

**Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?**

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

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

(Received 2 March 2007; published 5 June 2007)

Previous research [Lorenzo *et al.*, *Am. J. Phys.* **74**, 118 (2006)] demonstrated that the difference in performance between male and female students can be reduced and even eliminated, in consistent fashion, by using interactive engagement techniques in the introductory physics classroom. The present paper describes similar studies in a different, large research university and finds that the use of interactive engagement techniques does not necessarily reduce the gender gap. Furthermore, in the environments studied, there is a gap in learning gains between male and female students ( $p < 0.01$ ) whether partially or fully interactive classroom techniques are used. Our findings suggest that engaging students in interactive educational environments is not sufficient to reduce the gender gap, and we find instances where despite significant learning gains by all students, the gender gap is increased. There is indication that there are both student and instructor effects that impact the gender gap, which are the subjects of ongoing studies.

DOI: [10.1103/PhysRevSTPER.3.010107](https://doi.org/10.1103/PhysRevSTPER.3.010107)

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

**INTRODUCTION**

In a significant piece, Lorenzo, Crouch, and Mazur establish that, by using interactive engagement techniques at Harvard, the disparity in performance on the Force Concept Inventory (FCI) between males and females (gender gap) is reduced and even eliminated.<sup>1</sup> Furthermore, they demonstrate that in this environment the degree of interactivity in the classroom affects the degree to which the gender gap is reduced; the more interactive the class, the more the gender gap is reduced. The authors note that these classes are consistent with a variety of practices called for in the literature and that, because their results are consistent across five different instructors, the changes in student performance are due to the teaching strategies and pedagogical approach rather than the instructor.

Following that work, researchers at the University of Colorado have conducted similar studies examining the impact of interactive engagement techniques on the gender gap. We report on studies designed to replicate the original results in a first-semester course (mechanics) as well as studies from a second-semester course (electricity and magnetism). Our findings suggest that simply engaging students in interactive educational environments is not sufficient to reduce the gender gap, and we find instances where, despite significant learning gains by all students, the gender gap is increased. There is indication that there are both student and instructor effects that impact the gender gap.

**METHODS**

The studies at the University of Colorado (CU) occurred in the calculus-based, large-scale ( $N=400-600$ ) introductory physics sequence. Data were collected in six first-semester courses (spring 2004 to fall 2006) and five second-semester courses (fall 2004 to fall 2006), and represent over 3000 students with matched pre and post data. Conceptual learning was assessed with the Force and Motion Concept Evaluation (FMCE)<sup>2</sup> for the mechanics course and the Brief Electricity and Magnetism Assessment (BEMA)<sup>3</sup> in the electricity and

magnetism (EM) course. The first-semester courses in this study employed varying degrees of interactive engagement (IE) similar to those in the Harvard study. IE2 courses employed student discussions around ConcepTests,<sup>4</sup> online homework systems,<sup>5</sup> *Tutorials*,<sup>6</sup> and voluntary help-room sessions on problem-solving homework. The class environment is described in detail in prior work.<sup>7</sup> These are similar though not identical to the IE2 conditions of the Harvard study. IE1 courses were in the same environment but did not include the *Tutorials*; rather, these classes included traditional recitation sessions where TAs and students worked homework problems individually with occasional sessions of small-group work around materials from the textbook publisher. Six different faculty members participated in the study over six course offerings; the faculty varied widely in familiarity with physics education research.<sup>8</sup> The second-semester study only included IE2 courses, where *Tutorials* were used in conjunction with the other IE course elements. Five different faculty participated in the five different course offerings. From a coarse perspective, each of the IE1 courses was similar to the other IE1 courses in its use of tools and teaching strategies and the IE2 courses were also similar to one another in their implementation.<sup>8</sup> Some structural features of the courses studied are listed in Table I.

There are notable differences in the present study from the Harvard study. The student populations are different. While students post significant (middle- to high-gain<sup>9</sup>) normalized learning gains, the students start at approximately 30% vs the Harvard 65% and end with scores of approximately 65% vs 85% at Harvard. The measurement instrument used in the first-semester study was the FMCE rather than the FCI; notably, however, there is a high correlation for student performance<sup>10</sup> on these two instruments.<sup>11</sup> The class size is roughly three times that of the Harvard classes and the gender ratio [male to female ( $M/F$ )] is almost double that of the Harvard study, shown in Table II. Also, the CU classes did not explicitly employ the cooperative quantitative problem-solving techniques used in the Harvard classes, which had at least one more contact hour per week.

TABLE I. Summary of the introductory physics courses included in the study.

Group	Topic	Semester	Lectures	Sections
IE1	Mechanics	Fa04, <sup>a</sup> Sp05, Fa05	Peer instruction	Mostly traditional
IE2	Mechanics	Sp04, Sp06, Fa06	Peer instruction	Tutorials (Ref. 6)
IE2	EM	Fa04, Sp05, Fa05, Sp06, Fa06	Peer instruction	Tutorials (Ref. 6)

<sup>a</sup>Included some informal small-group activities.

**RESULTS**

The results of FMCE pre- and post-test scores for male and female students are averaged for each of the first semester courses and appear in Table II.

While there are individual cases where the gender gap is reduced, on average we see no statistically significant reduction in the gender gap in the IE2 or IE1 classes.<sup>13</sup>

Average results of the pre- to postinstruction gender gap for the two interactive engagement style courses are shown in Fig. 1, which parallels Fig. 1 from the Harvard study.<sup>1</sup>

Furthermore, while we do observe the effect that IE2 courses have higher normalized learning gains than IE1 courses, there is still a statistically significant disparity in these normalized gains between male and female students (see Fig. 2,  $p < 0.01$  via two-tailed  $T$ -test), in both IE1 and IE2 classes, unlike the Harvard study. Notably, there is less of a gender disparity in normalized learning gains for the IE2 case than the IE1 case. Additionally, the female students in IE2 environments on average make normalized gains as high as males in IE1 classes (Fig. 2); however, within treatment groups, the male students make greater learning gains than the female students.

In the second-semester study (see Table III) we observe a small gender gap on the pretest scores (averaging over all students for all five terms, the average pretest gap is  $1.8 \pm 0.5\%$ ). On the post-test we observe a statistically significantly higher gender gap (average =  $6.1 \pm 1.0\%$ ). As with the first semester, the males make more substantial learning gains than females. However, the gender gap that does appear is less substantial than those that appear in the first-semester course.

**DISCUSSION AND CONCLUSIONS**

Though desirable, we cannot conclude that interactive engagement techniques will eliminate or even necessarily reduce the gender gap. Of course, we do not deny the significant findings of the Harvard study; however, we believe prior results indicate that interactive engagement techniques are necessary, but not sufficient for addressing the gender gap. Depending upon how one interprets the notion of “teaching strategy” we disagree with the Harvard claim that there are no instructor effects (Ref. 1, p. 119). Rather, we believe that the instructor plays critical roles in at least two essential ways: (1) different instructors may enact these teaching strategies differently; and (2) instructors set the tone, norms of behavior, and attitude in the class both explicitly and implicitly. In the first case, while many instructors may engage in the use of ConcepTests, how, when, and why students engage will vary by instructor. In the second case, the culture of a classroom is established both by the practices selected and by how those practices are framed by the instructor.<sup>7,14</sup>

Ongoing studies of faculty practices delineate these instructor-based variations. For instance, while all implementations of the IE techniques in this study encourage student discussion, the purpose of these discussions varies by course offering—some courses emphasize getting the right answer while others emphasize the role of reasoning and sense making. Identifying which of these class norms are aligned with the findings of research on gender-based performance (such as those listed in Ref. 1) and how these norms and practices correlate with student performance are the subjects of current studies. Finally, we note that the students themselves play critical roles in establishing classroom norms and how effective IE practices are. The students in this study are clearly

TABLE II. FMCE (Ref. 2) data for the introductory calculus-based physics course at CU;  $N^{M(F)}$  is the number of males (females) for whom there are matching FMCE pre- and post-test data;  $S_i^{M(F)}$  is the average pretest score for males (females) [using Thornton’s scoring rubric (Ref. 10)]; and  $S_f$  is the average post-test score. The numbers in parentheses represent the standard error of the mean, not the standard deviation.  $S^M - S^F$  equals the gender gap (Ref. 12).

Group	Term	$N^M$	$N^F$	FMCE pretest score (%)			FMCE post-test score (%)		
				$S_i^M$	$S_i^F$	$S_i^M - S_i^F$	$S_f^M$	$S_f^F$	$S_f^M - S_f^F$
IE1	F04	216	80	36 (1.6)	25 (1.8)	11	74 (1.8)	58 (3.2)	16
	S05	146	61	31 (1.7)	22 (1.7)	9	61 (2.3)	52 (3.5)	9
	F05	210	74	35 (1.5)	24 (2.0)	11	63 (2.0)	48 (3.0)	15
IE2	S04	280	104	30 (1.3)	18 (1.4)	12	75 (1.5)	66 (2.8)	9
	S06	205	61	28.5 (1.4)	21.5 (2.2)	7	63 (1.9)	51 (3.8)	12
	F06	235	84	37 (1.6)	23 (1.8)	14	69.5 (1.8)	60 (3.2)	9

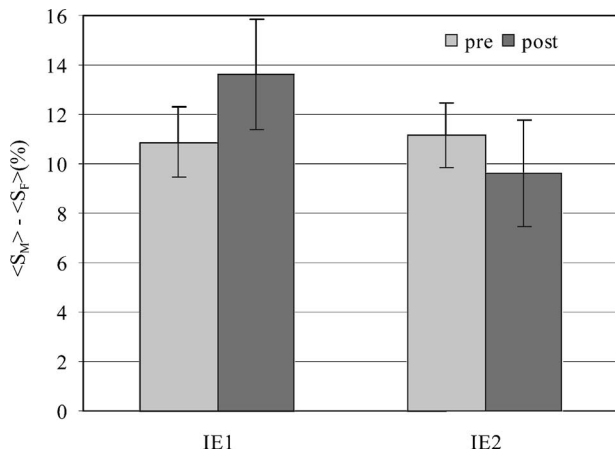


FIG. 1. Gender gap. Effect of instructional approach on gender gap in student performance on FMCE for CU students averaged across three semesters for IE1 (partially interactive) and three semesters IE2 (fully interactive). Error bars represent standard errors of the mean. There is no statistically significant shift in gap of performance by gender for either IE1 or IE2.

different from those in the Harvard study. The data presented in this paper are drawn from a population of students who start and end the class with notably lower scores on the conceptual survey (FMCE), which may be indicative that student pretest score (and prior knowledge) may be a critical factor in the observed effects of learning gain. Additionally, other research has suggested that the effectiveness of physics education research curricula may depend upon student population and background.<sup>7,15,16</sup> Understanding the role and impact of student background and preparation is the subject of current study.

Despite the difference in performance on the conceptual surveys, in no course in the Colorado study is there a gender gap in overall grade. We do observe competing effects by gender on student performance on exams and homework. Female students outscore their male counterparts on homework and participation scores; whereas the males perform better on the exams, in a fashion consistent with the gender gap observed on the conceptual surveys.<sup>17</sup> Notably, the homework is designed to be collaborative and is not time sensitive; whereas the exams are individual, more competi-

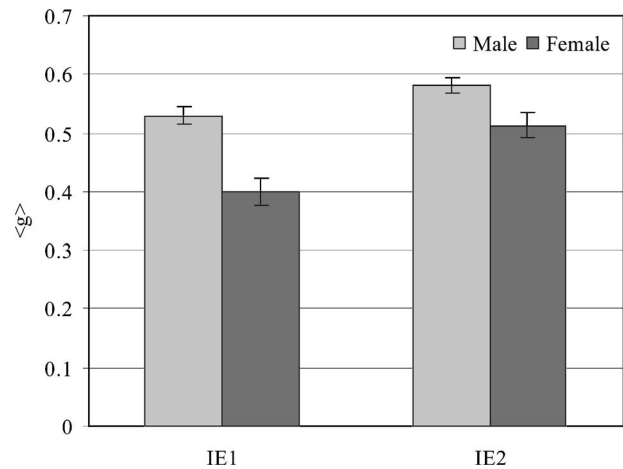


FIG. 2. FMCE normalized learning gain averaged over all matched students in IE1 and IE2 courses. Effect of instructional approach on normalized pre- to post-test gain  $\langle g \rangle$  on the FMCE for CU students. Error bars represent standard errors of the mean. There is a statistically significant difference ( $p < 0.01$ ) in the normalized learning gains of male and female students for both IE1 and IE2 cases.

tive, and time constrained. These features may contribute to the observed gender disparity and are consistent with the research on gender difference cited in Ref. 1.

In order to distinguish those factors of our introductory physics classes that contribute to or diminish the gender gap, we must certainly examine which techniques are used, but also more thoroughly examine how these interactive engagement techniques are enacted by faculty and students, and understand the broader class culture that frames these practices. These are subjects of current studies.

ACKNOWLEDGMENTS

We gratefully acknowledge valuable assistance from the members of the Physics Education Research group at Colorado, the inspiration of the original study authors, M. Lorenzo, C. Crouch, and E. Mazur and follow-up discussions with C. Crouch. This work is supported by the APS, AIP, AAPT PhysTEC project, the National Science Foundation, and the University of Colorado.

TABLE III. BEMA (Ref. 3) data for the second-semester calculus-based physics course at CU;  $N^{M(F)}$  is the number of males (females) for whom there are matching BEMA pre- and post-test data;  $S_i^{M(F)}$  is the average pretest score (using the BEMA scoring rubric) for males (females); and  $S_f$  equals the average post-test score. The numbers in parentheses represent the standard error of the mean.  $S_i^M - S_i^F$  equals the gender gap.

Group	Term	$N^M$	$N^F$	BEMA pretest score (%)			BEMA post-test score (%)		
				$S_i^M$	$S_i^F$	$S_i^M - S_i^F$	$S_f^M$	$S_f^F$	$S_i^M - S_i^F$
IE2	F04	249	79	26 (0.5)	23 (0.8)	3	60 (.9)	55 (1.5)	5
	S05	165	61	28 (0.8)	26 (1.2)	2	61 (1.2)	54 (1.9)	7
	F05	221	92	25 (0.7)	25 (0.8)	0	51 (1.1)	47 (1.7)	4
	S06	144	49	27 (0.7)	24 (1.0)	3	56 (1.3)	47.5 (2.1)	8
	F06	259	76	27 (0.6)	25 (1.1)	2	58 (1.0)	51 (1.9)	7

- <sup>1</sup>M. Lorenzo, C. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, *Am. J. Phys.* **74**, 118 (2006).
- <sup>2</sup>R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation, *Am. J. Phys.* **66**, 338 (1998).
- <sup>3</sup>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).
- <sup>4</sup>E. Mazur, *Peer Instruction: A Users Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- <sup>5</sup>CAPA, <http://www.lon-capa.org/>; Mastering Physics, <http://www.masteringphysics.com/>.
- <sup>6</sup>L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- <sup>7</sup>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).
- <sup>8</sup>Each course included a lead instructor and a backup instructor as described in Ref. 7. While team taught, the duties in the class administration were divided, and each professor took the lead role of the course only one time. An attempt was made to mentor future lead-faculty members by having them assume the backup roles the semester prior to assuming lead instructional roles. This instructional partnering ensured that the IE techniques were enacted seriously and with some consistency; however, variation in instructional practice did occur and is the subject of current investigation.
- <sup>9</sup>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).
- <sup>10</sup>R. Thornton, D. Kuhl, K. Cummings, and J. Marx (unpublished); K. Cummings, J. Marx, R. Thornton, and D. Kuhl (unpublished).
- <sup>11</sup>Combining these differences (student pre- and post-test range and the instrument used to evaluate student performance), we note that the CU students tend not to saturate the instrument, whereas the Harvard students appear to, which may be a contributing factor in the Harvard results.
- <sup>12</sup>Note the small ( $\sim$  five point), but consistent, difference between the fall and spring pretest scores. These indicate slightly different demographic pools for on- and off-sequence course offerings.
- <sup>13</sup>Of further note, we observe no consistent trend where the gender gap increases or decreases depending upon overall learning gain for the course (see Table II).
- <sup>14</sup>N. D. Finkelstein, Learning physics in context: A study of student learning about electricity and magnetism, *Int. J. Sci. Educ.* **27**, 1187 (2005).
- <sup>15</sup>M. S. Sabella, in *Proceedings of the 2002 Physics Education Research Conference*, edited by S. Franklin, K. Cummings, and J. Marx (PERC Publishing, New York, 2002), p. 79; <http://piggy.rit.edu/franklin/perc2002/Sabella.pdf>
- <sup>16</sup>V. Coletta and J. Phillips, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, *Am. J. Phys.* **73**, 1172 (2005).
- <sup>17</sup>Differences depend on courses, but, for example, in F04 males scored four points higher than females on in-class exams (statistically significant at  $p < 0.01$ ) but when averaged with homework and participation scores, women's and men's final course scores differed by less than half a point, not a significant difference.