Investigating the Source of the Gender Gap in Introductory Physics

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

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

Abstract. Our previous research showed that despite the use of interactive engagement techniques at our institution, the difference in performance between men and women on a conceptual learning survey persisted from pre to post test. This paper reports on a three-part follow-up study that investigates what factors contribute to the gender gap. First, we analyze student grades in different components of the course and find that men and women's course grades are not significantly different ($p>0.1$), but men outscore women on exams and women outscore men on homework and participation. Second, we compare average post test scores of men and women who score similarly on the pretest and find that there are no significant differences between men and women's average post test scores. Finally, we analyze other factors in addition to the pretest score that could influence the post test score and find that gender does not account for a meaningful portion of the variation in post test scores when a measure of mathematics performance is included. These findings indicate that the gender gap exists in interactive physics classes, but may be due in large part to differences in preparation, background, and math skills.

Keywords: gender, conceptual learning, introductory physics, physics education research
PACS: 01.40.Fk, 01.40.G-, 01.40.gb, 01.75.+m

INTRODUCTION

Recently, there has been great interest in the gender gap, the performance difference between men and women, in science. Several studies have suggested that both men and women learn more in interactive and engaging educational environments, but these techniques may disproportionately benefit women. [1,2] Researchers at Harvard were able to show that a pre-instruction gender gap was eliminated when fully interactive engagement techniques were employed. [3]

In our previous work [4], attempting to replicate the Harvard study, we found that engaging students in interactive educational environments did not always reduce the gender gap – IE techniques are not sufficient to reduce the gender gap. Our results suggest that a variety of factors are likely to contribute to men and women's differential performance. Not only which practices are used but how they are enacted appears to be critical. Furthermore, we hypothesize the background of the students plays a significant role [5], and this is the subject of our current investigation. Our preliminary follow-up studies involve three parts, and we report on each here.

We first question whether course components differentially affect males and females in our own introductory courses. By analyzing exam, homework, participation, and total course grades by gender, we find that there are differences between men and women on exams, homework, and participation, but these differences offset one another resulting in no difference between men and women's course grades.

In the second component of the study we ask: is the difference that we see in the post test scores between men and women a result of gender, or is the variation due to some other factor? We compare post test scores of men and women who have similar pretest scores. We see no substantial differences between men and women who score similarly on the pretest, indicating that pretest score is more relevant than gender.

The third part of the study asks what other factors influence the outcome of the post test. Preliminarily, we are interested in math skill and gender. Using multiple regression analysis we determine that gender does not account for a meaningful portion of the variation in post test scores. These results persist when we explicitly include a measure of mathematics performance. Our findings are in line with Meltzer, who found that mathematical skill correlated with student conceptual learning gains. [5]

TABLE 1. Analysis of students' course grades. Each column contains the difference between the average scores for men and women ($\langle S\rangle_M$ - $\langle S\rangle_F$). Standard error of the mean is shown in parenthesis. Those courses for which there is no participation listed included it in the homework grade, and participation alone could not be extracted. The * indicates statistically significant (via two-tailed t-test, $p<0.05$).

	Participation	**Homework**	**Exams**	**Course Grade**
Spring 04	-6.0 (1.4) *	-5.6 (1.6) *	5.2 (1.5) *	1.1(1.2)
Fall 04		-6.95 (1.9) *	4.0 (1.5) *	0.4 (1.4)
Spring 05		-6.0 (1.6) *	3.9 (1.4) *	0.5 (1.2)
Fall 05		-5.6 (1.8) *	3.9 (1.4) *	0.6 (1.3)
Spring 06	-2.0 (2)	-3.3 (1.8)	4.6 (1.5) *	1.7 (1.3)
Fall 06	-7.5 (1.6) *	-2.2 (1.7) *	4.4 (1.4) *	1.8 (1.3)
Spring 07	-2.7 (1.6)	-2.0 (2.0)	3.5 (1.3) *	1.5 (1.3)
Average	-4.6 (0.8)	-4.5 (0.7)	4.2 (0.5)	1.1 (0.5)

METHODS

The data in all studies were collected from seven offerings (spring 2004 to spring 2007) of the first semester, calculus-based mechanics course at the University of Colorado (CU). All seven classes used interactive engagement techniques, some to a higher degree than others. Each of the seven classes employed student discussions around ConcepTests[6], online homework systems[7], and voluntary help-room sessions on problem-solving homework. Four of the seven classes used *Tutorials in Introductory Physics* [8] during a one-hour per week recitation, while the remaining three classes held more traditional recitation sections.

Several measures were used to assess student performance in the course and preparation or background. Conceptual performance was assessed using the Force and Motion Concept Evaluation (FMCE) [9]. Only students with matched pre and post test data are included, N ~ 2100 students. An applied math test was used to assess students' math skills upon entering the physics course. This test has been administered for many years to incoming engineering students at CU to identify at-risk students going into the calculus I courses. We have these data for the subset of students who took a calculus course in the Applied Math department (N = 965). To analyze student overall performance in the physics class, we analyzed students' homework, exam, participation, and course grades.

RESULTS

Course Grades Analysis

Our initial investigation into the gender gap looks at course grades to determine which, if any, components of the course differentially affect men and women. For all seven semesters of the mechanics course men and women's scores are averaged on homework, participation, exams, and total course grade. Though each course varies in the weights they assign to different components of the course, typically exams make up 60% - 65% of the course grade, homework counts for 25% - 35%, and participation makes up the remainder. The difference between the average men and women's scores in each component ($\langle S\rangle_M$ - $\langle S\rangle_F$) is calculated for each class. These differences, along with the average differences, are shown in Table 1. For several courses the participation grade was included in the homework grade and could not be extracted.

In the past seven semesters there has been no significant (via two-tailed t-test, $p<0.05$) gender difference in total course grade. Men outscore women by about 4 points on exams and women outscore men by about 5 points each on homework and participation. These differences offset one another and result in course grades that are not significantly different.

Matched Pretest Analysis

The second part of the study looks at comparing students who have similar pretest scores. Students are binned by FMCE pretest score (each bin contains about equal numbers of students, N ~ 420), and then the average FMCE post test score is calculated for men and women in each bin. The results are plotted in Fig. 1. The same trends that are described below exist for a range of reasonable bin sizes.

We observe that students who have similar pretest scores have similar post test scores, regardless of gender. There were no statistically significant differences (as measured by two-tailed t-test, $p>0.1$) in any individual bin, i.e. between men and women who scored similarly on the pretest. Though the differences in each bin are not significant, males consistently score higher than females in all bins. We also see a correlation ($r = 0.562$) between FMCE pre and post test score.

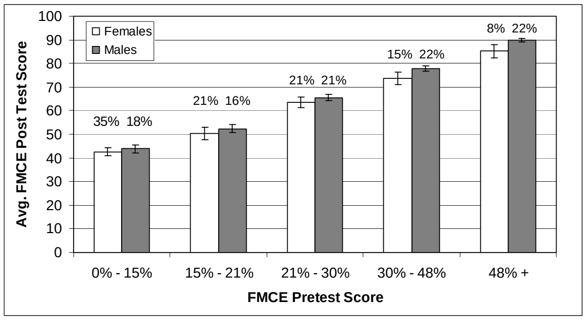

Figure 1. Average FMCE post test scores for women and men with matched FMCE pretest scores. The percentages above each bar represent the percentage of the women (or men) from the total in each bin. The error bars represent the standard error on the mean. There are no significant differences between men and women in any individual bin.

The error bars on the plot are only a lower limit on the actual error. These error bars do not account for sources of error other than statistical error, such as systematic error or sampling bias. They also do not account for differences in class practices from semester to semester and across different instructors. Regardless, we find that the differences between men and women are not significant.

We do find that a higher percentage of the women fall into the low pretest bins. The percentages above each bar in the plot represent the percent of the women, or men, from the total that fall into that bin. 56% of women versus 34% of men fall into the lowest two pretest bins, while 23% of women versus 44% of men fall into the highest two pretest bins.

The same trend exists for *normalized* learning gain; students with similar pretest scores have similar normalized gains, regardless of gender. We also see a correlation, albeit weaker ($r = 0.281$), between FMCE pretest score and normalized gain.

Multiple Regression Analysis

Our results above suggest that the FMCE pretest could be a significant factor in predicting students' post test score. Using multiple regression analysis, we can determine what other pre-factors are significant in regards to predicting the post test score. Preliminarily, we are interested in math skill and gender.

To analyze the impact of math skill and gender on post test score, we perform a stepwise multiple regression analysis where variables are included in the regression one at a time. We are interested in whether the addition of variables, applied math test score and gender, accounts for the variation in post-test. [10] The stepwise multiple regression analysis includes only those students for whom we have FMCE pre and post test data and applied math test data, N = 965.

We start by including only the FMCE pretest as an independent variable. The results are shown in Table 2. The FMCE pretest alone accounts for only 34% of the variation in post test scores. As a measure of students' math skills, we include an applied math test as an independent variable in the regression analysis. This regression has a multiple R value of 0.651. The additional variation accounted for by the math test is significant (via F-test, $p<0.01$). The FMCE pretest and the applied math test together account for 42% of the variance in post test scores.

TABLE 2. Multiple Regression. The FMCE post test is the dependent variable and the FMCE pretest, the applied math test, and gender are the independent variables.

	Multiple R	$\mathbf{R^2}$
FMCE pretest	0.583	0.340
FMCE pretest and applied math test	0.651	0.424
FMCE pretest, applied math test, and gender	0.655	0.429

We then include gender along with the FMCE pretest and applied math test in the regression of the FMCE post test. The multiple R value for this regression is R = 0.655, accounting for 43% of the variance in post test scores. The additional variation accounted for by gender is very small (although still statistically significant via F-test, $p<0.05$). Gender contributes to less than 1% of the variance in the observed post test scores. We find that the explicit inclusion of gender as a variable is not a principal factor in accounting for variation in students' post test scores when other measures of students' background and preparation are included.

The multiple regression analysis is repeated looking instead at the normalized learning gain. Although the multiple R values were lower (as we expect since post test score is more highly correlated with pretest score than normalized gain is), the same trends exist. The correlation of normalized gain with pretest alone is R=0.351. When adding math scores, the multiple R value becomes 0.487. Upon adding gender, it stays nearly the same, R=0.493. Thus, the multiple regression analysis shows that explicitly including gender does not additionally account for much of the variation in the learning measures.

DISCUSSION AND CONCLUSIONS

The analysis of course grades suggests that which course practices we engage in and how they are enacted do not differentially affect men and women in overall course grade. When we look at men and women with similar pretest scores, we see only small differences in average post test scores. This suggests that the gender gap we observe is due in large part to a gap in preparation between men and women. Women come in with lower pretest scores, and because of the significant correlation between pre and post test scores, women have lower post test scores.

Finally, when we conduct a multiple regression analysis, the applied math test and FMCE pretest score predict the post test score just about as well as when we explicitly include gender as an additional variable. These results again suggest that the gender gap is due in large part to differences between men and women in preparation and background.

In one sense, it may be interpreted that gender does not play a role in measures of student achievement – the variation in FMCE post test score may be attributed to other variables, notably pretest score and math achievement [11]. Furthermore, overall course grades are similar for male and female students. Such a stance would suggest that there is no explicit gender bias in the classes observed. Nonetheless, in these classes we observe a gap in performance by gender and observe instances where over the course of instruction, this gap is increased [4].

Another frame of analysis is that of implicit bias – that is, those components of a class that are most heavily weighted and essential for success disproportionately favor male students. While course grades are neutral overall, male students are more likely to score higher on exams (which are weighted more heavily in a typical class). While the classes studied are introductory courses with no expectation of prior knowledge of physics, those students who arrive to the class with greater background knowledge (higher pretest scores) are more likely to achieve high post test scores and greater learning gains. The class favors those students with stronger physics and math backgrounds – in this case, male students.

Such an arrangement of a class (or any social environment) plays to certain student backgrounds and when those backgrounds are correlated with particular demographic groups, it demonstrates bias. That is not to say this is an explicit or purposeful bias, but rather one that is the codified structure of systemic cultural bias. [12] Tatum refers to this as a 'smog of bias' [13] and others to the privileged preparation of some group (at the expense of others) as an "accumulated disadvantage" [14].

Recognizing that student preparation in physics or mathematics is a means by which this bias is propagated allows us as researchers and educators to proactively address the challenges of the gender gap in physics. Simply enacting research based reforms, or supporting current practices (the *status quo*) may improve aggregate student learning gains, but may also be promulgating the disparity of performance and lack of equity in our educational system.

ACKNOWLEDGMENTS

Thanks to PhysTEC (APS/AIP/AAPT), NSF CCLI (DUE0410744) and NSF CAREER (0448176). Thanks to Prof. Crouch, the CU Physics Department, and the PER at Colorado group for ongoing support.

REFERENCES AND NOTES

1. P. Laws et al. *Am. J. Phys.* **67**, S32 (1999).
2. M. Schneider *Phys. Teach.* **39**, 280 (2001).
3. M. Lorenzo et al. *Am. J. Phys.* **74**, 118 (2006).
4. S.J. Pollock et al. *Phys. Rev. ST PER* **3**, 010107 (2007).
5. D.E. Meltzer *Am. J. Phys.* **70**, 1259 (2002).
6. E. Mazur, *Peer Instruction: A Users Manual*, Upper Saddle River, NJ: Prentice-Hall, 1997.
7. CAPA, http://www.lon-capa.org/; Mastering Physics, http://www.masteringphysics.com/.
8. L.C. McDermott and P.S. Schaffer, *Tutorials in Introductory Physics*, Upper Saddle River, NJ: Prentice-Hall, 2002.
9. R.K. Thornton and D.R. Sokoloff, *Am. J. Phys.* **66**, 228 (1998).
10. Lomax, R.G., *An Introduction to Statistical Concepts for Education and Behavioral Sciences*, Mahwah, NJ: Lawrence Erlbaum Associates, 2001.
11. We also note that the identified variables only account for 42% of the variance in the data. Examining additional sources is the subject of current work.
12. Dancy, M., "The Myth of Gender Neutrality" in *2003 Physics Education Research Conference*, AIP Conference Proceedings 720, New York: American Institute of Physics, 2004, pp. 31-36.
13. B.D. Tatum, *Why are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race*, New York: Basic Books, 1997.
14. V. Valian, *Why So Slow: The Advancement of Women*, Cambridge: The MIT Press, 1999.