Peer instruction: From Harvard to the two-year college

Nathaniel Lasry
Physics Department, John Abbott College, Montreal, Qc H9X 3L9, Canada and School of Engineering and Applied Sciences, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138

Eric Mazur
School of Engineering and Applied Sciences, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138 and Department of Physics, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138

Jessica Watkins
School of Engineering and Applied Sciences, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138

We compare the effectiveness of a first implementation of peer instruction (PI) in a two-year college with the first PI implementation at a top-tier four-year research institution. We show how effective PI is for students with less background knowledge and what the impact of PI methodology is on student attrition in the course. Results concerning the effectiveness of PI in the college setting replicate earlier findings: PI-taught students demonstrate better conceptual learning and similar problem-solving abilities than traditionally taught students. However, not previously reported are the following two findings: First, although students with more background knowledge benefit most from either type of instruction, PI students with less background knowledge gain as much as students with more background knowledge in traditional instruction. Second, PI methodology is found to decrease student attrition in introductory physics courses at both four-year and two-year institutions. © *2008 American Association of Physics Teachers.*
[DOI: 10.1119/1.2978182]

I. INTRODUCTION

Peer instruction (PI) is a student-centered approach to teaching that has demonstrated effectiveness in university settings.[1–4] In PI, the progression of any given class depends on the outcome of real-time student feedback to ConcepTests—multiple-choice conceptual questions. Figure 1 shows the process of PI.

The number of instructors using PI increases every year as word spreads about the effectiveness of the method.[1–3] Research at Harvard University shows significant increases in students' conceptual understanding in comparison to traditional lecturing method.[1] However, there are few studies concerning the efficacy of PI within other student populations, particularly students in two-year colleges. A study of PI users who collected quantitative data on their students' learning found that most users report high conceptual gains in introductory physics courses;[2,3] however, no data were reported on two-year college courses.

We present student conceptual understanding and traditional problem-solving data from the first implementation of PI in an algebra-based introductory physics course at John Abbott College, a public two-year college in Montreal, Canada and compare with similar data from the first course in which PI was implemented at Harvard University. To determine the impact PI has on students with different prior knowledge, we also present data that controls for students' background physics knowledge and compares the resulting learning gains in traditional and PI courses at John Abbott College. Finally, we look at the effects of PI on student attrition to determine whether PI has an effect on student attrition in college courses at both the university and two-year college levels.

II. STUDY DESIGN

To determine the effectiveness of the method in a two-year college setting, we contrast results from the first PI course at Harvard University (Fall 1991)[1] with the first PI course taught at John Abbott College (Fall 2005). We examine only the first implementation in both institutions, as the effectiveness of any reformed instruction progresses as one gains familiarity with the approach and tailors materials to the specific student population. At Harvard, we compared results from traditional courses with the first PI course taught by one of the authors (EM).[1,4] At John Abbott College, we obtained results from both traditional and PI courses during the same semester. Students were randomly assigned to one of three sections: two PI sections ($N=83$) and one control section ($N=44$). The PI sections were taught by one of the authors (NL) and the traditional section was taught by another instructor, with similar characteristics, gender (male), age (± 3 yrs), teaching experience (± 1 yr). The PI sections used 3–4 ConcepTests per 75-min class, while the traditional section relied solely on lecturing. All sections followed the same course outline, were taught at similar paces, and had the same laboratory experiments.

III. METHODS AND RESULTS

A. PI in the college classroom: Conceptual learning

We measured students' conceptual understanding of Newtonian mechanics using the force concept inventory[5,6] (FCI) in both PI and traditional courses at John Abbott College and Harvard University. To measure how well students performed after instruction relative to their performance before instruction, we calculate the Hake normalized gains[7] for each
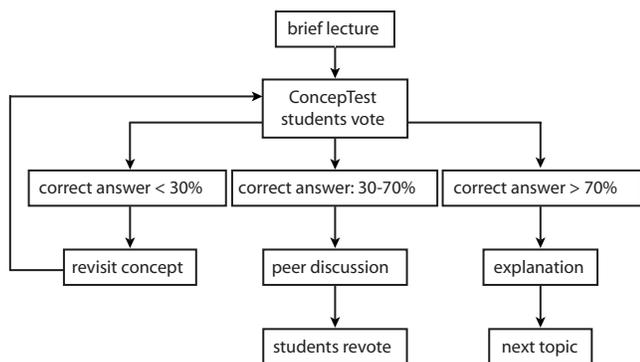
Fig. 1. The peer instruction implementation process. Note that the recommended percentages of correct answers for each step are estimates—how an instructor decides to proceed is highly dependent on the content and student population.

student. These sets of individual normalized gains for PI and control sections are then compared for statistical significance. However, the non-PI 1990 course at Harvard was not given the FCI before instruction. Inferring the preinstruction FCI class average from the average determined from the following three years at Harvard [1991: FCIpre=20.44 (70.9%); 1993: FCIpre=20.24 (69.8%); 1994: FCIpre =20.41 (70.4%)], we calculate the Hake normalized gains using prepost class average FCI scores. However, given we do not have individual student gains to compare; we cannot obtain statistical significance for differences between the non-PI 1990 and PI 1991 Harvard groups.

Table I shows the FCI scores before instruction (pretest) and after instruction (posttest), as well as the normalized gains for students in traditional and PI courses. Differences between the traditional and PI courses are considered significant for p values less than 0.05. At John Abbott College, no significant difference exists between the PI and control sections before instruction; however, the PI section achieves significantly greater normalized gains after instruction ($p <0.01$). In comparison, the pretest FCI scores at Harvard University ($>70\%$) are very different from those of the two-

Table I. Precourse (S_{pre}) and postcourse (S_{post}) force concept inventory results for PI and traditional courses. The FCI pretest was administered on the first day of class. In the traditional course at Harvard no pretest was given, so the average of the following years' pretest scores is listed; these years had a standard deviation of 0.55 percentage points. Note that the normalized gains at John Abbott are calculated by taking the average of the individualized student gains, while at Harvard, class averages are used, as no individual data exists for the pretest in the traditional course.

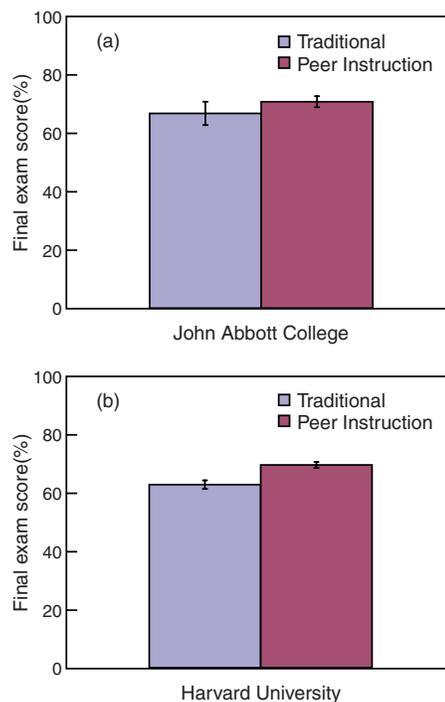|	N	S_{pre}	S_{post}	g
John Abbott College				
PI	69	42.6	68.6	0.50
Traditional	22	46.0	63.3	0.33
Difference		−3.4	5.3	0.17[a]
Harvard University				
PI	187	70.9	85.0	0.49
Traditional	124	(70.4)	77.4	0.23
Difference		—	7.5[b]	0.26

[a]$p<0.01$.
[b]$p<0.001$.

Fig. 2. Average final examination score for traditional and PI sections at John Abbott College, where students took the same exam and had the same graders. On the left, the average scores of common final examination problems from 1991 PI course and 1985 traditional introductory physics course at Harvard University. Note that the final examinations at John Abbott College and Harvard University are not the same.

year college students ($<50\%$), however, the student normalized gains in this college PI population ($g=0.50$) are nearly identical to the normalized class gain found at Harvard during the first implementation of PI ($g=0.49$).

B. PI in the college classroom: Traditional problem solving

We assess traditional problem-solving skills using a comprehensive final examination. At Harvard, PI students in the fall of 1991 were given the same final examination as a previous traditionally instructed cohort (fall 1985), and consisted of all algorithmic, nonconceptual questions.[4] At John Abbott, a committee of physics instructors (none of which were involved in this study) constructed the final examination for all introductory physics courses, including the PI and traditional sections. This exam was comprised of 90% algorithmic, nonconceptual questions. Each instructor marked a single exam question for the entire cohort. Therefore, the PI and control group at John Abbott took the same exam and had the same grader on every question. Furthermore, the graders of the exam questions were unaware of which students were in the traditional section and which were in PI sections.

Figure 2 compares the average grades between the control section and PI section on the common final examinations administered at John Abbott College. The average grades on a Harvard common final exam in a traditional course and PI course are also shown. The error bars indicate 95% confidence intervals around the average score. When the error bars overlap, the difference in the means is not statistically significant.

Table II. Individual average normalized gains on force concept inventory, grouped by instruction and pretest score. The fourth column shows the difference between high pretest and low pretest score groups within a given instructional strategy, the fourth row shows the difference between PI and traditional sections within low pretest and high pretest groups.

|	N	$S_{pre} < 40\%$	$S_{pre} \geq 40\%$	Difference
PI	69	0.39	0.67	0.28[b]
Traditional	22	0.26	0.38	0.12
Difference		0.12[a]	0.29[b]	

[a]$p < 0.1$.
[b]$p < 0.001$.

Fig. 3. On the left, attrition rates for traditional and PI sections at John Abbott College. On the right, attrition rates for 1991–1994 PI courses and 1990 traditional course at Harvard University.

At John Abbott College, the students in the PI courses ($N = 79$) obtained an average score of 68%, while the control group ($N = 35$) obtained an average score of 63%. Although the PI students scored higher than the traditionally taught students, the difference in these scores is not statistically significant. In comparison, the Harvard PI students also outperformed their non-PI peers on traditional exam problems. PI students at Harvard ($N = 222$) obtained an average score of 69%, while their predecessors taught with traditional lecturing ($N = 144$) scored 63% on average. The difference between the PI and traditional students is statistically significant ($p < 0.001$).

C. Effectiveness of PI: The role of background knowledge

To determine whether the effectiveness of PI depends on students' incoming knowledge, we analyze the FCI data at John Abbott College, separating students with different background knowledge. We first calculate the median preinstruction FCI score for students at John Abbott College (40%). We then separate both the PI and control sections into two groups: students who scored above the median on the pretest ($S_{pre} \geq 40\%$, "high background knowledge") and students who scored below the median ($S_{pre} < 40\%$, "low background knowledge"). Table II compares the normalized gains for these two groups.

In both the PI and control sections at John Abbott College, there is a difference in normalized gain between the low background knowledge and high background knowledge groups. The difference between the groups in the PI section is statistically significant ($p < 0.001$), while the difference in the control section is much smaller and not significant.

Both PI groups outperform their control group counterparts. PI students with high background knowledge achieve significantly more conceptual learning than high background knowledge students in the control section ($p < 0.001$). Low background knowledge PI students also obtain better conceptual gains ($g = 0.39$) than the low background knowledge students in the control section ($g = 0.26$), although the statistical significance of this difference is marginal ($p = 0.07$). Notably, low background knowledge PI students achieve learning gains which are virtually identical to those of higher incoming knowledge students in the control group.

D. Effectiveness of PI: Student retention

To determine whether the use of PI has a quantifiable effect on student retention in a course, we compare the number of students who completed PI and non-PI courses at both
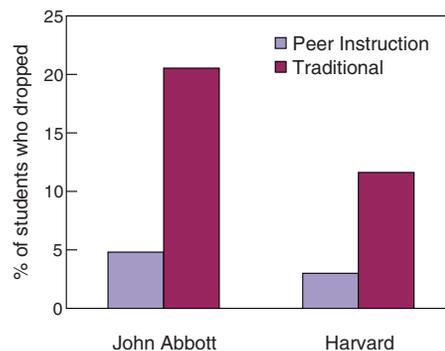
John Abbott College and Harvard. We determine completion of course by comparing how many students enrolled versus how many students completed the final examination.

Figure 3 shows the results at both John Abbott College and Harvard University. Of the initial number of students registered in the PI course at John Abbott ($N = 83$), less than 5% did not take the final exam ($N = 4$). Comparatively, 20.5% of the control section did not take the final examination (9 out of 44). The class size at Harvard is much larger, providing more data for comparison. Of the 224 students originally registered in the 1990 traditional mechanics course, 11.6% ($N = 26$) did not take the final examination. The following year, of the coincidentally identical number of students ($N = 224$) following the PI approach, no students were absent from the final examination. Over the next few years at Harvard, drop rates in PI sections were found to be consistently below 5% (1993: 4.9%; 1994: 3.5 %; 1995: 3.4%).

IV. DISCUSSION

A. Effectiveness of PI in the college classroom

As Table I shows, PI in the two-year college classroom is more effective at developing students' conceptual understanding than traditional lecture-based instruction. Although the control section at John Abbott achieved high-range normalized gain values on the FCI ($g = 0.33$) for traditional instruction, the PI normalized gains were significantly larger ($g = 0.50$, $p < 0.01$). Gains at the two-year college are similar to those found at Harvard during the first PI implementation, indicating that the benefits of PI on students' conceptual learning extend from the university to the two-year college. The difference between PI and traditional instruction found also replicates Hake's result[7] concerning the difference between interactive engagement methods and traditional instruction.

Although more time in PI class is devoted to conceptual learning and less time to traditional problem solving, PI students at a two-year college obtain similar grades in traditional problem solving. This result is once again comparable to what occurred at Harvard in 1991 when PI students obtained better grades on their final examination than their peers in 1985 after traditional instruction. These results confirm that providing a solid conceptual background allows students to be more effective in problem solving.[4] Therefore,

we can conclude that time spent on basic concepts does not reduce students' traditional problem solving ability even in two-year college populations.

B. Dependence of PI effectiveness on background knowledge

At John Abbott College, we find that higher background knowledge students display larger gains on the FCI, regardless of instruction. Previous research has shown that students with lower scientific reasoning ability or cognitive skills are more likely to be influenced by prior knowledge, and their learning of physics concepts will be correlated to their incoming knowledge.[8,9] At the community college level, research has also shown a relationship between scientific reasoning and FCI gains, in addition to a positive correlation between pre-instruction FCI scores and gains.[10] Our results confirm a similar relationship between prior knowledge and gains on the FCI in two-year colleges.

Although the normalized gain at John Abbott College depends on before-instruction FCI scores, it would be false to infer that PI should be used only with higher background knowledge students. Our results show that PI increases learning for all students regardless of background knowledge. Lower background knowledge students taught with PI achieve greater conceptual gains than their counterparts in the control section (0.39 vs 0.26; $p=0.07$). That there is no firm statistical difference ($p<0.05$) between the lower background PI group and lower background control group is most probably due to the small number of students ($N=40$ and $N=9$, respectively). Furthermore, lower background knowledge students achieve learning gains that match the gains of higher background knowledge students in the control section. Thus, PI is more effective than traditional instruction with both lower and higher background knowledge two-year college students.

We find no correlation between pre-instruction FCI scores and gains at Harvard University ($r=-0.04$) but the data at John Abbott College show a dependence of gain on pre-instruction FCI score ($r=0.60$), suggesting that this dependence is not universal. One possible explanation for the lack of dependence on pretest scores for Harvard students is that their high reasoning abilities compensates for less background knowledge.[8]

C. Peer instruction decreases student attrition

Previous studies suggest that active and cooperative learning methods are useful in reducing student attrition in science programs. Tobias cites the competitive, rather than cooperative, atmosphere in introductory science courses and the focus on skill performance instead of conceptual understanding as possible reasons for attrition of science majors.[11] Longitudinal research shows that courses using active or cooperative learning methods can reduce attrition in science and engineering programs.[12,13] Peer instruction is also helpful in reducing failure rates by decreasing the number of students who drop-out of an introductory physics course. We found large differences (by factor of 2–3) in the fraction of students dropping out between PI and non-PI courses both at John Abbott College and at Harvard University. This result suggests that engaging students through ConcepTests during a class and throughout a semester encourages students to stay in the course.

V. CONCLUSION

Peer instruction is an effective instructional approach not only at a top-tier university, but also at a two-year college. In both settings, PI increases conceptual learning and traditional problem solving skills. Although the effectiveness of PI depends on student background-knowledge at the two-year college, both higher and lower background-knowledge students benefit from the approach when compared to traditional lecture-based instruction. Finally, by engaging students in the course, PI reduces the number of students who drop the course.

[1] C. H. Crouch and E. Mazur, "Peer Instruction: Ten years of experience and results," Am. J. Phys. **69**, 970–977 (2001).

[2] A. Fagen, *Assessing and Enhancing the Introductory Science Course in Physics and Biology: Peer Instruction, Classroom Demonstrations, and Genetics Vocabulary* (Harvard U. P., Cambridge, MA, 2003), p. 186.

[3] A. Fagen, C. H. Crouch, and E. Mazur, "Peer Instruction: Results from a range of classrooms," Phys. Teach. **40**, 206–209 (2002).

[4] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).

[5] I. A. Halloun *et al.*, in *Peer Instruction: A User's Manual*, edited by E. Mazur (Prentice Hall, Upper Saddle River, NJ, 1995).

[6] D. Hestenes, M. Wells, and G. Swackhammer, "Force concept inventory," Phys. Teach. **30**, 141–151 (1992).

[7] 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**, 64–74 (1998).

[8] V. P. Coletta and J. A. Phillips, "Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability," Am. J. Phys. **73**, 1172–1182 (2005).

[9] V. P. Coletta, J. A. Phillips, and J. Steinert, "Interpreting force concept inventory scores: Normalized gain and SAT scores," Phys. Rev. ST Phys. Educ. Res. **3**, 010106-1–5 (2007).

[10] K. Diff and N. Tache, "From FCI to CSEM to Lawson's Test: A report on data collected at a community college," in *2007 Physics Education Research Conference*. Greensboro, NC.

[11] S. Tobias, *They're Not Dumb, They're Different: Stalking the Second Tier* (Research Corporation, Tucson, AZ, 1990).

[12] E. Etkina and A. Van Heuvelen, "Author response," Phys. Teach. **41**, 68 (2003).

[13] R. Felder, G. Felder, and E. J. Dietz, "A longitudinal study of engineering student performance and retention. V. Comparisons with traditionally-taught students," Bell J. Econom. **87**, 469–480 (1998).