed explicitly for the university level introductory physics courses. We chose to use four different conceptual inventories: the Force Concept Inventory (FCI), [9] the Brief Electricity and Magnetism Assessment (BEMA) [10], the Mechanics Baseline Test (MBT) [11], and the Conceptual Survey of Electricity and Magnetism (CSEM) [12]. While these only assess conceptual change in certain content areas, this information combined with the results from other assessment instruments gave us an indication of the results of changes we had made.

In addition, these assessment instruments are valid and reliable and the results can be compared to those at other universities. The other assessment instruments we used include scientific attitude and scientific reasoning inventories [13], locally written free-response pretests and posttests, and TA evaluation inventories [14]. In this paper, we report only on the results of the administration of the conceptual inventories. The results of the other assessments will be reported in other papers.

In addition to the changes being made in the large lecture classes, we wished to assess a laboratory-based, inquiry-based course (INQ) [15–18] that was developed with National Science Foundation (NSF) funding about 12 years ago and has been taught as a special section of the algebra-based course every semester since then.

There are many papers that report a significant rise in normalized gain when conceptual inventories are used to assess the use of PER-informed materials in the PER literature. So many that it is easy to forget (1) that simply the introduction of these materials, independent of the institutional environment and contextual factors, does not necessarily produce a significant rise in normalized gain [19,20], and (2) that conceptual inventories assess only one aspect of understanding and are not meant to be, and should not be, used as the sole indicator of the success (or failure) of particular materials or pedagogy. Conceptual inventories are not comprehensive assessment instruments and should not be used that way. However, they do reflect the degree to which an intervention applied within certain environmental and contextual factors has affected aspects of conceptual understanding. And it is useful to compare normalized gain results to those at other institutions, when interventions have been applied in similar or different environments and contexts.

We present our results from the implementation of PER-informed materials in the labs and recitations, with and without PER-informed materials in the lecture in our specific environment, to add to the collection of such results from the assessment of PER-informed implementations as evidenced by conceptual inventories. We also discuss the need for the use of other assessment instruments to develop a broader and deeper understanding of the effect of curricular changes. In our case, we also have results from free response pre- and post-testing and other assessment instruments that assess different aspects of instruction from the same study. The inclusion of those results in a single paper would make the paper much too long, and we report those results elsewhere [21].

The conceptual inventory assessment results are definitely informative and useful to our own institution, but we believe that it will also be informative to those at other institutions in similar situations. There is evidence that large research institutions are less likely to have adopted PER-informed instructional materials and practices [4]. Our results will inform other institutions of (1) the value of the introduction of PER-informed materials into the laboratories and recitations, with and without PER-informed instruction in the lecture, and (2) the value of performing a large-scale assessment across all of the introductory courses in order to inform decisions on instruction and curricular change.

In this paper, we discuss in Sec. II the student populations, the state of the introductory courses being assessed, the changes being made to the courses, and the teaching styles of the instructors, in Sec. III the administration of the assessment instruments, in Sec. IV the results, in Sec. V an analysis of the results, and in Sec. VI, we conclude.

II. THE DEPARTMENT AND STUDENT POPULATIONS

Texas Tech University (TTU) is a large university of about 32 000 students, with 26 000 of them undergraduates. The physics department has 20 tenured or tenure-track faculty and teaches about 2600 students (annual enrollment) in the introductory physics courses each year. This includes the calculus-based and algebra-based introductory physics classes. About 1800 of these students are in the calculus-based course and 800 in the algebra-based course. The introductory courses are usually taught by faculty, but may be taught by postdoctoral researchers, visiting faculty, or even graduate students on occasion.

A. The laboratories and recitation sections

Prior to Spring 2008, the introductory courses consisted of three hours of lecture and two hours of laboratory work each week. The labs were taught by teaching assistants (TAs). They were very traditional, "cookbook," in format and pedagogy and had not undergone significant change in many decades. The students would work through the labs and turn in a formal lab write up to the TA. There was no recitation.

After a transitional semester in Spring 2008, PER-informed laboratories and a one-hour weekly recitation section were introduced into the algebra-based courses in Fall 2008. At first, the labs and the recitations were held in three-hour blocks, with two hours of lab and one hour of recitation. This was done mostly to help with scheduling problems, as the recitations were added in. In the first course of the algebra-based sequence (ABI), the Module I of the Real Time Physics labs [22] was used exclusively. In the second course in the sequence (ABII), some Real Time Physics labs (from Modules 3 and 4) were used and some locally written PER-informed labs were used. By Fall 2010, the ABII labs were almost completely locally written PER-informed labs. They did not require a formal lab write up, but included laboratory homework. There were also biweekly quizzes in the recitation sections that included material from lecture, lab, and/or recitations.

Beginning in Spring 2009, a one-hour recitation was also implemented in the calculus-based courses, which we will refer to as calculus-based I and II (CBI and CBII). The CBI labs used some of the Real Time Physics laboratories and some of the traditional laboratories. The CBII course remained traditional labs. The labs in CBI remained partially Real Time Physics and partially traditional and the labs in CBII remained traditional until Fall 2010. Starting in Fall 2010, up to the present, the labs in the second course in each sequence (ABII and CBII) were almost completely locally written PER-informed labs. The labs in the first course in the calculus-based sequence used Real Time Physics labs exclusively in Fall 2010. After Fall 2010 to the present, PER-informed labs developed at the University of Illinois [23] were used. The University of Illinois labs were designed as part of the reform of their introductory courses and were designed as an adaptation of the approach of Real Time Physics [22], designed to address common misconceptions through active engagement of the students in the learning process.

With the introduction of PER-informed labs and recitation sections, the TAs were trained in different pedagogies than were used in the traditional labs. The traditional lab pedagogy had been random and varied, being left up to the individual instructors of each course, but were usually instructor-centered, "sage-on-the-stage" strategies. Most of the PER-informed labs were designed in a format that required interactive engagement (IE) during the lab to help guide the students. The labs did not require a formal write up, but included laboratory homework. There were also biweekly quizzes in all of the recitation sections that included material from lecture, lab, and/or recitations.

1. PER-informed labs

The locally written PER-informed labs were designed by instructors with knowledge of PER literature, other PER-based instructional materials, and pedagogical content knowledge. The labs consisted of five parts: Objectives, Overview, Explorations, Investigations, and Summary.

- (•) *The Objectives* listed the concepts and skills the students should understand and be able to demonstrate after completing the lab.
- (•) *The Overview* was a short summary of the purpose of the lab.
- (•) *The Exploration* section was the part of the lab designed to address common student difficulties and conceptions by posing appropriate questions to elicit, confront, and resolve the difficulties through questioning and discussion with other students and the TAs. The *Explorations* were qualitative measurements or, sometimes, qualitative problems or thought experiments designed to focus on concepts the students often have difficulty with, even after instruction. They were designed to focus on known difficulties and conceptions in an experimental setting, allowing students to make observations that might challenge or contradict their present conceptual understanding, and allow them to reshape their conceptual understanding through thought and discussion. The concepts were also chosen to support or relate to the *Investigation* section of the lab.
- (•) *The Investigation* section of the lab consisted of quantitative measurements and observations, taking data, graphing, analyzing, and interpreting it. Care was taken to elicit predictions before data were taken and to guide the data taking, graphing, and analysis, not being overly prescriptive, but allowing student thought input at each step. Still, it is the section that is more like a traditional lab.
- (•) In the *Summary*, the students were asked to focus on a particular part of the lab and summarize it in their own words.

There was also a lab homework to be completed and turned in at the next lab, but no formal lab report. A sample lab is included in Supplemental Material Appendix I [24]. Other examples of locally written PER-informed labs can be found on a curriculum development webpage [25].

2. Recitation sections

The recitation sections were about 50 min long and were usually group problem solving sessions monitored by the TA. The problems were chosen by the lab coordinator (s) and were often chosen from or were modified versions of published PER-informed problems, such as problems from Tasks Inspired by Physics Education Research (TIPER) [26], Ranking Task Exercises in Physics [27], books by Arnold Arons [28,29], and other sources. Sometimes the problems were textbook problems or modified textbook problems. The problems were chosen to be on content that had already been covered by all of the instructors teaching the course. The problems were chosen to cover concepts or skills that students often struggle with, even after instruction.

The students would work through the problems in groups, working on whiteboards, with the TA circulating, asking students questions, or answering questions from students. After students had a significant amount of time to work on the problem, the TAs checked on the students' understanding in different ways. Some TAs worked with groups individually, checking on their results both as they worked and as they finished, asking them to explain their results and asking further questions, as needed. Others called the class together and had groups present at the board and had a class discussion about the problems.

If there was time after the problem(s) for that week had been finished, the TAs entertained questions on homework or other questions students might have. The biweekly quizzes were also administered during the recitation sessions.

3. TA training

The TAs were trained and directed by the lab coordinator(s). Previous to the introduction of the PER-informed materials, the lab format was very traditional. The lab coordinator(s) were one of the instructors teaching a lecture section of the course. They met with the TAs weekly to "go over" the lab, which may or may not have included actually working through the lab. TAs would start the lab by giving a short presentation or summary of the lab, including important aspects of the experiment and the results that needed to be recorded and reported, as well as the expected format of the lab report. There was usually not a discussion of pedagogy or student conceptions and the instructional format was pretty much left up to the TAs. Lab reports were turned in at either the end of the lab or the next week, depending on the instructor. The TA training and lab format was quite varied and dependent on the instructor serving as lab coordinator in each course. There was no recitation section.

With the introduction of PER-informed labs, it was necessary to train the TAs in student-centered pedagogy. The lab TAs would spend time at their weekly meeting working in groups through the laboratory they would be teaching the following week, with the lab coordinator(s) modeling the teaching methods to be used by the TAs. In the Exploration parts of the lab, in particular, the TAs were taught to guide the students through questioning and discussion, eliciting and challenging the students conceptions and helping them to restructure their conceptions.

In the recitation sections, the TAs would spend time each week working in groups through the recitation materials they would be teaching the following week, with the recitation instructor circulating, asking questions and checking on their understanding. The instructors modeled the pedagogy to be used by the TAs. The TAs were taught to guide the students through questioning, not "telling" answers, but helping students to think through the questions themselves. They were taught how to help each group and to make sure everyone in the group contributed and was responsible for their own understanding. They were also taught how to guide whole class sessions, having groups or students present at the board and then lead class discussions. The use of the interactive-engagement methods was expected of them both in the recitations and in the Exploration parts of the laboratories.

The TAs received not just more training, but training in student-centered methods and pedagogical content knowledge with the transition to PER-informed materials. One could ask if that same training with the older materials would have produced a similar effect on the assessment outcomes. We do not have that data, but we will comment that the PER-informed laboratory materials themselves are aligned with student-centered pedagogy, while the prescriptiveness of the old labs was very aligned with a teacher-centered strategy. It would be hard to teach a new pedagogy if it was misaligned with the materials used.

B. The faculty

The majority of physics faculty members teach traditionally in a lecture-style format. Very few use PER-informed pedagogy or IE techniques. They focus primarily on the lecture and leave the recitations and laboratories to the TAs and lab coordinator(s). Although the labs and recitations were part of the course, the labs and recitations together were sometimes allotted as little as 10% of the grade. The lower allotments of the percentage of the grade for lab and recitation together were primarily in the calculus-based classes. However, some of the instructors allotted 20%–30% of the grade in those classes for lab and recitation (together). In the algebra-based classes, a higher percentage of the grade was allotted to the labs and recitation sections, with 20% the most common, although they ranged from 15% to 25%.

A few instructors interacted with the TAs in lab and recitations, contributing to the training of the TAs, the choice of materials, and content to be covered in recitations and the pedagogy to be used in lab and recitation. Most of the instructors who actively participated in the TA training were instructors who used PER-informed materials and instructional techniques in the lecture.

The instructors labeled by PER in this paper used PER-informed materials and teaching methods in the lecture.

C. The students

1. Calculus-based courses

Large lecture sections of the calculus-based course.—The calculus-based course consists primarily of engineering and computer science majors. The number of students registered for CBI, the first course in the sequence, each semester, is usually around 500, split among three lecture instructors. The number of students in the second course in the sequence, CBII, is around 400, split among two or three instructors. The instruction is primarily traditional lecture, with one one-hour recitation section and one two-hour lab, as described above. The labs and recitations are common among the three instructors each semester. Students from each of the lecture instructors are mixed in the labs and recitations.

Honors section.—There is one honors section of the calculus-based class that is taken by students in the TTU Honors College and by some of the physics majors. It is usually a small class, consisting of 10–24 students. Sometimes the honors students take the same laboratories as the large lecture sections and sometimes they do not, depending on the instructor. We have listed the scores that we have for honors students who did take the same laboratories as the students in the large lecture sections with traditional lecture instruction. We also list the data for one honors section that worked through locally written PER-informed labs based on Workshop Physics [30] and other PER materials combined with PER lecture instruction, separate from the other sections. That course had an integrated lab and lecture format. The number of students in these sections is small and we hesitate to draw significant conclusions from the data because of the small number of students assessed. The results from the honors sections are included for completeness.

2. Algebra-based courses

Large lecture sections of the algebra-based course.—The algebra-based class consists mostly of prehealth science majors, including premedical, predental, prephysical therapy, etc. The number of students registered each semester in the first course in the sequence is usually around 250–300 and has been around 100–150 in the second course in the sequence in recent semesters [31]. Except for the inquiry-based section of the course, the students are divided into two lecture sections taught by two lecture instructors each semester. The instruction is primarily traditional lecture, with one one-hour recitation section and one two-hour lab each week. The labs and recitations are common among the three instructors. Students from each of the lecture instructors are mixed in the labs and recitations.

Inquiry-based, laboratory-based section.—In addition to the changes being made in the large lecture classes, we wished to assess a laboratory-based, inquiry-based course [15–18] that was developed with National Science Foundation (NSF) funding about 10 years ago and has been taught as a special section of the algebra-based course every semester since then. It was developed explicitly for health science majors, taking their needs, learning styles, backgrounds and motivations into account. It is taught without a text in a Workshop-Physics style [30] environment and is an inquiry-based course in the manner of *Physics by Inquiry* [32], developed by the physics education group at the University of Washington, but at the algebra-based level. The materials were developed by modifying and adapting parts of existing materials designed for other populations and integrating them with new units in our own format, creating a course aimed specifically at health science majors.

The curriculum was designed to be taught in a laboratory-based environment with no lecture and no text; however, a text can be used. Students work through the units in groups, learning to develop both quantitative and qualitative models based on their observations and inferences and then using the models to make predictions and solve problems. The materials consist of the laboratory units, pretests, readings, and exercises. There are also homework sets, exams, and quizzes. The course covers approximately the same content as is covered in the other sections of the class, but with more of a focus on developing models based on experimentation and developing observational, analytic, and critical thinking skills in order to design experiments and work problems. The students sign up on a first-come, first-serve basis.

The FCI, MBT, BEMA, and CSEM were also administered to these students every semester, starting in Spring 2010.

III. ADMINISTRATION OF ASSESSMENT INSTRUMENTS

The conceptual inventories were administered as pre- and post-tests in the recitations over the course of this study. They were administered as a pretest at the beginning of the semester and as a post-test at the end of the semester. Students were allotted 45 min to take the assessments. The FCI and BEMA were administered as a pre- and a post-test every semester starting in Fall 2009, except one semester (Spring 2011) when BEMA was administered as a post-test only. We also have FCI data from semesters prior to Fall 2009 for select classes. The MBT was administered online as a pre- and a post-test in Spring 2010 and as a post-test only in Fall 2010 and Spring 2011. The CSEM was administered as a pre- and a post-test in Spring 2010 and as a post-test only in Fall 2010 and Spring 2011.

Taking the assessment counted as part of the students' laboratory or recitation participation grade. Depending on the class, one to three points were deducted from their participation grade if they did not take the assessment. In addition, starting in Fall 2010, up to five points toward their laboratory or recitation grade were awarded to students, based on their performance on the assessment. The number of points was determined by the percentage correct on the post-test. The students were not told their score and the assessment was not discussed with the students. They were simply told the number of extra points they received, if they asked. For most students this was three points or less. This constituted not more than 1% of their total course grade. We report any difference in the students' scores that could be accounted for by the year the assessment was taken. We present some of the ABI and CBI results chronologically in the next sections to address this concern.

The online administration of MBT and CSEM was terminated after one semester, Spring 2010, due to concerns about the efficacy of online testing, as well as concerns about too much assessment in the introductory courses.

IV. RESULTS

We report results for each of the four conceptual inventories and present post or normalized gain scores in Figs. 1–7. A more detailed version of this section's results and analysis can be found in Supplemental Material Appendix II [24]. The detailed version includes pretest, post-test, and

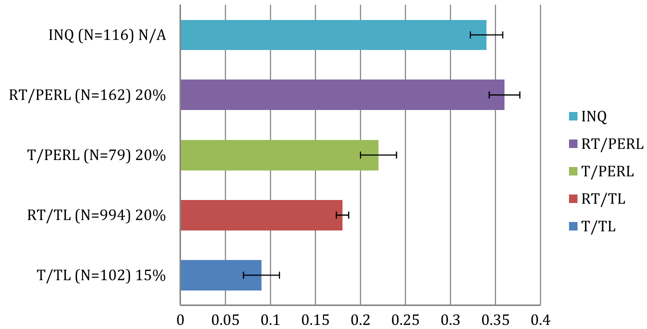

FIG. 1. Algebra-based FCI normalized gain by laboratory and teaching style. The data are listed by *laboratory/teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an H.

normalized gain scores, additional tables, figures, and details of analysis. The normalized gain [33] g was calculated using $g = (\text{postscore} - \text{prescore})/(100 - \text{prescore})$.

The error bars represent the standard error of the mean.

A. Force Concept Inventory

The normalized gain results for the FCI for the algebra-based and calculus-based courses are presented in Figs. 1 and 2. We have combined all of the data by laboratory and teaching style and present the means and standard error for each lab and lecture teaching style. For the algebra-based course, data from Fall 2006 through Fall 2011 are included, but the FCI was not administered every semester until Fall 2009. For the calculus-based course, the data includes Fall 2008 through Spring 2012, except for Spring 2009. We report the normalized gain only for those students who took both the pretest and post-test. The percent of total grade allotted to the combined lab and recitation part of the course is also given.

The data in the graphs is labeled by *laboratory to teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an H.

The FCI data from traditionally taught labs (T) were collected before the recitation sections were introduced. FCI data from all other lab styles was taken after the recitations were implemented.

1. Algebra-based FCI

For the algebra-based course, we have FCI data with traditional (T) and PER-informed Real Time Physics labs (RT) with both PER-informed (PERL) and traditional (TL) lecture teaching styles. This gives us information on the effectiveness of the RT labs and recitation compared to T labs and PERL vs TL lecture styles.

For comparison with other scores across the country, we refer to a large survey paper by Hake [33]. In that paper, it is reported that most students taught in a traditional lecture format have normalized gain scores of about 0.15 and students taught in an IE format generally have scores in the 0.30–0.60 range, known as the IE region on a “Hake” plot. Most PER-informed materials, such as those listed in the PER User’s Guide [34], employ IE methods and are also designed to address many of the alternative conceptions found on the FCI, so it is also expected that the use of PER-informed materials will result in an increased normalized gain.

The distributions were determined to be normal based on histograms of the data and we used a student’s T test to determine if the data were significantly different. All of the results in the figures are significantly different for comparison of data with nonoverlapping error bars at the $p < 0.05$ level by a student’s T test for all of the algebra-based scores.

From the algebra-based data, it is very clear that the introduction of RT labs significantly increases the gain on the FCI, independent of the lecture style. However, the gain is greater for the PERL lecture style. It is also clear that a change in lecture style only from TL to PERL increases the FCI scores significantly, the gain being greater for RT labs.

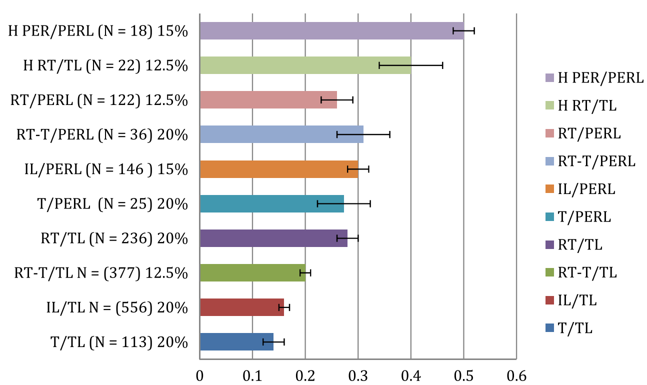

FIG. 2. Calculus-based FCI normalized gain by laboratory and teaching style. The data are listed by *laboratory/teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an H.

It is only with the implementation of both RT labs and recitations and PERL lecture instruction that the FCI normalized gain is above 0.30 and in the interactive-engagement region. The mean FCI normalized gain for the inquiry-based (INQ) class is 0.34 ± 0.02 and not significantly different from that for the large lecture sections when both RT labs and PERL instruction are implemented. Both the INQ class and the large lecture sections with PERL instruction in the lecture and RT labs and recitation sections demonstrate the effectiveness of research-based instruction.

An important point here is that the FCI scores, while increased 100% above the T and TL scores when RT labs and recitations were introduced without changing the lecture instruction (RT and TL), are still closer to traditional scores than to the IE region. This is important information for universities who try to implement PER-informed techniques in the labs and recitations only, without faculty concurrence to change instructional methods in the lecture also. The gains are not as large as they would be if faculty would overcome their hesitancy and resistance to the adoption of PER-informed techniques in the lecture.

2. Calculus-based FCI

The data from the calculus-based classes is more complex. The transition to PER-informed labs in the calculus-based classes was hindered and very much opposed. The first introduction of PER-informed labs was part RT and part traditional labs (RT-T). RT labs only were run for only one semester before the introduction of labs developed at the University of Illinois (IL). So there are four lab styles T, RT-T, RT, and IL that have been used in the large lecture classes. Each of the lab styles has been run with both PERL and TL teaching techniques.

In addition, we report scores for the honors and majors sections when they took the same labs as the large lecture sections and for one honors section taught PER or PERL, separate from the other sections. While, as stated previously, we hesitate to draw significant conclusions from this data because of the small number of students assessed, it is interesting that (1) the normalized gain of the honors class with PER-informed instruction is higher than the traditionally taught honors classes, as is true with the regular classes, and (2) that the honors section normalized gain is higher than that of the regular classes, independent of instructional methods. We don't have an explanation for the second result, except to suggest that we suspect it has to do with student expectations and motivations and needs further study. We discuss this also as a possible explanation for the differences in the algebra-based and calculus-based scores in Sec. V, Discussion.

We also indicate the percent of the total grade allotted for lab plus recitation because it varied somewhat, from 12.5% to 20%. However, we do not think we can draw any conclusions about the effect of the percentage grade allotted for lab and recitation, as we can in the second semester BEMA data presented in Sec. IV C 2.

The large lecture sections barely achieve normalized gains of 0.30 with PERL instruction (any type of lab) or RT labs with TL instruction. The PERL data and the RT or TL instruction data are not significantly different. The T, IL, and RT-T labs with TL instruction are not significantly different from each other by a student's T test at the $p < 0.05$ level, but they are each significantly different from the RT or TL labs at the $p < 0.0001$ level.

B. Mechanics Baseline Test

The MBT was also administered as part of this study. However, it was not administered every semester and was sometimes administered only as a post-test or only online. It was administered online in Spring 2010 and only as a post-test in Fall 2010 and Spring 2011.

The post-test results are presented in Figs. 3 and 4, with the online results separated out. The online results were lower for the TL lecture classes, but not for the PERL lecture classes. Possible reasons for this are discussed in the Supplemental Material [24], along with statistical details.

There is not as much published data for comparison to scores at other universities as there is for the FCI. Students have been reported to score on the order of 15% points lower on the MBT than the FCI and it has been considered to be a harder test and to have more problem solving in it, as it requires some math skills and some critical thinking skills [11].

1. Algebra-based MBT

The online scores differ on the order of 10%–15% points between in-class and online testing for the large lecture sections. However, the scores for the INQ course are not significantly different online and in class. These differences

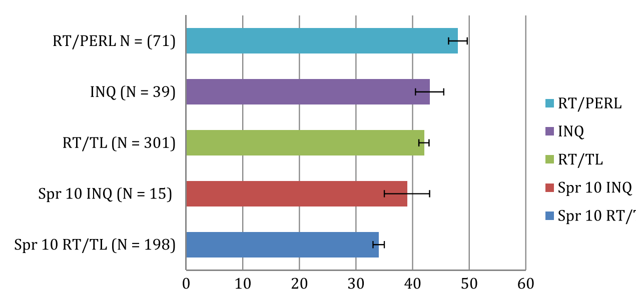

FIG. 3. Algebra-based MBT post by laboratory and teaching style. The data are listed by *laboratory/teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an H.

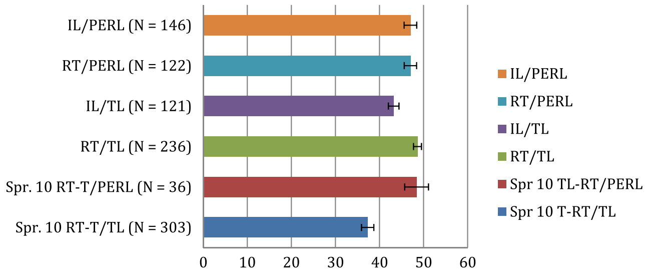

FIG. 4. Calculus-based MBT post by laboratory and teaching style. The data are listed by *laboratory/teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an *H*.

are addressed in the Supplemental Material [24]. There is not much comparison data for the algebra-based course. Scores as high as the high sixties have been observed in algebra-based classes using Peer Instruction (PI) [35] at Harvard University [36].

2. Calculus-based MBT

As with the algebra-based scores, the RT-T or TL students who took the assessment online scored about ten points lower than students in any of the classes who took the assessment in class. Discussion of this is in the Supplemental Material Appendices [24]. The scores of students who took the assessment in class are for the most part in the high forties. This is at the low end of scores published nationally [11,36], which range from the forties to the high seventies. Classes taught traditionally fall on the lower end of that scale and classes taught by Peer Instruction (PI) at Harvard University fall at the high end of the scale. The TTU scores are seven or eight points below the TTU FCI scores.

C. Brief Electricity and Magnetism Assessment

BEMA was administered as a pretest and a post-test every semester from Fall 2009 through Spring 2012 in the calculus-based course, except for Spring 2011, when it was administered as a post-test only. In the algebra-based course, it was administered as a pre- and a post-test every semester from Fall 2009 through Fall 2011, except for Spring 2011, when it was administered as a post-test only.

In the Spring of 2011, in the large lecture sections in the algebra-based course, we had evidence of cheating on the BEMA post-test in the form of a TA talking with students during the assessment and some students with identical high grades, inconsistent with the rest of their work. We removed all of the data from that TA's sections and obvious

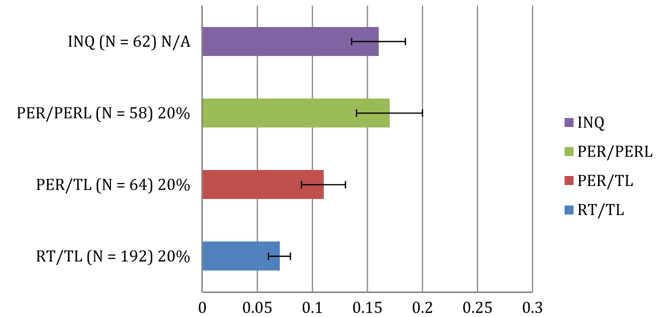

FIG. 5. Algebra-based BEMA normalized gain by laboratory and teaching style. The data are listed by *laboratory and teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an *H*.

cheating from other sections. However, we do not know how widespread the cheating was and if we have removed all of it, so we have chosen not to present the data from Spring 2011 for the algebra-based course.

The normalized gain results for BEMA are presented in Figs. 5 and 6. Other graphs and data can be found in the Supplemental Material Appendices [24]. We have combined all of the data by lab and teaching style and present the means and standard error. We report the normalized gain only for those students who took both the pretest and post-test. The percent of total grade allotted to the combined lab and recitation part of the course is also given. This is particularly relevant in the calculus-based classes.

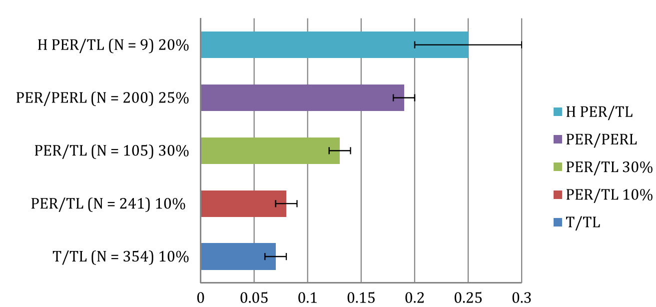

FIG. 6. Calculus-based BEMA normalized gain by laboratory and teaching style. The data are listed by *laboratory or teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an *H*.

1. Algebra-based BEMA

In the algebra-based classes, we had already been using RT labs with some locally written PER-informed labs when we began the assessment using BEMA. We do not have a comparison to the algebra-based courses taught with traditional labs. The results are not particularly high, with the highest gain at 0.17 and the highest post-test at 35.7% (post-test details in Supplemental Material Appendices [24]). The locally written, PER-informed labs (PER) have a higher gain and post-test than the RT labs with TL instruction, and even greater with PERL instruction. We do not have the RT labs with PERL instruction for comparison. The INQ and PER or PERL scores are not significantly different.

There are not a lot of comparison scores in the literature for the algebra-based course. Most of the research using BEMA has been done with calculus-based classes. Typical post-test scores reported for calculus-based students are in the 40%–50% range for traditionally taught students and around 60% for students taught nontraditionally with research-based materials [37–39]. We have found one algebra-based score of 0.38 gain and 51% post-test score posted on a Physics Teacher Education Coalition (PhysTEC) website [40]. While our scores are not particularly high, they have increased in the large section algebra-based classes, as we introduced locally developed PER-informed laboratories and used PER-informed instruction in the lecture.

2. Calculus-based BEMA

In the calculus-based classes, we have data with traditionally (T) taught labs and locally developed PER-informed laboratories (PER). The distributions were determined to be normal based on histograms of the data and we used a student's T test to determine if the data were significantly different. All of the results in the figures for comparison of data with nonoverlapping error bars are significantly different at the $p < 0.005$ level by a student's T test. Since we have a record of the percentage the lab and recitation together counted towards the total course grade, it is interesting to examine the difference between lab and recitation counting as 10% or 30% of the grade. The BEMA scores with T labs were not significantly different from the PER labs when the labs plus recitation were allotted 10% of the total grade with TL lecture instruction. If the percentage grade allotted to the labs and recitation was raised to 30%, the PER-BEMA scores were higher and significantly different from the T scores with 10% allotted to labs and recitation and TL instruction. With PER, instead of TL lecture instruction, the gain rises to close to 0.20 and the post-test to close to 40% (post-test details in Supplemental Material Appendices [24]). These scores are consistent with scores reported for traditionally taught students at other universities across the country [37–39]. The honors physics class has a gain of 0.25 and a post-test of 43%.

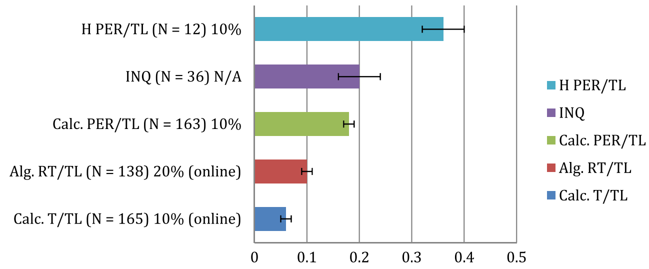

FIG. 7. Calculus and algebra-based CSEM normalized gain by laboratory and teaching style. The data are listed by *laboratory or teaching style* ($N =$ number *of students*) *percentage total grade allotted to laboratories plus recitation*. Lab styles are labeled by traditional (T), Real Time Physics (RT), combination RT and T (RT-T), developed at the University of Illinois (IL), and locally written PER-informed (PER). The lecture teaching styles are labeled by traditional lecture (TL), PER-informed lecture (PERL), and inquiry-based instruction (INQ). Honors sections are labeled with an H.

D. Conceptual Survey of Electricity and Magnetism

The CSEM was administered as a pretest and a post-test in Spring 2010 and Spring 2011 and as a post-test only in Fall 2010. The Spring 2010 assessment was administered online and we do not know if that was the reason for lower scores. We have chosen not to present the Spring 2011 data, due to the issues with cheating, as discussed with the Spring 2011 BEMA scores. We also do not have CSEM scores with PER instruction and we do not have CSEM scores with 30% of the total grade allotted to labs and recitation. As a result, our CSEM scores are not comprehensive. There has been research demonstrating statistically indistinguishable gains on BEMA and CSEM [41], so it is plausible that the CSEM results in categories not assessed would be similar, but we do not know that for sure.

We present the normalized gain scores in Fig. 7, including all of the available algebra-based and calculus-based scores on one plot. In this case, the online results were lower and the other results, calculus-based or algebra-based, were not significantly different from each other, with scores close to a gain of 0.2. The honors students scored higher. Graphs of the post-test scores and comparisons to common post-test scores for students nationally [42,43] can be found in the Supplemental Material Appendices [24].

V. DISCUSSION

This project is important because it provides data on the introduction and implementation of PER-informed materials into the labs and recitation sections at a large university and the results of a large-scale assessment across all of the introductory courses at a large university. We also have these results under the conditions of faculty not unified and in agreement on the implementation of PER-informed materials, with most faculty teaching traditionally in the lecture portion of the course. Because there were a few course instructors who did use PER-informed materials in the lecture and had significant interactions with the lab coordinator and TAs in the labs and recitations, the study reflects the impact of changes when PER-informed materials are introduced in the laboratory and recitation, with and without changes in the lecture part of the course.

A. Major findings

The data from the FCI in the algebra-based course clearly demonstrate an increase in conceptual understanding, as measured by the FCI, due to the implementation of PER-informed materials and instruction in the laboratories and recitations, and due to changes from TL to PERL teaching methods in the lecture. Changes in the laboratories and recitations only or changes in the lecture only, both significantly increase normalized gains above T or TL instruction, with the change to PERL instruction in lecture only having a somewhat greater effect than changes in the labs only. Only with the implementation of PER materials in both the lecture and the laboratory and recitations are normalized gains above 0.30, in the IE region on a Hake [25] plot observed in the large lecture classes.

Results from the calculus-based classes are similar; the normalized gain increases with either implementation of PER-informed materials and instruction in the laboratories and recitations or changes from TL to PERL teaching methods in the lecture. In this case, the completely PER-informed laboratories (RT) with traditional instruction increased the conceptual understanding as much as PERL instruction in the lectures with any kind of lab. However, unlike the algebra-based course, the combination RT and PERL was not significantly higher than either of those changes independently. The PER or PERL honors section, taught with integrated lab and lecture instruction and completely separate from the other sections, had the highest normalized gain on the FCI.

The FCI gains in the large lecture algebra-based classes with PER-informed instructional techniques in both lab and recitation and lecture were at least as high or higher than in the calculus-based courses (except for the honors sections). As the implementation of the interventions was similar in the two groups, it is worth further research into the reasons for the differences to see if other factors, such as goals, learning styles, expectations, or motivations, play a role.

The data from BEMA indicate the same increase in normalized gain that we see with FCI, as PER-informed labs are added and with PERL instructional methods in both the lecture and the laboratories. However, neither the algebra-based nor the calculus-based classes achieve more than 0.20 in normalized gain or above 40% on the post-test, except for the honors sections. The scores, for the most part, are significantly below other published scores from universities introducing new curricula, such as instruction from the Matter and Interactions [44] curricular materials [37–39].

We did, however, in the BEMA study, see the impact of the percentage of the total grade allotted to the laboratories and recitation sections. With TL instruction in the lecture the changes to the labs and recitation sections made a significant difference, if the labs and recitations accounted for 30%, as opposed to 10%, of the total course grade.

The MBT data are fairly flat and the CSEM data all fall within the ranges seen across the country.

B. On the use of conceptual inventories and further assessment

In this paper, we have presented the results of conceptual inventories as indicators of the effectiveness of changes made in the curricula and instructional techniques in the laboratories and recitations separately and together with changes made in the lecture part of the course. The conceptual inventories are designed to assess conceptual understanding and it is reasonable to expect that PER-informed changes to instruction would result in increased gain in conceptual understanding. However, one would hope that PER-informed changes would result in changes in many skills, from lab skills to computational skills, to problem solving and critical thinking skills, not just changes in conceptual understanding. One might argue that conceptual inventories, while giving significant information on changes in conceptual understanding, are not the best instruments to assess some of these other skills.

Based on conceptual inventory scores, the inquiry-based class and the large lecture classes with PER-informed labs and PERL lecture instruction performed at the same level. Based only on conceptual inventory data, the two methods of instruction lead to identical results. But is this the whole story? How would the two methods of instruction compare if other assessment instruments were used to assess other skills?

Too often, conceptual inventory results are presented as if they are comprehensive results, the main factor in determining whether instruction has been effective or not. We believe that this use of conceptual inventory results is unfortunate and that different and more comprehensive assessment instruments need to be developed by the PER community. We support researchers developing assessments that go beyond conceptual inventories, assessing problem solving, lab skills, and other important aspects of instruction [45,46].

In our project, we also administered a series of free-response pretests and post-tests in the labs and recitations over four semesters. The questions required written explanations, covered lab, recitation, and lecture concepts and also assessed lab skills and problem solving. They were administered biweekly and were not comprehensive. They did, however, give us snapshots of students' understanding and abilities throughout the course. This research gave us different information on the students' abilities when different instructional methods were used. The results of that research are presented in other papers [21].

VI. CONCLUSIONS

We conclude that when PER-informed materials are introduced through the labs and recitations, independent of the lecture style, in a large university setting, there is an increase in students' conceptual understanding, as measured by PER developed conceptual inventories. There is also an increase in the results on conceptual inventories, if PER-informed instruction is used in the lecture. The highest normalized gains were achieved by the combination of PER-informed lectures and laboratories in large class settings and by a hands-on, laboratory-based, inquiry-based course and a PER or PERL taught honors section, both in small class settings. We hope that these results will motivate change at our own and similar institutions and be informative to PER researchers studying barriers to change and those working on assessments beyond conceptual inventories.

ACKNOWLEDGMENTS

We thank Kelvin Cheng and Amy Pietan for their work on the statistical analysis that contributed to this paper. We also thank the National Institutes of Health (NIH) for their support of this project, for the funding of NIH Challenge Grant No. 1RC1GM090897-02.

---

[1] The results of physics education research are not always introduced into curricula by taking one set of materials developed and implementing only that set of materials in the exact manner as the developers intended. It is common for instructors to adopt pieces of curricula or materials or instructional strategies and interweave those pieces into their own course. Sets of materials have also been developed locally by PER practitioners based on their knowledge of the field (their knowledge of the research based on talks and workshops presented at national meetings, their own research, and the literature). We operationally define PER-informed instructional materials and pedagogy as those developed by instructors well informed and well read in PER or researchers in PER that may include pieces of published PER-based curricula, materials, or instructional strategies, problems from those materials or from PER-based texts or literature and PER-based strategies that have been demonstrated to be effective in some context. We use the term PER informed because it is not the adoption of a single PER-based curricula or instructional method, but a compilation of materials locally by instructors well read and informed in the use of various PER-based curricula, materials, and instructional strategies.

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

[3] C. Henderson and M. Dancy, Impact of physics education research on the teaching of introductory quantitative physics in the United States, Phys. Rev. ST Phys. Educ. Res. **5**, 020107 (2009).

[4] C. 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).

[5] C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj, Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process?, Phys. Rev. ST Phys. Educ. Res. **8**, 020104 (2012).

[6] J. Tagg, Why does the faculty resist change?, Change: The Magazine of Higher Learning, **44**, 6 (2012).

[7] National Institutes of Health (NIH) Challenge Grant No. 1RC1GM090897-01, An Assessment of Multimodal Physics Lab Intervention Efficacy in STEM Education, to assess for interventions to the laboratory curriculum at Texas Tech University, PI's Beth Thacker and Kelvin Cheng.

[8] We are not aware of any comprehensive assessment instruments in general use designed for the undergraduate introductory physics courses. Instruments do exist for other populations, such as Advanced Placement (AP) physics exams for high school students, the Graduate Record Exam (GRE) for graduate students, and the Major Field Test for Physics for physics majors. We are interested in the development of such an examination, but designed for the purpose of providing feedback to instructors about their students' performance and use in assessing changes to courses and curricula.

[9] I. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53**, 1043 (1985).

[10] L. 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).

[11] D. Hestenes and M. Wells, A mechanics baseline test, Phys. Teach. **30**, 159 (1992).

[12] D. Maloney, T. O'Kuma, C. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).

[13] In particular, we used the Colorado Learning Attitudes about Science Survey (CLASS), W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring

student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **2**, 010101 (2006); and The Classroom Test of Scientific Reasoning (CTSR), A. E. Lawson, The development and validation of a classroom test of formal reasoning, J. Res. Sci. Teach. **15**, 11 (1978).

[14] We used the Reformed Teaching Observation Protocol (RTOP), D. L. MacIsaac and K. A. Falconer, Reforming physics education via RTOP, Phys. Teach. **40**, 479 (2002).

[15] B. Thacker, A. Diaz, and A. M. Eligon, The Development of an Inquiry-based Curriculum Specifically for the Introductory Algebra-based Physics Course, arXiv:physics/0702247.

[16] J. Wilhelm, B. Thacker, and R. Wilhelm, Creating constructivist physics for introductory university classes, *Electronic J. Sci. Educ.* **11**, 2 (2007).

[17] National Science Foundation—Course, Curriculum, and Laboratory Improvement Grant CCLI #9981031, Workshop Physics with Health Science Applications, for development of materials for the introductory, algebra-based physics course with a health science based content focus in an interactive, computer-based laboratory setting, PI's Beth Thacker and Anne Marie Eligon.

[18] National Science Foundation—Course, Curriculum, and Laboratory Improvement Grant CCLI-EMD #0088780, Humanized Physics—Reforming Physics Using Multimedia, and Mathematical Modeling, for development of materials for the introductory, algebra-based physics course with health science based content using mathematical modeling, and multimedia, co-PI's R. G. Fuller, V. L. Plano Clark, B. Thacker, N. L. Beverly, C. D. Wentworth, and M. W. Plano Clark.

[19] M. H. Dancy and C. Handerson, Beyond the individual instructor: Systemic constraints in the implementation of research-informed practices, AIP Conf. Proc. **790**, 113 (2005).

[20] N. 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).

[21] B. Thacker, K. West, G. Chapagain, V. Rusuriye, and H. Dulli, Lessons from a Large-Scale Assessment: Results from Free Response Pre- and Post-testing in Electricity and Magnetism, *Phys. Rev. ST Phys. Educ. Res.* (to be published); B. Thacker, G. Chapagain, M. Ellermann, D. Pattillo and K. West, The Effect of Problem Format on Students' Responses, *Am. J. Phys.* (to be published).

[22] D. R. Sokoloff, R. K. Thornton, P. W. Laws, *Real Time Physics Active Learning Laboratories Modules 1,3 and 4* (Wiley, New York, 2004), 2nd ed.

[23] S. Lance Cooper, B. M. Dick, and A. Weissman, *Physics 211 Laboratory Experiments, Department of Physics, College of Engineering, University of Illinois at Urbana Champaign* (Stipes Publishing, Champaign, IL, 2004).

[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevSTPER.10.020104 for a sample lab (Appendix I) and a detailed version of results and analysis (Appendix II).

[25] B. Thacker, Curriculum, http://www.phys.ttu.edu/~batcam/Curriculum.html, 05/23/14.

[26] C. J. Hieggelke, D. P. Maloney, T. L. O'Kuma, and S. Kanim, *E&M TIPERs: Electricity & Magnetism Tasks* (Addison-Wesley, Reading, MA, 2005).

[27] T. L. O'Kuma, D. P. Maloney, and C. J. Hieggelke, *Ranking Task Exercises in Physics: Student Edition* (Addison-Wesley, Reading, MA, 2003).

[28] A. B. Arons, *Teaching Introductory Physics* (Wiley, New York, 1996).

[29] A. B. Arons, *A Guide to Introductory Physics Teaching* (Wiley, New York, 1990).

[30] P. W. Laws, *Workshop Physics Activity Guide* (John Wiley & Sons, New York, 1999).

[31] We are presently studying the drop in the number of students from the first to the second semester of the algebra-based course. This drop is striking because most of the students taking the first semester need to take the second semester in order to meet the requirements of their major. We have noted an increased number of students taking the second semester course in the summer sessions and it is possible that some are taking the course at different institutions in the summer. This is still being researched.

[32] C. Lillian McDermott and the Physics Education Group, *Physics by Inquiry* (John Wiley & Sons, New York, 1996), Vols. I and II.

[33] R. R. Hake, Interactive-Engagement vs Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses, Am. J. Phys. **66**, 64 (1998).

[34] PER User's Guide, http://perusersguide.org/, 04/25/13.

[35] E. Mazur, *Peer Instruction: A User's Manual* (Addison-Wesley, Reading, MA, 1996).

[36] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

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

[38] M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. Chabay, L. Ding, M. P. Haugan, M. Jackson Marr, B. Sherwood, and M. F. Schatz, Tale of two curricula: The performance of 2000 students in introductory electromagnetism, Phys. Rev. ST Phys. Educ. Res. **5**, 020105 (2009).

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

[40] Seattle Pacific University Project Report, Physics Teacher Education Coalition, http://www.phystec.org/institutions/seattle-pacific/assessment.php, 10/5/12.

[41] S. J. Pollock, Comparing student learning with multiple research-based conceptual surveys: CSEM, and BEMA, AIP Conf. Proc. **1064**, 171 (2008).

[42] D. 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).

[43] See, for example, Ball State University Project Report, Physics Teacher Education Coalition, http://www.phystec.org/institutions/ball-state/assessment.php, 10/5/12.

[44] R. W. Chabay and B. A. Sherwood, *Matter and Interactions* (Wiley, New York, 2010 and 2011), Vols. 1 and 2.

[45] Targeted Poster Session, Moving Beyond Conceptual Inventories, PERC 2011, August, 2011, Omaha, Nebraska, http://www.compadre.org/per/perc/2011/Detail.cfm?id=4114.

[46] M. D. Caballero, E. F. Greco, E. R. Murray, K. R. Bujak, M. Jackson Marr, R. Catrambone, M. A. Kohlmyer, and M. F. Schatz, Comparing large lecture mechanics curricula using the Force Concept Inventory: A five thousand student study, Am. J. Phys. **80**, 638 (2012).