

Gender disparities in second-semester college physics: The incremental effects of a “smog of bias”

Lauren E. Kost-Smith, Steven J. Pollock, and Noah D. Finkelstein

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

(Received 28 April 2010; published 3 September 2010)

Our previous research [Kost *et al.*, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009)] examined gender differences in the first-semester, introductory physics class at the University of Colorado at Boulder. We found that: (1) there were gender differences in several aspects of the course, including conceptual survey performance, (2) these differences persisted despite the use of interactive engagement techniques, and (3) the post-test gender differences could largely be attributed to differences in males’ and females’ prior physics and math performance and their incoming attitudes and beliefs. In the current study, we continue to characterize gender differences in our physics courses by examining the second-semester, electricity and magnetism course. We analyze three factors: student retention from Physics 1 to Physics 2, student performance, and students’ attitudes and beliefs about physics, and find gender differences in all three of these areas. Specifically, females are less likely to stay in the physics major than males. Despite males and females performing about equally on the conceptual pretest, we find that females score about 6 percentage points lower than males on the conceptual post-test. In most semesters, females outperform males on homework and participation, and males outperform females on exams, resulting in course grades of males and females that are not significantly different. In terms of students’ attitudes and beliefs, we find that both males and females shift toward less expertlike beliefs over the course of Physics 2. Shifts are statistically equal for all categories except for the Personal Interest category, where females have more negative shifts than males. A large fraction of the conceptual post-test gender gap (up to 60%) can be accounted for by differences in males’ and females’ prior physics and math performance and their pre-Physics 2 attitudes and beliefs. Taken together, the results of this study suggest that it is an accumulation of small gender differences over time that may be responsible for the large differences that we observe in physics participation of males and females.

DOI: [10.1103/PhysRevSTPER.6.020112](https://doi.org/10.1103/PhysRevSTPER.6.020112)

PACS number(s): 01.40.Fk, 01.40.gb

I. INTRODUCTION AND BACKGROUND

According to a recent National Science Foundation (2009) report [1], females now earn over half (55%) of the bachelor’s degrees and just under half (47%) of the doctoral degrees awarded in the sciences. Despite reaching parity of males and females in the sciences overall (and in many individual disciplines, including biology and chemistry), females still earn only 21% of the bachelor’s degrees awarded in physics. Physics has one of the lowest representations of females, and is comparable to representation in computer science and engineering (both 19% female). These national trends are reflected at the University of Colorado at Boulder (CU), where we see a similar low fraction of females earning bachelor’s degrees in physics. This under-representation of females in physics continues to be a cause for concern. In our previous work [2,3], we began to address this disparity in participation by examining gender differences in the first-semester, calculus-based mechanics course. This course serves as an introduction to physics and is a critical first step towards pursuing a physics degree. In prior studies we found that in our first-semester, introductory mechanics course:

(1) There are gender differences in students’ performance on conceptual surveys, attitudes and beliefs about physics and about learning physics, math and physics background, and high school preparation.

(2) The gender differences in conceptual performance persist from pre- to post-test despite the use of interactive engagement techniques, including *Peer Instruction* [4] and the

Tutorials in Introductory Physics [5].

(3) The conceptual post-test gender difference we observe can largely be accounted for by differences in males’ and females’ math and physics background and their incoming attitudes and beliefs about physics.

We argued that the background and preparation of students, especially background in conceptual physics, impacted not only how individual students performed in the course, but also influenced the gender differences that persisted over the semester. In the current study, we continue our work in characterizing the gender differences we observe in participation and performance by examining the second-semester, calculus-based electricity and magnetism (E&M) course. This course is particularly interesting as fewer students (male or female) have significant exposure to the E&M content before coming to the course. The goal of this work is to further develop an understanding of gender differences in participation and performance in physics by identifying gender differences in the second-semester course and determining whether these differences can be accounted for by background differences of males and females.

The benefits of using interactive engagement (IE) techniques have been consistently demonstrated [6]. At CU, classes that use IE techniques have average normalized learning gains ($\langle g \rangle$) [6] on conceptual assessments that range from 32% to 64%. While it is clear that IE methods improve student learning gains, it is less apparent that IE techniques eliminate the gender gap. Some research has suggested that females may benefit more from an interactive pedagogy than

males [7,8], and many of the recommendations made for increasing the participation of females in physics and in the sciences [9–11] align with IE techniques. Researchers at Harvard University found that a preinstruction gender gap on the Force Concept Inventory (FCI) [12] was eliminated over the course of an interactive and engaging introductory physics course [13]. They went on to claim that because their results were consistent over five different instructors, the elimination of the gender gap was not dependent on the instructor, and was due to the pedagogical strategies that were used in the class. Despite these encouraging findings, the elimination, or even reduction, of the gender gap does not appear to be universal. As stated above, at CU, we find that the gender gap on the Force and Motion Conceptual Evaluation (FMCE) [14] persists from pre- to post-test despite the use of fully interactive engagement techniques [2,3]. At the University of Minnesota, where cooperative problem-solving and “context-rich” problems [15,16] are used in the introductory physics course for scientists and engineers, researchers found that the FCI gender gap also persisted from pre- to post-test [17].

At Harvard, Lorenzo, *et al.* [13] argued that the level of interactivity in the course affects the differential performance of males and females, that is, the more interactive a class is, the smaller the post-course gender difference will be. They also argued that the reduction of the gender gap is independent of the instructor, although we found that the instructor and the choices that the instructor makes may impact the gender gap [3]. In addition to the use of interactive techniques and the instructor, there are other contextual factors in the classroom that can contribute to the gender differences we observe in the course, such as content covered, student demographics, climate, or how different activities are framed. At this point, it is unclear which of these factors are critical in reducing or eliminating gender disparities. To further explore some of the contextual factors that may be key, in this study we examine in detail the gender differences in the second-semester, E&M course.

The second-semester introductory physics course is different in many ways from the first-semester course. Most apparent are the differences in physics content. The first-semester course (Physics 1) covers mechanics, including Newton’s laws, work, energy, momentum, and waves. The second-semester course (Physics 2) covers electricity and magnetism, including electric fields, Gauss’s law, circuits, magnetic fields, and electromagnetic waves. Not only is the content covered in the two courses different, but student familiarity with the content also varies between the two courses. As reported previously [3], 72% of students in Physics 1 have taken one year of high school physics, most likely a mechanics course. Only 14% of students have taken two years of high school physics, suggesting that only a small fraction of students may have seen much E&M content in high school. Another way in which Physics 2 differs from Physics 1 is in the student population, which will be addressed in more detail later. Notably, there are fewer non-science and undeclared majors in Physics 2 than in Physics 1. Though *Tutorials* are used in both Physics 1 and 2, there are a greater number of individual tutorials in Physics 2 that require the use of equipment. Several studies have found

differences in how male and female students engage with lab equipment [18–20], and these differences could have more of an impact on student learning in Physics 2 than in Physics 1.

While there has been a lot of interest in looking at the performance of males and females in physics, most of the research studies [13,17,21–23] have focused on the first-semester, mechanics course. One exception is Meltzer’s study of a “hidden variable” in electricity and magnetism conceptual test performance [24]. Meltzer found that while students’ conceptual pretest scores were not correlated with their normalized learning gains on the Conceptual Survey in Electricity (CSE), students’ preinstruction mathematics skill was correlated with CSE gains, suggesting that differences in learning gains between two populations may be due in part to different incoming math skill, rather than different pretest scores or abilities to learn physics concepts.

With our present study of gender differences in the E&M course, we begin to tease apart some of the similarities and differences between performance in Physics 1 and Physics 2, and to get a sense of which, if any, contextual factors differentially impact males and females. Using data gathered from ten semesters of Physics 2 and more than 3500 students, we explore three gender gaps observed in the second-semester, E&M course: retention, performance, and attitudes and beliefs. We are particularly interested in whether we observe the same trends in Physics 2 as we saw in Physics 1, even though the contexts of the two courses are quite different. This continued exploration of gender differences and their possible contextual dependencies will inform future interventions designed to address gender disparities in the classroom. In this paper, we address the following research questions:

- (1) What fraction of students, and of physics majors specifically, are retained from Physics 1 to Physics 2? Are there differences between students who continue in the introductory sequence and those who do not?
- (2) How do the performance, attitudes and beliefs, and preparation of males and females in the second-semester introductory physics course compare?
- (3) To what extent can prior factors help explain or account for the persistence of the gender gap in the second-semester class?

In summary, we find no differences in the retention rates of males and females from Physics 1 to Physics 2 for students overall, though there are small differences in retention rates of male and female physics majors, with males systematically more likely to continue and less likely to drop out than females. The trends we see in terms of male and female course grades in Physics 2 match those observed in Physics 1 except for some notable cases in which males and females have significantly different course grades. Despite males and females having similar E&M conceptual pretest scores at the beginning of Physics 2, males outperform females at the end of the semester by about 6 percentage points. This post-test gender gap can largely be attributed to differences in males’ and females’ prior physics performance (FMCE post-test, BEMA pretest, and Physics 1 exam grades), mathematics standardized test performance, and students’ attitudes and beliefs. A multiple-regression model of students’ conceptual performance suggests these prior factors can account for up

to 60% of the observed gender differences. Taken all together, our current study of the second-semester physics course indicates that there is not one single factor that can explain the under-representation of females in physics, but it is rather the building up of small differences between males and females over time that may be responsible for the large disparities in participation of males and females in physics.

II. RESEARCH METHODS AND STUDENT POPULATION

The data in the following studies were collected from ten offerings (from Fall 2004 to Spring 2009) of the second-semester calculus-based introductory electricity and magnetism (E&M) course at the University of Colorado (CU). These are large-enrollment courses that typically have 300–500 students. All ten classes used interactive engagement (IE) techniques. Each of the ten classes employed student discussions around ConcepTests (*Peer Instruction*) [4] in lecture, online homework systems [25], and voluntary help-room sessions on problem-solving homework. In addition, all ten classes used *Tutorials in Introductory Physics* [5] and Learning Assistants [26] during a 1 hr/week recitation. There is no laboratory associated with this course. A more detailed description of the course structure can be found in prior work [27]. In our previous work on the first-semester course, we categorized courses as IE 1 (partially interactive, no *Tutorials*) or IE 2 (fully interactive, use *Tutorials*) [2]. All ten classes in this study are categorized as IE 2. Though we categorize all of these classes as IE 2, we also recognize that there are a variety of faculty teaching these classes who have differing levels of experience and familiarity with the interactive engagement methods that are employed. Though the curriculum looks the same, we know that how the curriculum is enacted can be very different [28].

The ten classes included in this study were taught by seven different instructors [29]. Instructors (and semesters) will be identified by a letter. When the same instructor taught this second-semester course multiple times, they will be identified by a letter followed by a number to indicate the number of times that they have taught. For instance, the first time professor A taught this course would be labeled A1, the second time A2, and so forth. Instructors who only taught the course once during these ten semesters will be identified by a letter only. All instructors except one (Professor F) were male.

The student population in the second-semester introductory course is about one-quarter female, just as is the case in Physics 1. Over half of the students are declared engineering majors and about 20% are other science majors. Only about 8% of the students who are enrolled in the introductory E&M course are declared physics majors. This is a slightly different student population than we see in Physics 1. Not surprisingly, there are fewer nonscience and undeclared majors in Physics 2 than in Physics 1. Also, a larger fraction of the students are engineering majors in Physics 2 than in Physics 1. We also see significant differences in student major by gender ($p < 0.001$), as shown in Table I. Namely, females are less likely to be engineering majors, and about

TABLE I. Frequencies for gender, student declared major, and ethnicity for all students in the study, that is, students who enrolled and received a grade in the second-semester introductory physics course between Fall 2004 and Spring 2009.

Gender ($N=3895$)				
	%			
Male	75.3			
Female	24.7			
Major ($N=3894$)		%	% of males	% of females
Physics	8.2	9.0	6.0	
Engineering	57.6	60.3	49.1	
Other science	21.8	17.6	34.7	
Nonscience	5.3	5.0	6.4	
Undeclared or other	7.0	8.1	3.9	
Ethnicity ($N=3662$)		%	% of males	% of females
Asian	8.7	7.8	11.6	
African American	1.3	1.2	1.4	
Hispanic	6.0	6.0	6.0	
Native American	0.7	0.7	0.7	
White	81.4	82.6	77.7	
Foreign	2.0	1.8	2.5	

twice as likely to be other science majors, which is similar to trends observed in Physics 1. Unlike in Physics 1, where the percentage of male and female physics majors were not different (5.6% of males and 5.2% of females), in Physics 2 a higher fraction of the male students are physics majors as compared to female students (9% of males versus 6% of females). This difference is significant ($p < 0.01$). Looking at student ethnicity, over 80% of the students are white, about 9% are Asian, and less than 10% are African American, Hispanic, or Native American. There are only small differences in ethnicity by gender. Further, we see no differences between the ethnicity distributions of students in Physics 2 compared to Physics 1.

Of primary interest in this study is to what degree males and females differ on measures of background and preparation and to what degree these differences contribute to the observed gender gap. Conceptual performance, as measured by the Brief Electricity and Magnetism Assessment (BEMA) [30], serves as the focus of the study. The BEMA post-test score for each student is used as a measure of the student's conceptual understanding of physics at the end of the semester. Only students with matched pretest and post-test data are included. In two of the ten semesters the BEMA was not given to students as a pretest, and was only given at the end of the semester as a post-test [31]. Though we have post-test data for these semesters, we do not include them in most of the analyses in this paper, as we cannot match individual students' pre- and post-test scores. In another two of the ten semesters, the BEMA was only given to half of the students in the class. We also have excluded these semesters in some of the analyses in this paper. The number of students with

TABLE II. Average course grades for males and females who did and did not take the BEMA. Course grades are on a 0.0–4.0 scale.

	Males			Females			Differences	
	<i>N</i>	Mean	SD	<i>N</i>	Mean	SD	M–F	<i>p</i> value
Students without pre- and/or post-BEMA	500	1.93	1.1	114	1.96	1.2	–0.03	0.773
Students with pre- and post-BEMA	1257	2.85	0.8	447	2.80	0.8	0.05	0.307
Differences	(BEMA–no BEMA)	0.92			0.84			
	<i>p</i> value	<0.001			<0.001			

matched pretest and post-test data from the remaining six semesters is 1704. Additional evaluation of student performance in the course is captured by homework, exam, participation, and course grades, which were collected from the instructor in each semester.

Data have been gathered [32] on students' background knowledge and their preparation for college physics. Prior academic performance is captured by students' high school grade point average (GPA), while the BEMA pretest scores and FMCE post-test scores from the previous term (when available) are used to measure students' prior conceptual understanding of physics. The math portion of the Scholastic Aptitude Test (SAT-math) and the math portion of the American College Test (ACT-math) were used as measures of students' prior knowledge of mathematics [33]. Scores on each of the math tests were similarly correlated with the BEMA post-test ($r \approx 0.35$) and were also highly correlated ($r = 0.71$) with each other (for the almost 2000 students who took both tests). To get a measure of prior math knowledge for almost every student and to avoid having multiple variables that contained the same information, the scores on the two tests were combined. The scores for each test were first normalized (converted to z -scores [34]). For students who took only one of the two tests, the z -score on that test was used to measure mathematics knowledge. For the smaller number of students who took both tests, the combined math score is an average of the z -scores for each test. Student course preparation for college physics is measured by how many years of high school physics and calculus a student had taken. Data were not available on the grade that students received in their high school courses.

In addition to students' prior content knowledge, data were also collected on their attitudes and beliefs about physics and about learning physics. Attitudes and beliefs are measured by the Colorado Learning Attitudes about Science Survey (CLASS) [35]. The CLASS questions are classified into eight categories of student beliefs. The survey is made up of 42 statements and students respond on a Likert-like scale. Each response is coded favorable, neutral, or unfavorable based on whether the response agrees or disagrees with the expert response. Students are then given a percent favorable and a percent unfavorable score in each category. Favorable pretest scores on each category are used as measures of students' incoming beliefs. Favorable post-test scores and shifts (post–pre) are used as measures of students' attitudes and beliefs at the end of the semester and to measure change in attitudes and beliefs, respectively. The CLASS is adminis-

tered at the beginning and end of Physics 1 and Physics 2. For those students who took both the pre- and post-CLASS in Physics 1 and Physics 2, the correlation between the post-Physics 1 CLASS score and the pre-Physics 2 CLASS score is 0.73, suggesting that students' attitudes and beliefs are fairly stable over the winter and summer breaks. This analysis will primarily use results from the survey administered in Physics 2.

As discussed in our prior paper [3], we note that several assessments used throughout the study only measure student *performance* on these instruments—however, we use them as a proxy measurement of student understanding and actual attitudes and beliefs upon entry and exit. We recognize that these instruments may be measuring more, such as test-taking ability, and may differ by gender. Several studies have shown that the context of questions and the format of questions could disadvantage males or females unequally [36–38]. While these studies question the validity of these instruments, we note that (a) we are using the standard measures that have been adopted by the community, and (b) we are analyzing *shifts* on these instruments, which allows us to normalize students against themselves.

The BEMA is administered the first and last weeks of the semester during recitation, and only those students that attend both weeks take the pre- and post-BEMA. This nonrandom sampling could introduce bias into our results. To understand the bias of our sample we compare students who did and did not take both the pre- and post-BEMA. Of the 2318 students who took Physics 2 during the semesters included in this study (and in the semesters where the BEMA was offered to all students both pre- and post-instruction), 1704 students (74%) took both the pre- and post-BEMA. Comparing the populations that did and did not take both the pre- and post-BEMA, we find that females were more likely to take the BEMA than males: 80% of females took the BEMA, while only 72% of males took the BEMA. The course grades (on a scale from 0.0 to 4.0) for males and females in each group are shown in Table II. The average course grades of students who took the BEMA are higher than the course grades of students who did not take the BEMA ($p < 0.001$). While this is a source of bias, there is no significant gender gap in course grades for either of the two groups. By focusing on the BEMA as a measure of learning, we limit the sample of students included in the analysis and exclude primarily those with lower course grades. But, the similarity in gender gaps in course grades for the two groups suggests that the estimate of the gender differences in con-

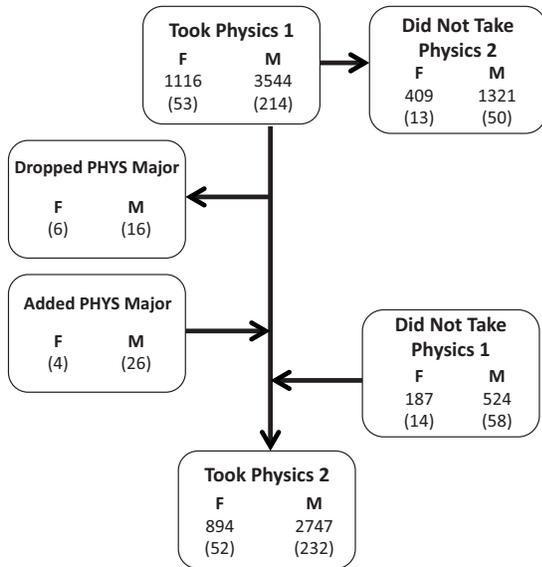


FIG. 1. Tracking students through the introductory physics sequence. The chart above shows the numbers of males and females who took Physics 1 (between spring 2004 and spring 2008) and Physics 2 (between fall 2004 and spring 2009). The numbers in parentheses are the number of male and female *physics majors* (PHYS) at each step. Males and females are about fractionally equal at every step of the chart, except in the percentage of physics majors in Physics 2.

ceptual performance provided by the BEMA may be a reasonable estimate of the gender difference for all students.

III. RESULTS: TRACKING STUDENTS FROM PHYSICS 1 TO PHYSICS 2

Before looking at differences between males and females in the second-semester course, we look at which students continue from Physics 1 to Physics 2 and whether there are differences by gender. For this analysis, we only included students who took Physics 1 between Spring 2004 and Spring 2008. We find that the majority of students who do take Physics 2, take it within a year of taking Physics 1. By only looking through Spring 2008, we ensure that the majority of students included in our analysis will have taken Physics 2 if they were likely to do so. Figure 1 shows the number of males and females (physics majors are listed in parenthe-

ses) at each step of the progression from Physics 1 to Physics 2. Of the students who took Physics 1, 37% of both males and females did *not* take Physics 2. We see no gender difference in the percentage of students who do not go on to Physics 2. Of the students who took Physics 2, about 20% of both the males and females did not take Physics 1. Looking both at the number of students who drop out of the introductory sequence and who join in the sequence after Physics 1, we find no gender differences.

Next, we concentrate on physics majors, that is, those students who were declared physics majors the semester that they took Physics 1 or 2, regardless of whether they remained physics majors until graduation. As we mentioned above, there is not a significant difference in the percentage of male and female physics majors who take Physics 1. Also, there is not a significant difference ($p > 0.8$) in the percentage of males and females who were declared physics majors in Physics 1, but who never took Physics 2, and presumably changed their major. 25% of female physics majors and 23% of male physics majors in Physics 1 never took Physics 2. On the left side of Fig. 1 is information about students who took Physics 2, but who changed their major between Physics 1 and Physics 2. Of students that were declared physics majors in Physics 1, 11% of the females and 7% of the males changed their major to something other than physics between Physics 1 and 2 (not a significant gender difference). Of students that were not declared physics majors in Physics 1, 0.4% of females and 0.8% of males switched their major to physics between Physics 1 and Physics 2 (not a significant gender difference). These small but consistent gender differences in the numbers of students who continue through the introductory sequence result in a significantly smaller percentage of females who are physics majors in Physics 2 compared to males. In Physics 2, 6% of the females are physics majors and 8% of the males are physics majors. Looking at the trajectories of males and females from Physics 1 to Physics 2, we see no significant gender differences, but the small, nonstatistically significant differences combine such that there is a smaller percentage of females than males who are physics majors in Physics 2.

We also compare the Physics 1 course grade and FMCE scores of students who did and did not go on to Physics 2. These comparisons are found in Table III. We begin by comparing students within each gender. Not surprisingly, the students who did take Physics 2 have higher course grades in Physics 1 than the students who did not take Physics 2. The

TABLE III. Gender gaps in course grades and FMCE for those students who took Physics 1 but then did and did not take Physics 2. The differences in the bottom row are (Physics 1 and 2 – Physics 1 only). The asterisk (*) indicates that the difference is significant at the $p < 0.05$ level.

	Physics 1 course grade			FMCE pretest			FMCE post-test		
	M	F	M-F	M	F	M-F	M	F	M-F
Physics 1 only	2.03	1.90	0.13*	28.1	21.0	7.1*	59.0	48.1	11.1*
Physics 1 and 2	2.83	2.74	0.09*	34.4	23.1	11.3*	69.6	57.0	12.6*
Differences	0.80*	0.84*		6.3*	2.1		10.6*	8.9*	

TABLE IV. Gender gaps in Physics 1 CLASS (% favorable) pretest, post-test, and shifts for those students who took Physics 1 but then did and did not take Physics 2. The differences in the bottom row are (Physics 1 and 2 – Physics 1 only). The asterisk (*) indicates that the difference is significant at the $p < 0.05$ level.

	Physics 1 CLASS pretest			Physics 1 CLASS post-test		
	M	F	M–F	M	F	M–F
Physics 1 only	62.7	62.0	0.7	56.0	52.3	5.7*
Physics 1 and 2	66.4	64.2	2.2*	61.3	57.5	3.8*
Differences	3.7*	2.2		5.3*	5.2*	

difference is significant ($p < 0.05$) and about the same for both males and females. The effect sizes [39] of the differences are 0.77 for males and 0.81 for females, both relatively large effect sizes. Having found sizeable differences between the course grades of students who did and did not take Physics 2, we now look at FMCE scores, beginning with the pretest. Males who did take Physics 2 had significantly higher ($p < 0.05$) FMCE pretest scores than males who did not take Physics 2. The difference was about 6% (an effect size of 0.29). However, the FMCE pretest scores of females who did and did not take Physics 2 are not significantly different ($p = 0.1$). This suggests that despite the two groups of females being equally (un)prepared for Physics 1 in terms of incoming conceptual performance, some of the females continued on in physics while other females did not. It appears that FMCE pretest score is an indicator of whether males move on to Physics 2, but it is not an indicator for females. There are also differences on the FMCE post-test score for both males and females comparing those who did and did not take Physics 2. The post-test differences between those students who only took Physics 1 and those students who went on to Physics 2 are larger for both males (11%) and females (9%) than were the pretest differences. The effect sizes of the post-test differences are 0.38 for males and 0.31 for females.

We now look at the gender differences for students who do and do not go on to Physics 2. We might expect that for those students who continued on in the introductory physics sequence, we would not observe the same gender differences that we observed for the Physics 1 class as a whole, that is, the gender differences observed in Physics 1 may be primarily due to those students who drop out of the introductory sequence by Physics 2. We find that this is not the case in general. In terms of course grade, the gender difference in course grade for students who did take Physics 2 is slightly smaller than the gender difference for the students who did not take Physics 2. But, the situation is reversed when looking at the FMCE pre- and post-test. The gender gap on the FMCE pre- and post-test for students who did take Physics 2 is larger than for students who did not take Physics 2. The gender gap (and its increase from pre- to post-test) persists even when only looking at this special subpopulation of students who continued on to Physics 2.

In addition to looking at performance measures to compare students who do and do not go on to Physics 2, we can also look at students' attitudes and beliefs, as measured by

the CLASS instrument. In Table IV we present the CLASS pre- and post-test scores for males and females who did and did not take Physics 2. These data are collected at the beginning and end of Physics 1. As with prior studies, we observe an overall negative shift in student attitudes and beliefs for all students. Again, we might expect that those students who go on to Physics 2 would have more favorable attitudes and beliefs than those students who do not go on. We do find that students who take Physics 2 have more favorable attitudes and beliefs both at the beginning and end of the semester than students who do not take Physics 2 (though the difference in pretest scores for females is not significant). Though these differences between students who do and do not take Physics 2 are significant, the effect sizes of the differences are small, between about 0.1 and 0.3.

In summary, despite the gender differences that we see at the end of Physics 1 (in terms of FMCE post-test score), we find that males and females are continuing through the introductory sequence (and not continuing) at the same rate. The same is true of the physics majors, though we do see a smaller percentage of female physics majors in Physics 2 compared to males; 6% of females versus 8% of males are declared physics majors in Physics 2. As evidenced by the Physics 1 grades and FMCE post-test scores of males and females who continue on to Physics 2, the females who are taking Physics 2 are less prepared than the males. Females who continue in the introductory sequence also have less favorable attitudes and beliefs than the males who take Physics 2. Having examined the gender differences in retention from Physics 1 to Physics 2, in the next section, we take a closer look at the performance gender differences in the second-semester physics course.

IV. RESULTS: IDENTIFYING DIFFERENCES BY GENDER

A. College course performance differences

We now focus our attention on students in the second-semester introductory course by examining conceptual mastery, course grades, DFW rates (grades of D, F, or withdrawal), and attitudes and beliefs.

1. Conceptual surveys

We first look at students' conceptual performance as measured by the BEMA. Figure 2 presents the pre- and post-test gender gaps for each semester included in the study. Recall

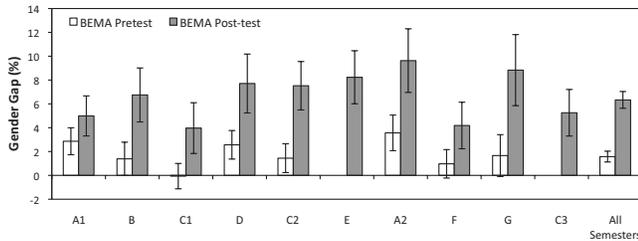


FIG. 2. Pre- and post-test gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) by semester. The data shown here includes all students who took the pre- and post-BEMA. These data represent seven different instructors and over 2500 students. The instructor is indicated along the x axis. Instructors who taught more than once are labeled with a letter and a number. The error bars represent the standard errors of the mean.

that in two of the ten semesters (Semesters E and C3) students were given the BEMA only at the end of the semester as a post-test. We include the post-test gender gaps for these semesters, even though we have no pretest data. In five of the eight semesters of pretest data there is not a significant gender difference in pretest scores ($p > 0.05$). Males and females do not score significantly differently on the BEMA pretest in the majority of the semesters that the BEMA pretest has been given. In the remaining three semesters, where there is a statistically significant pretest difference, the gender gap is only between 2.6% and 3.6%. This is much smaller than the gender gaps that we observe on the FMCE pretest at the beginning of Physics 1, which are between 6% and 14% (about 10% on average) [3]. Taking the BEMA pretest score as a measure of preparation, it seems that males and females are equally prepared for Physics 2 in terms of exposure to E&M content.

Despite equal preparation of males and females, the BEMA post-test gender gap is statistically significant ($p < 0.05$) in all ten semesters. Males scored significantly higher on the BEMA post-test than females in all semesters. Even in those semesters where there was no significant difference on the pretest, males and females performed differently on the post-test. Recall that all ten semesters used fully interactive engagement methods, including *Peer Instruction* [4] and *Tutorials* [5]. Despite the use of these IE methods, the gender gap increases from pre- to post-test in all semesters of Physics 2. The post-test gender gap ranges from 4.0% to 9.6%. On average, the effect size [40] of the pretest gender gap is 0.17, and the effect size of the post-test gender gap is 0.39.

The normalized gain [6],

$$\langle g \rangle = \frac{\langle \text{post} \rangle - \langle \text{pre} \rangle}{100 - \langle \text{pre} \rangle},$$

for these classes is between 0.33 and 0.48. Over the eight semesters in which both the pre- and post-test were given, the average normalized gain is 0.40. These gains match the range of normalized learning gains reported for classes at other institutions that use the Matter and Interactions [41] curriculum [42]. While the normalized gains for the course are in line with gains of other reformed courses, we do see differences by gender. The individual normalized gains [43]

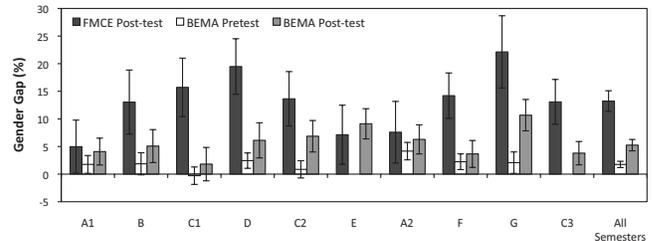


FIG. 3. FMCE and BEMA gender gaps ($\langle S \rangle_M - \langle S \rangle_F$) by semester. The gender gaps for each semester are shown above ($N = 1487$) for only those students who took both the FMCE (in the first-semester course) and the BEMA (in the second-semester course). The error bars represent the standard errors of the mean.

of males and females are statistically different ($p < 0.01$). Females have an average normalized gain of 0.35 (over all semesters) while males have an average normalized gain of 0.42. It appears that females learn a smaller percentage of what they did not already know coming into Physics 2 than males.

From Fig. 2, it appears that a gender gap is created over the course of Physics 2. Males and females come into the course with the same level of E&M conceptual understanding, and at the end of the course, the males are performing better on the BEMA than the females. However, if we look at the FMCE post-test gender gaps for those students who took both the BEMA and the FMCE, shown in Fig. 3 [44], we find that in most semesters the post-FMCE gender gaps are larger than the post-BEMA gender gaps. Further, in our classes the average pretest scores on the BEMA are 25.3% for females and 26.8% for males. These scores are close to “informed guessing.” Most students only take one year of high school physics, which is most likely a mechanics course, so most students, male or female, probably have not been exposed to much E&M when they come into Physics 2. One interpretation is that the pretest gender gap is masked by lack of conceptual exposure to the subject. The BEMA does not measure what students know on the pretest, and in fact, if the FMCE is taken as the measure of conceptual performance upon entering Physics 2, it appears that we may be reducing the gender gap from Physics 1 to Physics 2.

2. Course grades

In addition to looking at student performance on conceptual surveys, we can also look at how males and females performed in the course overall and on each of the components of the course. For each of the ten semesters of the E&M course males’ and females’ scores are averaged on participation, homework, exams, and total course grade. In all of these courses, exams make up 60–70% of the course grade, homework counts for 25% to 35%, and participation makes up the remainder, between 0% and 10%. The difference between the average scores of males and females in each component ($\langle S \rangle_M - \langle S \rangle_F$) is calculated for each semester. These differences for each semester, along with the average differences across all semesters, are shown in Table V.

On average in Physics 2, females outscore males by about 6% on participation and by about 5% on homework, but

TABLE V. Analysis of students' course grades. Each column contains the difference between the average scores for males and females ($\langle S \rangle_M - \langle S \rangle_F$). Error (shown in parentheses) is computed from the standard errors of the mean for males and females added in quadrature. The asterisk (*) indicates that the difference is statistically significant at the $p < 0.05$ level. In Semester D no participation credit was given.

	Participation (%)	Homework (%)	Exams (%)	Course GPA (4 pt. scale)
Semester A1	-6.4(1.0)*	-7.0(1.5)*	3.6(1.6)*	-0.04(0.12)
Semester B	-6(3)*	-4(2)	6.1(1.9)*	0.08 (0.13)
Semester C1	-7.4(1.6)*	-3.3(1.8)	4.2(1.5)*	-0.03(0.12)
Semester D		-3(3)	2.5 (1.7)	0.06 (0.13)
Semester C2	-7.2(1.6)*	-5.5(1.9)*	4.8(1.7)*	0.07 (0.12)
Semester E	-5(2)*	-4.0(1.7)*	2.4 (1.6)	-0.01(0.13)
Semester A2	-4.9(1.7)*	-2.6(1.9)	3.1(1.4)*	0.02 (0.11)
Semester F	-11.5(1.8)*	-8.2(1.9)*	-1.5(1.6)	-0.32(0.11)*
Semester G	-2.7(1.9)	-3.7(1.7)*	7.7(1.4)*	0.20(0.09)*
Semester C3	-6.1(1.7)*	-5.3(1.8)*	4.2(1.4)*	0.02 (0.11)
Average	-6.1(0.6)*	-4.7(0.6)*	3.8(0.5)*	-0.002(0.04)

males outscore females by about 4% on exams. This is very similar to what we reported in studies of Physics 1 [3]. These differences on participation, homework, and exams offset one another resulting in course grades of males and females that are not statistically different. This happens on average and in most individual semesters. There are two notable differences from this trend. In Semester G, the differences in males' and females' homework and participation scores are on the smaller side, and the difference in exam scores is the largest that we report. This results in course grades of males and females that are significantly different. Males have course grades about 0.2 grade points higher than females (on 0 to 4.0 scale).

Another semester that stands out is Semester F, the only semester in which there was a female professor. In this semester females have much higher participation and homework scores than males. The differences of 12% on participation and 8% on homework are the largest differences in this data set. In addition, the exam scores of males and females are not significantly different. This leads to females having higher course grades than males by about a third of a letter grade. This is the largest gender difference that we have seen in course grades (including Physics 1 and Physics 2), and the only time we have seen females with statistically significantly higher course grades than males. Though females outperformed males in the course overall in this semester, there was still a small (4%), but statistically significant, BEMA post-test gender gap.

3. DFW rates

Another way to compare course grades is to look at DFW rates of males and females. The DFW rate for each semester is the percentage of students that received a grade of D, F, or W (withdrew from the course [45]). Table VI lists the DFW rates for males and females in each semester and an average DFW rate over all semesters. We first look to see if there are differences in the DFW rate from semester to semester for

males and for females. There are not significant differences in the female DFW rate by semester ($p > 0.3$, via χ^2 test), but there are significant differences in the male DFW rate by semester ($p < 0.01$). We use pairwise χ^2 tests to determine in which semesters males have significantly different DFW rates. The only significant differences involve semester G being different from semesters A1 and F. Overall, the DFW rates of males and females do not change much from semester to semester, and less than 15% of both males and females receive grades of D, F, or W.

We next compare the DFW rates *between* males and females to determine if they are significantly different in any

TABLE VI. DFW and W rates for Males and Females in Each Semester. The DFW rate is the percentage of students who receive a grade of D, F, or W (withdrew from the course). The W rate is the percentage of students who withdrew from the course. On average, the DFW and W rates of males and females are not significantly different.

	DFW Rates (%)		W Rates (%)	
	Males	Females	Males	Females
Semester A1	17.5	16.7	2.7	4.6
Semester B	16.9	16.7	2.4	5.1
Semester C1	15.5	17.6	1.5	4.8
Semester D	12.6	14.7	3.0	2.1
Semester C2	12.5	9.3	2.1	4.1
Semester E	15.7	12.5	3.8	2.5
Semester A2	16.3	15.2	4.1	2.7
Semester F	20.4	8.4	3.7	0
Semester G	8.5	9.5	3.4	4.3
Semester C3	12.3	10.6	0.3	0
Average	14.8 (1.1)	13.1 (1.1)	2.7 (0.4)	3.0 (0.6)

semester. In most semesters, the DFW rates of males and females are not significantly different ($p > 0.4$). The one exception is Semester F, where the DFW rate for males was 20% and the DFW rate for females was 8%. This is the only semester where the DFW rates of males and females were significantly different ($p < 0.01$). We saw above that females had higher course grades in this semester than males. But if we look at just the percentage of males and females who withdrew from the course in Semester F, we find that none of the females withdrew, but 4% of the males withdrew that semester (a significant difference, $p = 0.04$). The only other semester where there was a significant difference in the numbers of males and females that withdrew from the course was semester C1, where a higher percentage of females withdrew compared to males ($p = 0.04$).

4. Attitudes and beliefs

In addition to looking at performance in the second-semester introductory course, we can also explore how the attitudes and beliefs of males and females change over the course of the semester and whether there are any gender differences. In our previous work looking at the CLASS [3], we found that both males and females shifted towards less expert-like attitudes and beliefs over the course of the *first-semester* introductory physics course, and females had more negative shifts in all categories than males. This is reflected in the pretest scores that are collected at the beginning of the second-semester course. In all categories except Sense-Making females have significantly lower average pretest scores than males ($p < 0.05$). This means that females come in to the second-semester course reporting less expertlike attitudes and beliefs about physics and learning physics than males. If we look at the shifts in students' attitudes and beliefs over the course of Physics 2, we see different results than we saw in Physics 1. In our previous work [3], we found that females had more negative shifts than males in all categories and overall. The shifts that we saw in Physics 1 were between about -5% and -15% . Figure 4 shows the shifts of males and females in Physics 2. We see in Fig. 4 that the shifts over the course of Physics 2 are considerably smaller than those from Physics 1, but still zero or negative. The shifts over the course of Physics 2 are between 0% and -6% . We do not find any significant gender differences in the shifts of males and females except in the Personal Interest category, where females have more negative shifts than males. The Personal Interest category has one of the largest pretest gender differences, which, in combination with the gender differences in shifts over the semester, results in an 11% post-test gender difference in the Personal Interest category, larger than any other category. In summary, the attitudes and beliefs of students do shift towards less expertlike beliefs over the course of Physics 2 [46], but the shifts are much smaller than in Physics 1. Also, there are fewer differences in shifts between males and females in Physics 2 than in Physics 1.

B. Background differences

In the previous section we reported the observed differences in males' and females' performance and attitudes in the

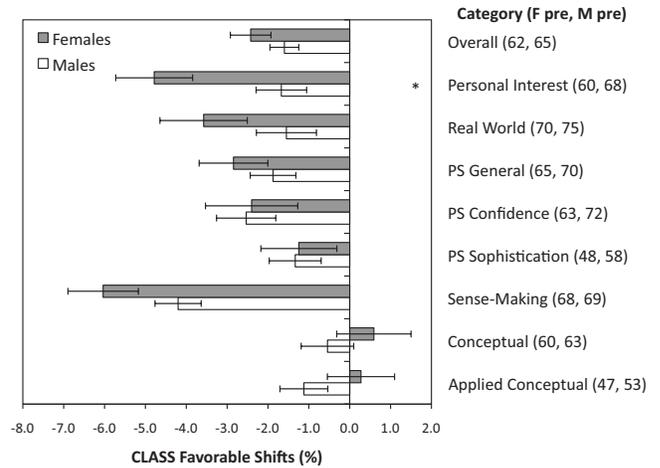


FIG. 4. Average shifts (post-pre) for males and females in Physics 2 on each of the CLASS categories. Note that all shifts are negative or zero, meaning both male and female students shift toward less expertlike attitudes and beliefs about physics or remain the same. The asterisk (*) indicates that the difference in shifts for males and females is significant ($p < 0.05$). Values in parentheses (on the right hand side) are female and male average pretest scores. The *pretest* scores of males and females are significantly different in all categories except sense-making.

second-semester course. Here, we examine the background and preparation of males and females in Physics 2. As part of students' background and preparation we look at both high school factors and data from Physics 1, for those students who took Physics 1. Male and female averages for each of the background variables and the gender differences for each are presented in Table VII for the population of Physics 2 students. Note that not all data are available for all students, as is the case in any course. As a consequence of missing data the reported averages may be biased due to sampling error. We present them regardless as they are the best estimates we have of the values for all students who enroll in Physics 2.

Just as we saw when comparing the FMCE scores of students who went on to Physics 2 to students who did not in Sec. III, we find that the measures of students' physics and math background found in Table VII are higher for this population of Physics 2 students than they were for the population of Physics 1 students [3]. Despite the higher level of preparation of these Physics 2 students, Table VII shows that males have significantly higher ($p < 0.05$) values than females on almost all variables. Females take less high school physics than males and score lower on the SAT- and ACT-Math tests. We also see that females perform worse in Physics 1 (as discussed above) than males. Females in Physics 2 had lower FMCE pre- and post-test scores and had lower grades in Physics 1 than males. The only background variables in which males do not outperform females are high school GPA, where females outscore males, and years of high school calculus, where males and females are not significantly different. Similar results have been found by other researchers [47,48], who concluded that males and females were equally prepared for the introductory physics courses. We suspect that overall measures of high school grades and

TABLE VII. Male and female average values for variables that were collected. The range of possible scores for each variable is shown in parentheses. The effect size is calculated as $ES = (\langle S \rangle_M - \langle S \rangle_F) / SD$, where the SD for all students is used. Significant differences exist between males and females on almost all of the variables.

	Males			Females			Differences		
	N	Mean	SD	N	Mean	SD	M-F	ES	p value
High school GPA (0.0–4.0)	2816	3.604	0.4	918	3.796	0.3	-0.192	-0.50	<0.001
Yrs. high school physics	2564	1.105	0.6	824	1.021	0.6	0.084	0.15	<0.001
Yrs. high school calculus	2538	0.876	0.6	818	0.899	0.6	-0.023	-0.04	0.3
SAT–Math (200–800)	1964	655.2	66	680	640.1	73	15.1	0.22	<0.001
ACT–Math (1–36)	2259	28.63	4	791	28.26	4	0.37	0.10	0.012
Math combined (z -score)	2817	0.116	0.9	923	-0.04	1	0.156	0.17	<0.001
FMCE Pretest (0–100)	1575	34.5	22	559	23.2	15	11.3	0.54	<0.001
FMCE Post-test (0–100)	1575	70.3	27	559	57.3	28	13.0	0.47	<0.001
Physics 1 Grade (0.0–4.0)	2624	2.81	0.8	847	2.71	0.9	0.10	0.12	0.003
Physics 2 CLASS Pretest (0–100)	1480	65.0	18	625	61.9	18	3.1	0.17	<0.001
Physics 2 CLASS Post-test (0–100)	1480	63.4	19	625	59.5	19	3.9	0.20	<0.001
BEMA Pretest (0–100)	1485	26.8	10	515	25.3	9	1.5	0.17	0.001
BEMA Post-test(0–100)	1485	57.7	16	515	51.4	15	6.3	0.39	<0.001

enrollment in a calculus course are not as important to performance in an introductory physics course as are enrollment in high school physics (exposure to relevant content) and performance on standardized math tests (measures of mathematics performance). By such metrics, females are less prepared for Physics 1 and Physics 2 than males.

V. RESULTS: CORRELATION OF STUDENT BACKGROUND WITH STUDENT CONCEPTUAL PERFORMANCE

Having identified several background variables that vary by gender (high school classes taken, standardized test scores, Physics 1 performance, and BEMA pretest), we next want to know which of these variables is associated with performance on the BEMA post-test and could potentially account for some of the post-BEMA gender difference that we observe. One way to determine whether a background variable can help account for the BEMA post-test gender gap is to group students according to the background variable and then compare the average BEMA post-test scores of students in each group. In this way, we can control for students background score, only comparing students that are similar on that measure. We would normally begin by looking at the BEMA pretest, but the lack of spread in BEMA pretest scores precludes an analysis of this sort. Any reasonable sort yields the same 6% spread that exists in the overall score.

Rather than looking at the BEMA pretest, we can use the FMCE post-test as a measure of prior conceptual understanding. In Fig. 5, we have divided students into five groups by FMCE post-test score. The groups are divided such that an equal number of students is in each bin. We then calculate the average BEMA post-test score for the males and females in each bin. As is seen in Fig. 5, males and females with the same FMCE post-test score have BEMA post-test scores that

are not significantly different in all five of the bins. Males and females who score similarly on the FMCE post-test in Physics 1, score similarly on the BEMA post-test in Physics 2. Though the difference is not significant in any individual bin, males outperform females in four of the five quintiles. The percentages located above each bar in Fig. 5 indicate the percentage of females (or males) that fall into that bin. The distributions of males’ and females’ among the five bins are not equal. More than half of the females are in the lowest two bins, while just about half of the males are in the highest two bins. Figure 5 suggests that by taking into account the FMCE post-test scores of males and females, we can account for a large part of the gender gap in BEMA post-test scores.

A similar analysis could be repeated for each of the background variables in Table VII separately, but ultimately, we want to know how much of the BEMA post-test gender gap can be accounted for by *all* of the background variables together. We explore this question in the following section.

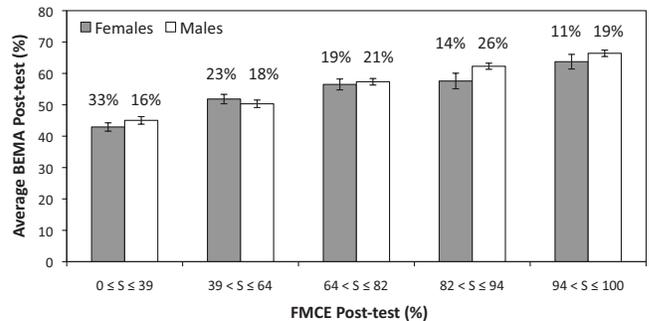


FIG. 5. Average BEMA post-test scores for females and males with matched FMCE post-test scores ($N=1117$). The percentages above each bar represent the percentage of the females (or males) from the total in each bin. The error bars represent the standard error on the mean. The differences between males and females are not significant ($p > 0.05$) in all five of the bins.

VI. RESULTS: ESTIMATION OF THE IMPACT OF STUDENT BACKGROUND ON THE GENDER GAP

We investigate whether the background differences between males and females (discussed in the Sec. IV) can account for the gender difference that we observe in BEMA post-test scores. We model students' BEMA post-test scores using a multiple-regression analysis, which describes the relationship between a student's post-test score and the values of several background variables for that student. Using this relationship, we estimate the difference in post-test scores for males and females with all background variables being held equal. In this way, we will determine how much of the gender gap can be accounted for by factors other than gender.

The post-test scores are modeled according to the equation,

$$BEMAPOST = b_0 + b_1 \times FEMALE + \sum_{k=2}^N b_k \times VAR_k,$$

where *BEMAPOST* is the post-test score on the BEMA, *FEMALE* is a dummy variable that is 1 for females and 0 for males, and *VAR_k* are the other background variables that are included in the model and any cross terms between *FEMALE* and other background variables. *b_k* are the coefficients for each term, and the multiple-regression analysis gives estimates for these coefficients. The coefficient of the *FEMALE* variable (*b₁*) gives the difference between a male's and a female's scores, with all other factors being equal. It is this coefficient that we are ultimately interested in.

As in our previous work, we are modeling students' BEMA post-test scores rather than their absolute or normalized gain because we are primarily interested in reducing the gender gap in post-test scores. By modeling the post-test, we can determine what factors influence the post-test score and could therefore contribute to the gender gap. Each of the possible confounding variables is included in the regression analysis. Variables are entered sequentially in order to find the parsimonious combination of factors that best predicts the post-test score for each student. The best model will be judged based on the size of the coefficients, the increase in multiple *R*² (the fraction of variation in post-test scores that is accounted for by the variables in the model), and the significance of variable coefficients.

As stated above, not all data were available for all students. With this being the case, only a subsample of students who took the second-semester introductory course was used in the multiple-regression analysis. Recall that only 1704 of the 3895 students who enrolled in Physics 2 between Fall 2004 and Spring 2009 took the BEMA pretest and post-test. Of these 1704 students, complete data (meaning all background variables presented in Table VII) were available for only 637 students. These 637 students make up the first sample used for the analysis. This sample of students is labeled *S*₁. All of the students in *S*₁ took the FMCE in Physics 1, so we can use their FMCE post-test score as a measure of their prior understanding of mechanics content. Students' grade in Physics 1 could also be used as a measure of mechanics understanding. If we use Physics 1 grade rather than FMCE post-test score, then we have a second sample of 907

TABLE VIII. Average course grades for students included in the first regression sample (*S*₁) and those who are not in the *S*₁ sample. Course grades are on a 0.0 to 4.0 scale. The asterisk (*) indicates that the differences are significant (*p* < 0.05).

	Males	Females	M-F
In <i>S</i> ₁ sample	2.96	2.92	0.04
Not in <i>S</i> ₁ sample	2.46	2.47	0.01
In <i>S</i> ₁ – Not in <i>S</i> ₁	0.50*	0.45*	

students. This second sample, *S*₂, has more students since not everyone takes the FMCE in Physics 1, but everyone receives a grade in Physics 1. We run the regression analysis using both of these samples.

For both samples, it is important to keep in mind that the samples used are not representative of all students who enroll in Physics 2. We can see from Table VIII and IX that the students included in each of the samples have higher course grades than students not in the samples. In all cases, the differences are about half of a letter grade. Though we are sampling students with higher course grades, the gender difference in course grades for both samples is not significantly different from zero, as was the case when looking at the class overall. It appears that the samples used in the regression analyses may be good estimates of the gender *differences* for all students.

If we look further at the BEMA pre- and post-test gender gaps for all students who took the BEMA, students in the *S*₁ sample and students in the *S*₂ sample, we see that the gender differences across all three samples of students are very similar. These data are presented in Table X. This suggests again that the samples used in the regression analyses are reasonable representative of the gender *differences* in the entire population of students.

The results of the regression analysis for sample *S*₁ are shown in Table XI. Three models are reported, starting with a bivariate model that includes only gender and then additional variables are added in each successive model. The table contains the coefficient estimates (*b_k*) and *p* values for the coefficients in each model as well as the model-level statistics. The variables that are entered in each successive model are not only significant, but they also increase *R*² substantially (the additional variance explained by each model is significant via *F* test at the *p* < 0.01 level). The *R*² for the

TABLE IX. Average course grades for students included in the second regression sample (*S*₂) and those who are not in the *S*₂ sample. Course grades are on a 0.0 to 4.0 scale. The asterisk (*) indicates that the differences are significant (*p* < 0.05).

	Males	Females	M-F
In <i>S</i> ₂ sample	2.93	2.86	0.07
Not in <i>S</i> ₂ sample	2.39	2.43	-0.04
In <i>S</i> ₂ – Not in <i>S</i> ₂	0.54*	0.43*	

TABLE X. BEMA pre- and post-test gender gaps for all students, students in the first regression sample (S_1), and students in the second regression sample (S_2). The asterisk (*) indicates that the differences are significant ($p < 0.05$).

	All students			S_1 sample			S_2 sample		
	M	F	M-F	M	F	M-F	M	F	M-F
BEMA pretest	26.8	25.3	1.5*	26.8	25.0	1.8*	26.7	25.0	1.7*
BEMA post-test	57.7	51.4	6.3*	58.4	51.7	6.7*	57.5	50.8	6.7*

final model is 0.34, such that the variation in the independent variables explains 34% of the variation in post-test scores [49].

We are interested in the difference between males' and females' post-test scores after controlling for several prior factors. In Model 1, where only *FEMALE* is included as an independent variable, the gender difference is 6.8 points. This is just the average difference in post-test scores between males and females in this sample. In Model 2, several covariates that are correlated with the post-test are added. When previous physics performance (BEMA pretest and FMCE post-test), previous math performance (combined math score), and previous attitudes and beliefs (CLASS pretest) are controlled for, the gender difference drops to 2.6 points. Already, there is a substantial reduction in the gender difference once previous physics and math performance and attitudes and beliefs are accounted for.

To get a final estimate of the gender difference, we turn to Model 3. In this model, variables are added to take into account the semester that students took Physics 2. Controlling for semester is important for two reasons. First, by including a variable that controls for the semester that students took physics, some dependence among students due to taking

physics at the same time is eliminated. Second, the average post-test scores are different in each semester. Including a semester variable will account for any differences that happen by semester which contribute to the post-test scores. Although we have no further information about specific aspects of each semester that could contribute to the differences, by including the semester variables we can see if there are differences once other prior factors are accounted for. The base case in Model 3 is semester A1 (meaning there is no variable included for this semester). This means that the coefficients of each semester variable give the average difference between semester A1 and that semester after all other variables have been accounted for. For example, controlling for prior physics performance, math performance, and attitudes and beliefs, the average difference between semester A1 and semester C1 is -4.7 points. This is the only difference that is significant, but this analysis only allows a statistical comparison between semester A1 and all other semesters. It does not allow us to compare semesters B and C1, for instance. There could be other significant differences between post-test scores by semester.

With Model 3, a final estimate of the difference between a male's and a female's post-test scores, controlling for several

TABLE XI. Coefficient estimates and multiple-regression model statistics for each multiple-regression model. The S_1 sample was used for this regression analysis.

	Model 1		Model 2		Model 3	
Model-level statistics						
Multiple R^2	0.04		0.32		0.34	
F statistic p value	<0.001		<0.001		<0.001	
Residual standard error	15.5		13.1		13.0	
Predictors	Coeff.	p value	Coeff.	p value	Coeff.	p value
Intercept	58.4	<0.001	29.1	<0.01	30.5	<0.01
FEMALE	-6.8	<0.001	-2.6	0.03	-2.6	0.02
BEMA Pretest			0.25	<0.01	0.27	<0.01
FMCE Post-test			0.24	<0.01	0.23	<0.01
Combined math score			2.1	<0.01	2.0	<0.01
CLASS pretest			0.07	0.03	0.07	0.04
Semester B					0.78	0.66
Semester C1					-4.7	<0.01
Semester D					-2.1	0.27
Semester C2					-0.52	0.76
Semester F					-0.08	0.97

TABLE XII. Coefficient estimates and multiple-regression model statistics for each multiple-regression model. The S_2 sample was used for this regression analysis.

	Model 1		Model 2		Model 3		Model 4	
Model-level statistics								
Multiple R^2	0.04		0.34		0.37		0.40	
F statistic p value	<0.001		<0.001		<0.001		<0.001	
Residual standard error	15.9		13.2		12.9		12.6	
Predictors	Coeff.	p value						
Intercept	57.5	<0.01	18.0	<0.01	20.7	<0.01	40.6	<0.01
FEMALE	-6.8	<0.01	-4.6	<0.01	-4.7	<0.01	-3.2	<0.01
BEMA Pretest			0.32	<0.01	0.33	<0.01	0.30	<0.01
Physics 1 GPA			6.7	<0.01	6.9	<0.01		
Combined math score			2.1	<0.01	1.8	<0.01	0.92	0.10
CLASS pretest			0.16	<0.01	0.14	<0.01	0.12	<0.01
Semester B					1.7	0.28	1.5	0.33
Semester C1					-7.1	<0.01	-6.8	<0.01
Semester D					-4	0.02	-3.8	0.02
Semester C2					-1.4	0.34	-1.2	0.39
Semester F					-0.48	0.76	-0.67	0.66
Physics 1 Avg. Exam Score							8.0	<0.01

other factors, can be estimated. This difference is 2.6 points. This is a substantial reduction from the 6.8 point difference that is observed just by subtracting the average male and female post-test scores. Controlling for student background in this way, we can account for 62% of the observed gender gap using this final model.

We can also include Physics 1 course grade in the final model, in addition to the FMCE post-test. Though there is an increase in R^2 when Physics 1 course grade is added, because there is not a large gender difference in Physics 1 grade, including it in the model does not lower the coefficient of *FEMALE*, but rather increases it slightly to 3.1 points. We do not include Physics 1 grade in the final model because when it is included, math score and CLASS pretest are no longer significant predictors of BEMA post-test. Because each of these variables, math performance and prior attitudes and beliefs, are somewhat more explanatory and straightforward than Physics 1 grade (which is a combination of exams, homework, and participation), we chose to keep them in the final model.

We also attempted to include years of high school physics, students' declared major in Physics 2, ethnicity, and interaction variables between *FEMALE* and all other variables. None of these variables significantly contributed to the model beyond those variables already included in the final model. We suspect that this is primarily due to correlations between these variables and variables already included in the final model.

Table XII presents the results of the regression analysis using the S_2 sample. Recall that for this sample of students we used Physics 1 grade, rather than FMCE post-test, as a measure of prior mechanics conceptual understanding. For

this sample, we report four models, starting with a bivariate model and then adding variables in each successive model. The R^2 for the final model is 0.40, such that the variation in the independent variables accounts for 40% of the variation in post-test scores.

Again, we are interested in the coefficient of the *FEMALE* variable. In Model 1, where only *FEMALE* is included, the gender difference is 6.8 points, as we saw above. In Model 2, when covariates are included in the analysis, the gender difference drops to 4.6 points. We note here, that when Physics 1 grade is included in the model, rather than FMCE post-test, less of the gender gap can be accounted for. This is not surprising, since there is not as large of a gender difference on Physics 1 grade as there is on FMCE post-test. We also include variables controlling for the semester that each student took Physics 2 in Model 3. Using Model 3, we can estimate the difference between a male's and female's scores when controlling for prior physics course performance, prior math performance, and prior attitudes and beliefs to be 4.7 points. This is a smaller reduction in the gender gap than we saw using sample S_1 . Controlling for these background factors, we account for about 30% of the observed gender gap in BEMA post-test scores.

Because the gender difference in Physics 1 course grade is small, we include an average Physics 1 exam score variable in lieu of Physics 1 grade in Model 4. We want to see if more of the gender gap can be accounted for by exam score, which has a larger gender gap than course grade. The average exam score is calculated by first converting each of the Physics 1 exam scores (three midterm exams and the final exam) to z -scores, and then computing the average exam z -score for each student in the sample. Converting to z -scores is a way

to normalize the exam scores since each exam has a different average score. From Table XII, the average exam score is a significant predictor of BEMA post-test, and it reduces the FEMALE coefficient from -4.6 points to -3.2 points. When we use only the exam component of Physics 1 grade, we find that the resulting gender gap is only 3.2 points, approaching what we found with the S_1 sample.

Again, in this sample, we included years of high school physics, students' declared major in Physics 2, ethnicity, and interaction variables between FEMALE and all other variables in the regression model. None of these variables significantly contributed to the model beyond those already included in the final model. This result is likely due to correlations between these variables and variables already included in the final model.

VII. DISCUSSION AND CONCLUSIONS

In this study, we have examined in detail three gender differences in the second-semester introductory physics course: retention, performance, and attitudes and beliefs. This has allowed us to expand our understanding of gender differences at our institution. We began by tracing the trajectories of students from Physics 1 to Physics 2. We found that, overall, males and females continued and did not continue from Physics 1 to Physics 2 at the same rate. This may be largely due to course requirements of engineering and science majors, most of whom are required to take both Physics 1 and Physics 2. However, we find differences when we focus on physics majors. While the gender differences in how many students did not take Physics 2 and how many students added and dropped the physics major are not significant, the differences are in a consistent direction such that the percentage of female physics majors in Physics 2 is significantly less than the percentage of male physics majors. We are disproportionately losing female physics majors as compared to male physics majors, an issue that needs to be further investigated.

Looking at performance in the second-semester course, we find that despite apparently equal precourse E&M content exposure, males outperform females on the BEMA at the end of the semester. Though this may demonstrate bias in our courses, we argue that the BEMA pretest does not accurately measure precourse differences between males and females. And in fact, when we use the FMCE post-test as a pre-Physics 2 measure, we find that the gender gap may be reduced over the Physics 2 semester.

We also examine the course grades of males and females, as another measure of performance in the course. As we found in Physics 1, the total course grades of males and females are generally not different, as females outperform males on homework and participation, but males outperform females on exams. This trend holds true for all semesters examined except two. In one semester, the gender differences on homework and participation were small, and males considerably outperformed females on the exams, resulting in significantly higher course grades for males. The other inconsistent semester was Semester F, when females had significantly higher course grades than males. In this semester,

there was no significant difference in the exam scores of males and females, and females considerably outperformed males on homework and participation. Semester F was also the only semester (in the past 25 semesters of Physics 1 and Physics 2 in which we have been collecting data) in which a female faculty member was the lecture instructor [50]. While the impact of a female faculty member on gender differences in the introductory physics courses needs to be further investigated, there is evidence that a female role model can influence the performance of females in science and mathematics [48,51].

In addition to analyzing retention and performance, the third gender gap that we examined is in students' attitudes and beliefs. Just as in Physics 1, we find that both males and females shift toward less expertlike attitudes and beliefs over the course of Physics 2. However, the negative shifts that we observe in Physics 2 are between 0% and -6% . This is smaller than the shifts in Physics 1, which are typically between -5% and -15% . In all categories except one, Personal Interest, males and females do not have significantly different shifts. In the Personal Interest category, males have about a -2% shift while females have about a -5% shift. Because of the large pretest gender difference in the Personal Interest category and the significant gender difference in shifts, the Personal Interest category has the largest gender difference at the end of Physics 2, a difference of 11%. What was, at the beginning of Physics 1, an 8% gender difference in the Personal Interest category has increased to an 11% gender difference after just two semesters of introductory physics. It seems that we are differentially negatively impacting females' interest in physics.

In trying to understand the possible sources of the gender disparities that we observe in E&M course performance, we used a multiple-regression analysis to determine which factors contribute to students' post-test scores and could account for portions of the gender gap. We find that about 60% of the gender difference can be accounted for by differences in males' and females' prior conceptual performance on both the FMCE and the BEMA, prior math performance, and precourse attitudes and beliefs about physics. That is, the gender gap in BEMA post-test scores is reduced from about 7% to about 3% when these measures of student background are controlled for. This result is the case when we use the FMCE post-test score as a measure of Physics 1 performance. We can instead use students' Physics 1 course grade, and when we do that, we find that less of the gender gap can be accounted for, only about 30%. If we use students' Physics 1 exam average in place of the FMCE post-test, we find that about 53% of the BEMA post-test gender gap can be accounted for. These differences in how much of the gender gap can be accounted for by different variables may suggest that the gender gap we observe is in part an issue of testing. The FMCE post-test and average exam grades (both tests) can account for a higher fraction of the BEMA post-test gender gap than can total course grade (made up of tests, homework, and participation). We observe repeated gender differences in performance on tests, which are high stakes, sequestered, time-sensitive tasks. These trends, along with survey data that we have collected showing differences in males' and females' physics self-efficacy [52], suggest that

stereotype threat [53,54] may be playing a role in our courses, and affecting females' performance on tests, even tests that are explicitly used only for diagnostic purposes. The impact of stereotype threat and the alleviation of the threat through self-affirmation [55–57] are the focus of current research studies. Preliminary results suggest that self-affirmation can reduce, or in some cases eliminate, the gender gap [58]. This supports our hypothesis that stereotype or identity threat is impacting females' performance in our courses.

From this work we can draw several conclusions. First, interactive engagement is not sufficient for eliminating, or even reducing, the gender gap. As suggested by our prior work, and further emphasized by this work, we need to explore the contextual factors in our classrooms that can impact the gender gap. By examining gender differences in Physics 2, we begin to investigate the impact of different contextual factors on the gender gap. Student familiarity with the course content may be an important factor in the gender gap, as is suggested by the smaller postcourse gender differences in Physics 2, compared to Physics 1. However, students' familiarity with the current course content is not the only factor that contributes to course performance. Physics 1 performance is also a significant predictor of BEMA post-test score. Our prior work also suggested that the instructor may be a factor in the gender gap, as the post-test gender gap varied semester to semester. This is further supported by the current results, which also hint that the gender of the professor may play a role in the gender gap. These, and other, contextual factors need to be further investigated to determine if and how they influence gender disparities in the classroom.

We have seen from this work that differences in males' and females' backgrounds can account for much of the difference we observe at the end of the Physics 2 semester. This finding suggests that females are coming into our courses underprepared, and leaving our courses underprepared for future courses, as compared to males. In some sense, because the post-BEMA gender gap is smaller than the post-FMCE gender gap, we may say that females are catching up to the males. On the other hand, we may conclude that females are getting more and more behind males as they move through the introductory sequence, since they perform worse on tests of mechanics conceptual understanding and subsequently perform worse on tests of E&M conceptual understanding.

Further, gender differences in students' personal interest in physics seem to be increasing as students work through the introductory physics sequence. Females are more likely to leave the physics major than males. While none of these differences is particularly large on its own, females are consistently lagging behind males. Valian refers to this building up of deficiencies as an "accumulated disadvantage" [59]. Small, consistent differences can build up and accrue over time to result in large disparities.

Rather than identifying a single factor that is responsible for gender disparities in physics participation, we find small gender differences across several different factors, including retention, performance, and attitudes and beliefs. Female students consistently fall behind males in each of these areas as they move through the introductory physics sequence. This pattern of disadvantage suggests a systematic culture in which males are privileged over females. Tatum refers to this cultural bias as a "smog of bias" [60], a smog that surrounds us and that we constantly breathe in, though at times we may be unaware that it even exists. Understanding that retention, performance, and attitudes and beliefs are some of the mechanisms by which the cultural bias is maintained and reinforced is a first step towards alleviating the gender disparities in physics. By creating new cultural norms in our classrooms that are inclusive and supportive of all students [61], we may begin to construct physics classrooms and physics cultures in which males and females can participate equally.

ACKNOWLEDGMENTS

We gratefully acknowledge valuable assistance from the members of the Physics Education Research group at Colorado, the Discipline-Based Education Research group at Colorado, and the following physics faculty: D. Anderson, M. Dubson, V. Gararie, M. Gustafson, M. Holland, H. Lewandowski, J. Nagle, C. Rogers, and S. Zhong. We would also like to thank our colleagues in the Psychology department: G. Cohen, T. Ito, and A. Miyake. This work is supported by the APS, AAPT PhysTEC project, the National Science Foundation, and the University of Colorado. This material is based on work supported by the National Science Foundation under Grant No. REC 0448176, CAREER: Physics Education and Contexts of Student Learning and NSF Grant No. 0833364, I3: Towards a Center for STEM Education.

-
- [1] National Science Foundation, Division of Science Resources Statistics, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2009*, NSF 09-305, (Arlington, VA; January 2009). Available from <http://www.nsf.gov/statistics/wmpd/>
- [2] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Phys. Rev. ST Phys. Educ. Res.* **3**, 010107 (2007).

- [3] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009).
- [4] E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- [5] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- [6] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for

- introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [7] P. Laws, P. Rosborough, and F. Poody, Women's responses to an activity-based introductory physics program, *Am. J. Phys.* **67**, S32 (1999).
- [8] M. Schneider, Encouragement of women physics majors at Grinnell college: A case study, *Phys. Teach.* **39**, 280 (2001).
- [9] *Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics, and Engineering*, edited by S. V. Rosser (Teacher's College Press, New York, 1995).
- [10] National Research Council, *To Recruit and Advance: Women Students and Faculty in Science and Engineering* (National Academies Press, Washington, DC, 2006).
- [11] D. Halpern, J. Aronson, N. Reimer, S. Simpkins, J. Star, and K. Wentzel, *Encouraging Girls in Math and Science* (NCER 2007-2003). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education (2007). Retrieved from <http://ncer.ed.gov>
- [12] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- [13] M. Lorenzo, C. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, *Am. J. Phys.* **74**, 118 (2006).
- [14] R. 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).
- [15] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *Am. J. Phys.* **60**, 627 (1992).
- [16] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- [17] J. Docktor and K. Heller, "Gender Differences in Both Force Concept Inventory and Introductory Physics Performance" in *Proceedings of the 2008 Physics Education Research Conference*, edited by C. Henderson, M. Sabella, and L. Hsu (AIP Conference Proceedings, New York, 2008) p. 15.
- [18] G. Carter, S. L. Westbrook, and C. D. Thompkins, Examining Science Tools as Mediators of Students' Learning about Circuits, *J. Res. Sci. Teach.* **36**, 89 (1999).
- [19] J. B. Kahle, L. H. Parker, L. J. Rennie, and D. Riley, Gender Differences in Science Education: Building a Model, *Educ. Psychol.* **28**, 379 (1993).
- [20] C. P. Russel and D. P. French, Factors affecting participation in traditional and inquiry-based laboratories, *J. Coll. Sci. Teach.* **31**, 225 (2002).
- [21] R. H. Tai and P. M. Sadler, Gender differences in introductory undergraduate physics performance: University physics versus college physics in the USA, *Int. J. Sci. Educ.* **23**, 1017 (2001).
- [22] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, *Sci. Educ.* **91**, 847 (2007).
- [23] P. M. Sadler and R. H. Tai, Success in introductory college physics: The role of high school preparation, *Sci. Educ.* **85**, 111 (2001).
- [24] D. E. Meltzer, "The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores," *Am. J. Phys.* **70**, 1259 (2002).
- [25] CAPA, <http://www.lon-capa.org/>; Mastering Physics, <http://www.masteringphysics.com/>
- [26] V. Otero, N. D. Finkelstein, R. McCray, and S. Pollock, Who is responsible for preparing science teachers? *Science* **313**, 445 (2006).
- [27] 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).
- [28] C. Turpen and N. D. Finkelstein, Not all interactive engagement is the same: Variations in physics professors' implementation of *Peer Instruction*, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020101 (2009).
- [29] In general, two professors are assigned to each of the large-enrollment introductory physics courses at CU. One professor (the lead) is responsible for the lecture part of the course and the other professor (the back-up) is responsible for the recitation part of the course. The back-up instructor usually does not have much, if any, interaction with students, while the lead instructor is who the students see in lecture. This being the case, we will only talk about the role of the lead instructor in these courses.
- [30] 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).
- [31] The pretest BEMA data for this course are reasonably stable (mean scores between 25% and 28%, standard deviations are about 9%) from semester to semester, which is why some faculty have chosen not to give the BEMA pretest to their students.
- [32] Data on students' high school experience (high school courses, high school GPA, SAT-math, and ACT-math scores), student demographic data (gender, ethnicity, and declared major), and course data (semester of enrollment and course grade) were collected from the Office of Planning, Budget, and Analysis at CU-Boulder. None of these data were self-reported.
- [33] While we recognize that there are other ways of quantifying a students' mathematics background, the most objective measures that were available to use were SAT-and ACT-Math scores and years of high school calculus. We found in the regression analyses that SAT-and ACT-Math scores were better predictors of student performance than years of high school calculus, so we chose to use those measures as a proxy for mathematics background.
- [34] For example, a student's SAT-math score S was converted to a z -score using the transformation $z_S = (S - \langle S \rangle) / SD$, where $\langle S \rangle$ is the average SAT-math score for all students and SD is the standard deviation of the SAT-math scores.
- [35] W. K. Adams, K. K. Perkins, N. 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).
- [36] L. McCullough, Gender, context, and physics assessment, *J. Int. Women's Stud.* **5**, 20 (2004).
- [37] T. L. McCaskey, M. H. Dancy, and A. Elby, "Effects in assessment caused by splits between belief and understanding" in *Proceedings of the 2003 Physics Education Research Conference*, edited by J. Marx, S. Franklin, and K. Cummings (AIP Conference Proceedings, New York, 2004) p. 37.
- [38] E. Hazel, P. Logan, and P. Gallagher, Equitable assessment of students in physics: importance of gender and language back-

- ground, *Int. J. Sci. Educ.* **19**, 381 (1997).
- [39] The effect size for males is calculated as $ES_M = (\langle S \rangle_{Phys1\&2} - \langle S \rangle_{Phys1}) / SD_M$, where $\langle S \rangle_{Phys1\&2}$ and $\langle S \rangle_{Phys1}$ are the average values for males who took both Physics 1 and 2, and for males who took Physics 1 only, respectively. SD_M is the standard deviation for all males who took Physics 1. A similar calculation is used for the effect size of the females.
- [40] The effect size is calculated as $ES = (\langle S \rangle_M - \langle S \rangle_F) / SD$, where the SD for all students is used.
- [41] R. Chabay and B. Sherwood, Restructuring the introductory electricity and magnetism course, *Am. J. Phys.* **74**, 329 (2006).
- [42] M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. W. Chabay, L. Ding, M. P. Haugan, M. J. Marr, B. A. 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).
- [43] The normalized learning gain that we are using here is the gain for each individual student, calculated as $g = (\text{post} - \text{pre}) / (100 - \text{pre})$.
- [44] The post-test scores for the FMCE and BEMA are slightly different. The mean FMCE post-test score is 65% and the mean BEMA post-test score is 56%. Even when we normalize the gender gaps by these post-test scores, the trends are the same.
- [45] At CU, students can drop a course within the first two and half weeks of the semester and the course will not appear on their transcript. If students drop the course after two and a half weeks, a grade of W (withdraw) will appear on their transcript. It is only these students, who receive a grade of W for the course, that are included in this analysis.
- [46] We presented here only average CLASS shifts over all Physics 2 semesters. While there are some differences in CLASS shifts by semester, most semesters show trends similar to the averages we have presented here. One exception is Semester A2, in which both males and females showed significant positive gains overall.
- [47] Z. Hazari, P. M. Sadler, and R. H. Tai, Gender differences in the high school and affective experiences of introductory college physics students, *Phys. Teach.* **46**, 423 (2008).
- [48] S. E. Carrell, M. E. Page, and J. E. West, "Sex and science: How professor gender perpetuates the gender gap," NBER Working Paper 14959 (Cambridge, MA; May 2009). Available from <http://www.nber.org/papers/w14959>
- [49] In our previous work (Ref. [3]), a similar multiple-regression analysis on the FMCE posttest yielded an R^2 value of 0.44. Similar analyses conducted by other researchers have R^2 values of 0.16 (Ref. [21]), 0.28 (Ref. [22]), and 0.36 (Ref. [23]). The R^2 value reported here of 0.34 is within this range.
- [50] Though not part of the data set used for this analysis, Professor F taught again in a more recent semester. We did not observe the same trends with respect to course grades or DFW rates the second time that she taught as the first time. While we did not find consistent evidence of a positive impact of a female professor on the performance of females in the course, we still believe it is an issue that deserves further study.
- [51] D. M. Marx and J. S. Roman, Female Role Models: Protecting Women's Math Test Performance, *Pers. Soc. Psychol. Bull.* **28**, 1183 (2002).
- [52] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, "Unpacking Gender Differences in Students' Perceived Experiences in Introductory Physics" in *Proceedings of the 2009 Physics Education Research Conference*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conference Proceedings, New York, 2009), p. 177.
- [53] C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance, *Am. Psychol.* **52**, 613 (1997).
- [54] S. J. Spencer, C. M. Steele, and D. M. Quinn, Stereotype threat and women's math performance, *J. Exp. Soc. Psychol.* **35**, 4 (1999).
- [55] C. M. Steele, *Advances in Experimental Social Psychology*, edited by L. Berkowitz (Academic Press, New York, 1988), p. 261.
- [56] G. L. Cohen, J. Garcia, N. Apfel, and A. Master, Reducing the racial achievement gap: A social-psychological intervention, *Science* **313**, 1307 (2006).
- [57] A. Martens, M. Johns, J. Greenberg, and J. Schimel, Combating stereotype threat: The effect of self-affirmation on women's intellectual performance, *J. Exp. Soc. Psychol.* **42**, 236 (2006).
- [58] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, "Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation" (to be published).
- [59] V. Valian, *Why So Slow: The Advancement of Women* (MIT, Cambridge, 1999).
- [60] B. D. Tatum, *Why Are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race* (Basic Books, New York, 1997).
- [61] B. L. Whitten, S. R. Foster, M. L. Duncombe, P. E. Allen, P. Heron, H. M. Zorn, L. McCullough, K. A. Shaw, and B. A. P. Taylor, What Works? Increasing the Participation of Women in Undergraduate Physics, *J. of Women and Minorities in Science and Engineering* **9**, 239 (2003).