Gender Differences in Both Force Concept Inventory and Introductory Physics Performance

Jennifer Docktor and Kenneth Heller

School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455

Abstract. We present data from a decade of introductory calculus-based physics courses for science and engineering students at the University of Minnesota taught using cooperative group problem solving. The data include 40 classes with more than 5500 students taught by 22 different professors. The average normalized gain for males is 0.4 for these large classes that emphasized problem solving. Female students made up approximately 20% of these classes. We present relationships between pre and post Force Concept Inventory (FCI) scores, course grades, and final exam scores for females and males. We compare our results with previous studies from Harvard [2] and the University of Colorado [3,4]. Our data show there is a significant gender gap in pre-test FCI scores that persists post-instruction although there is essentially no gender difference in course performance as determined by course grade.

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

INTRODUCTION

Men typically outperform women on science achievement tests, even among students with similar math and science course backgrounds [1]. This *gender gap* in performance is small in early grades and becomes more pronounced through high school and college.

One response to this gap has been to alter the classroom environment to encourage the interactive engagement of all students. Recent studies at both Harvard [2] and the University of Colorado [3,4] have examined the effects of such methods of instruction on the gender gap in large calculus-based introductory physics courses.

The study at Harvard concluded that using Peer Instruction, *Tutorials in Introductory Physics*, and cooperative problem-solving activities decreased and in some cases eliminated the gender gap in Force Concept Inventory (FCI) exam scores. The University of Colorado (CU) found that implementing methods similar to those used at Harvard did not consistently reduce the gender gap in scores on the Force and Motion Concept Evaluation (FMCE). The variation in the gap at CU was attributed to instructors' differences in implementation. CU also suggested that differences in student background and preparation might play a role in conceptual learning gains.

In this paper we compare preliminary results from a decade of the University of Minnesota introductory physics course for science and engineering students taught using cooperative group problem solving.

COURSE DESCRIPTION

The first semester of calculus-based physics for scientists & engineers at the University of Minnesota has an average fall term enrollment of 800 students with an average class size of 185. The course consists of three 50-min lectures, a 2-hr laboratory, and one 50-min problem-solving discussion session per week. The laboratory and discussion sessions have 16-18 students per class and are taught by graduate student teaching assistants using Cooperative Problem Solving and "context-rich" problems appropriate for group work [5,6].

Cooperative Problem Solving is a structured environment in which students practice solving problems with their peers. Students work in groups of three (or if necessary, four) members with the same groups for laboratory and discussion sessions. In discussion sections, students co-construct a solution to a single context-rich problem. They are not permitted to consult their textbook or notes, but use their existing knowledge to explain, clarify, and justify their thinking to the other members in the group. The groups are always assigned by the instructor with first groups of a semester chosen essentially at random. Subsequent groups are structured to have students of mixed-performance based on their individual test performance (one high, medium, and low-ability student). When possible, all groups are structured to include at least two females. Groups change following each test (about 4 times a semester). During the discussion session preceding this test, each group co-constructs a single solution to a problem that is graded and counts as one-fourth of their total test score. To facilitate instructor intervention to improve group functioning, students are assigned a role of Manager, Checker/Recorder, or Skeptic and this role rotates with each session.

Context-rich problems include a motivation and a realistic context. The problems avoid physics cues and do not explicitly define physics quantities. They often include extraneous information or require reasonable assumptions to be made by the solver. For group work, problems are designed to be difficult for an individual to solve during the session but very manageable for a group.

In the discussion session, groups solve a single context-rich problem while a teaching assistant provides coaching as necessary. Near the end of a session, a representative from each group writes a part of their group solution on the board and the teaching assistant facilitates a short whole-class discussion. Students receive a complete solution to the problem when they leave the room. For the laboratory, students solve the problems (usually 2 per lab session) individually before coming to lab. They then work in their groups to reconcile their solutions and check their results from nature in laboratory activities. All students in a given discussion section are also in the same lab section and attend the same lecture.

New teaching assistants have an eight day orientation for teaching with Cooperative Problem Solving during the two weeks prior to the fall term [7]. They receive ongoing support in a weekly seminar during the semester from experienced graduate students designated as mentor TAs. These mentor TAs also visit their classes and offer suggestions to improve teaching. Experienced TAs have a half day orientation before the beginning of the fall semester and can request coaching from the mentor TAs. All TAs meet with the lecturer once per week.

DATA COLLECTION AND ANALYSIS

The Force Concept Inventory (FCI) is routinely given to introductory physics students in the first and last week of the term. It does not count toward the course grade. From 1993-1996 the original version of the FCI was administered and from 1997-2007 the revised FCI was used. Only the revised FCI scores are included in this analysis. To avoid possible differences between on and off-sequence student populations, only the fall term classes are included.

To avoid possible biases, the classes included in this analysis have pre-post matched data for more than half of the class. Of the 50 fall classes from 1997-2007, 40 meet this criteria. Gender is self-reported; however, very few (N=14) students have been excluded because of missing gender information. A few scores (N=60) are not included because more than three questions were left blank on the pre-test or post-test.

The final sample includes 5,636 students (1,261 females and 4,375 males) from 40 classes with 22 different instructors. On average, females make up 22% of these classes. The average dropout rate is 7% and failure rate (Ds and Fs) is 4%. The distribution of post-test FCI scores for males exhibits a clear ceiling effect, and for this reason non-parametric statistics are reported.

RESULTS

Female students in the calculus-based physics course have an average pre-test score on the FCI that is substantially lower than their male counterparts. Both groups exhibit similar absolute learning gains, preserving the gender gap after instruction.

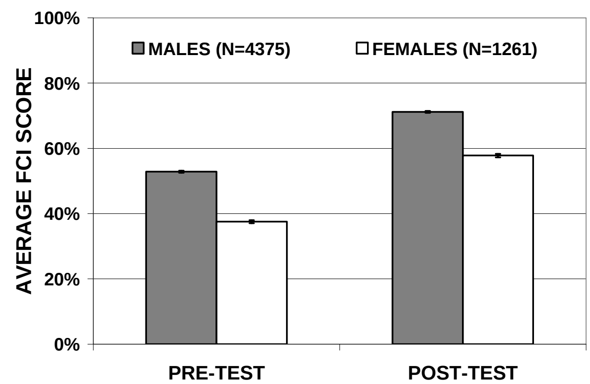

FIGURE 1. FCI pre-test and post-test scores by gender. Error bars represent the standard error of the mean.

As shown in Figure 1, the pre-test gender difference in mean FCI scores is 15.3±0.5% and the post-test gender difference decreases only slightly to 13.4±0.6%. Both results are statistically significant ($p<0.0001$ from a nonparametric test). Both groups show similar learning gains during the term (about six questions on the FCI).

Figure 2 shows a trend of increasing pre-test FCI scores for both genders from 1997-2007 but relatively stable post-test scores.

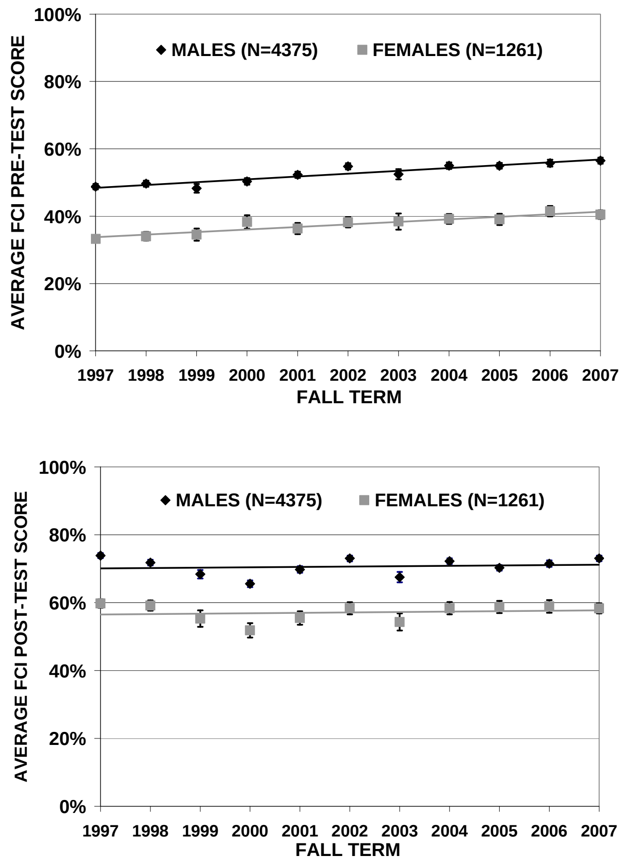

FIGURE 2. a) FCI pre-test scores and b) FCI post-test scores as a function of time.

The change in the gender gap for a class ranges from –8%±3% (decreased gap) to +7%±5% (increased gap) with an average reduction of -1.9±0.6%. As seen in Figure 3, these differences by class and instructor are consistent with statistical fluctuations. Even though different instructors implement cooperative group problem solving differently, no significant instructor dependent gender gap change was observed.

The gender gap decreases with increasing post-test score probably due to a ceiling effect for the males. Harvard observed a reduction or elimination of the gender gap with interactive engagement methods [2]. This is consistent with our data for their higher post-test (and pre-test) scores. The University of Colorado observed no significant reduction of the gender gap with interactive engagement methods [3,4]. This is also consistent with our data for their lower post-test (and pre-test) scores.

There is a significant correlation between FCI pre-test and post-test that is somewhat higher for males than females (Spearman's $\rho = 0.740$ for males and $\rho = 0.630$ for females). The relationship between FCI pre-test and course grades is also stronger for males ($\rho = 0.454$ for males and $\rho = 0.330$ for females).

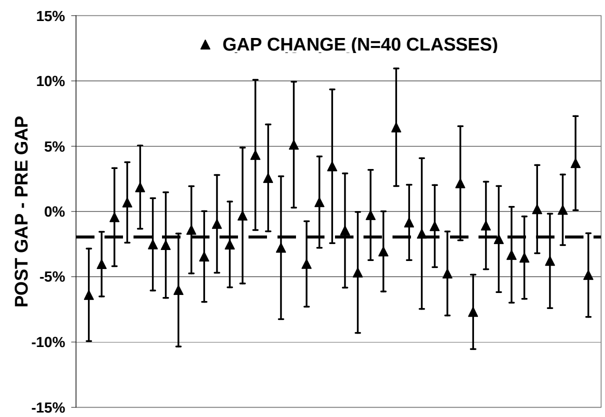

FIGURE 3. The FCI gender gap change for different instructors; these variations are consistent with statistical fluctuations. The dashed line is the average gender gap change.

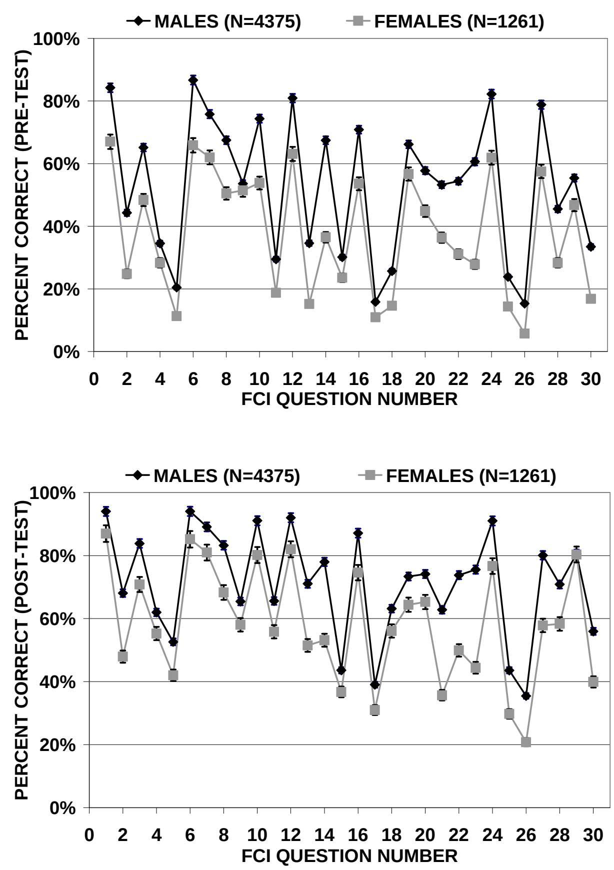

FIGURE 4. Performance by gender on each FCI question. On average, males outperform females on every question of the FCI exam for both a) pre-test and b) post-test.

As shown in Figure 4, males outperform females on every question of the FCI. The largest gender gap

(>30%) is observed for FCI items 14 (airliner dropping a bowling ball) and number 23 (rocket). This is true for both the pre-test and post-test and agrees with previous results [8].

Despite substantial gender differences on the FCI, there is little difference in course performance as measured by grades. We observe that males' grades average 1.5±0.2% higher than females (significance $p<0.0001$ from a nonparametric test). This is consistent with other results [9] that indicate males perform one-sixth of a letter grade higher than females in university physics, a difference too small to be educationally significant. Because a grade includes laboratory reports and other participation, achievement might be offset by diligence. A more direct measure of achievement is the final exam score. The difference in final exam scores is somewhat higher (3.9±0.6%) but is still small. Figure 5 shows the final exam distribution by gender.

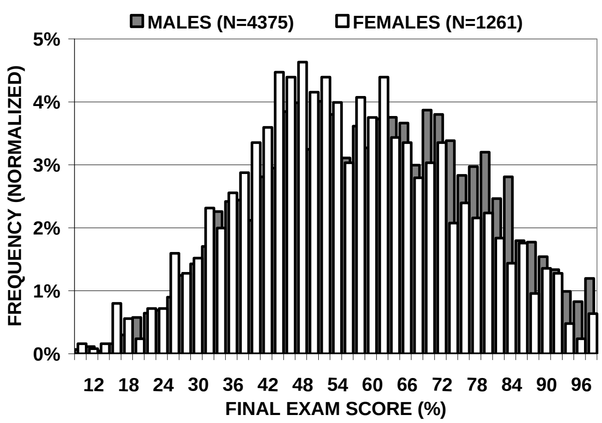

FIGURE 5. Normalized final exam distribution by gender.

SUMMARY

For the calculus-based course for scientists and engineers at the University of Minnesota taught using cooperative problem solving, our data show a significant gender gap in pre-test Force Concept Inventory scores that persists post-instruction although there is essentially no gender difference in course performance as determined by course grade. We also observe a trend that pre-test scores over the past decade have been increasing whereas post-test scores resulting from our instruction remain stable. The change in the gender gap by class or instructor is consistent with statistical fluctuations.

Future work will examine the role of background factors such as high school physics and math experience as well as mathematical skills on FCI exam scores and course performance.

ACKNOWLEDGMENTS

We are grateful to the many members of the University of Minnesota PER group of the past and present who administered and analyzed conceptual surveys and background data for introductory physics courses. This work is supported in part by NSF DUE-0715615.

REFERENCES

1. J.B. Kahle and J. Meece, "Research on gender issues in the classroom," in *Handbook of Research on Science Teaching and Learning,* edited by D.L. Gabel, New York: Macmillan, 1994, pp. 542-557.
2. M. Lorenzo et al. *Am. J. Phys.* **74**(2), 118-122 (2006).
3. S.J. Pollock et al. *Phys. Rev. ST – PER* **3**, 010107, 1-4 (2007).
4. L.E. Kost, S.J. Pollock, and N.D. Finkelstein, "Investigating the source of the gender gap in introductory physics" in *2007 Physics Education Research Conference*, edited by L. Hsu et al., AIP Conference Proceedings 951, American Institute of Physics, Melville, NY, 2007, pp. 136-139.
5. P. Heller et al. *Am. J. Phys.* **60**(7), 627-636 (1992).
6. P. Heller and M. Hollabaugh, *Am. J. Phys.* **60**(7), 637-644 (1992).
7. F. Lawrenz, P. Heller, R. Keith, and K. Heller, *J. Coll. Sci. Teach.* **22**(2)**,** 106-109 (1992).
8. L. McCullough and D. E. Meltzer, "Differences in male/female response patterns on alternative-form versions of FCI items," in *Proceedings of the 2001 Physics Education Research Conference*, edited by S. Franklin et al., pp. 103-106.
9. R.H. Tai and P.M. Sadler, *Int. J. Sci. Educ*. **23**(10), 1017-1037 (2001).