

Reducing the gender gap in the physics classroom

Mercedes Lorenzo

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

and IES Universidad Laboral, Avenida de La Mancha, s/n, 02080 Albacete, Spain

Catherine H. Crouch^{a)}

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

and Department of Physics, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081

Eric Mazur

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

(Received 26 July 2005; accepted 2 December 2005)

We investigate if the gender gap in conceptual understanding in an introductory university physics course can be reduced by using interactive engagement methods that promote in-class interaction, reduce competition, foster collaboration, and emphasize conceptual understanding. To this end we analyzed data from the introductory calculus-based physics course for non-majors at Harvard University taught traditionally or using different degrees of interactive engagement. Our results show that teaching with certain interactive strategies not only yields significantly increased understanding for both males and females, but also reduces the gender gap. In the most interactively taught courses, the pre-instruction gender gap was gone by the end of the semester. © 2006 American Association of Physics Teachers.

[DOI: 10.1119/1.2162549]

I. INTRODUCTION

In the United States, females are underrepresented in science and technology-related courses and careers^{1,2} and their average scores are lower than males on science tests at the secondary and post-secondary level.^{3,4} This gender gap is also reported at the secondary level in 28 countries in Europe, North America, Asia, Oceania, and the Middle East.^{3,5-7} Although the gender gap has been closing in most scientific and technological fields, the largest gender disparity in both achievement and professional representation remains in physics.¹⁻³

Extensive research on gender differences in science achievement has been carried out since the 1980s. The results of this research suggest that the following teaching strategies help narrow the gender gap:

- (1) Integration of everyday experiences and interests that are relevant to both genders into the content and context of instruction.⁷⁻⁹
- (2) Assessment and use of students' prior knowledge to construct new knowledge.⁷
- (3) Interactive environments that enhance cooperation and communication in the classroom among the students and between the students and the instructor.^{7,10}
- (4) Alternation between group discussion and structured teaching. Females perform better when they are able to articulate their thoughts verbally and males perform better when their learning experience is structured.¹¹
- (5) Activities that decrease competitiveness.^{5,7,12,13}
- (6) Diverse and frequent assessment practices and feedback.^{2,5,7}
- (7) Activities that foster students' understanding.^{5,14-16}
- (8) Application of physics to a broader world-view.¹⁶

Some studies suggest that females benefit especially by the

use of active pedagogies.^{12,13} Although active engagement benefits both genders, females tend to learn more when they express ideas in words through discussion, whereas males prefer working independently.¹⁷

The interactive engagement (IE) methods used in this study (Peer Instruction,^{18,19} the *Tutorials in Introductory Physics*,²⁰ and cooperative quantitative problem-solving activities) make use of many of the strategies listed above [particularly strategies (3)–(7)]. We therefore investigated if these teaching methods narrow the gender gap in the large calculus-based introductory physics course for non-majors at Harvard University.

II. METHODOLOGY

The introductory calculus-based physics course for non-majors was taught at Harvard from 1990 to 1997 (no data were collected in 1992). This course, which covers Newtonian mechanics, meets twice a week in a large lecture hall for a total of 3 h of instruction and once a week for 1 to 2 h in smaller sections of 15–20 students directed by teaching assistants. During the seven years of the study, the average enrollment in the course was 202; the student gender ratio in the course averaged 1.7 (male/female).

A. Teaching approaches

The teaching approach evolved from traditional lectures to a highly interactive teaching style. In 1990 both the lectures and the sections were taught using a traditional lecture format. In 1991 traditional lectures were replaced by Peer Instruction (PI). Peer Instruction modifies the traditional lecture by alternating between short (10–15 min) mini-lectures and conceptual questions discussed by students in small groups to directly address conceptual difficulties during class

Table I. Classification of the courses studied according to the teaching approach used in the lectures and in the sections; T (traditional), IE1 (partially interactive), and IE2 (fully interactive).

Group	Year	Teaching approach	
		Lectures	Sections
T	1990	Traditional	Traditional
IE1	1991–1995 (except 1992)	Peer Instruction	Traditional
IE2	1996–1997	Peer Instruction	<i>Tutorials</i> ^a and cooperative problem solving activities

^aReference 20.

time. Students are expected to read before class about the day's topic in the textbook, so that they will be better able to participate in these discussions.

Over the following five years, we improved the implementation of PI. In 1993 and 1994, the set of conceptual questions and the in-class questioning/discussion strategy were refined. In 1995 we began using a research-based mechanics text being developed by Mazur.²¹ The text introduces concepts before developing the mathematical framework and conceptual obstacles are directly addressed in a real-world context that matches the interests of males and females. Until 1996 we ensured that students read before class by giving multiple-choice reading quizzes at the beginning of class; in 1996 and 1997, students completed written assignments about the reading before class. A more detailed description of Peer Instruction and its past and current implementations is in Ref. 19.

Although the lectures from 1990 to 1995 actively involved students once Peer Instruction was implemented, the weekly 1.5-h section meetings consisted primarily of traditional presentations by a teaching assistant to students seated in rows. In 1996 these traditional sections were replaced with weekly 2-h workshops during which students worked in groups of three or four seated around tables. During the first hour of the workshop, students completed the *Tutorials*,²⁰ which emphasize conceptual reasoning and hands-on activities rather than standard quantitative problem solving and provide a structured opportunity for students to discuss concepts with one another. Under the supervision of teaching assistants, the *Tutorials* help the students construct conceptual models and apply them to real-world situations. The second half of the workshop is devoted to reinforcing the students' problem-solving skills through cooperative quantitative problem-solving activities similar to those developed by Heller *et al.*²² Table I summarizes the teaching approaches used and classifies the courses studied into three groups: traditional (T), partially interactive (IE1), and fully interactive (IE2).

Five instructors taught these courses; one of the instructors used all three approaches in different years, and others used only one or two of the three approaches. We find consistent results for each approach regardless of the instructor. Consequently, we conclude that the observed changes in students' performance are not caused by changes in the instructor, but rather due to the variation in teaching strategy.

B. Assessment methods and data collection

To investigate how interactive teaching methods affect the gender gap in conceptual understanding, we used the Force

Concept Inventory (FCI).²³ This test is widely used to assess students' understanding of Newtonian mechanics in introductory physics courses and to evaluate the effect of changes in instructional practice.²⁴ Students were given 45 min to complete the test, once at the beginning of the semester (pretest) and again three months later after instruction in mechanics was completed (posttest). No pretest data are available for the traditionally taught course.

Between 1990 and 1994 we used the original version of the test consisting of 29 questions. In 1995, the test was revised to eliminate ambiguities. Starting in 1995, we used the revised version, which contains 30 questions. The two versions of the test have 27 questions in common, although some of these questions were slightly reworded. Because the assessment instrument changed slightly during this study, we analyzed the FCI scores by scoring all questions on each version of the test and by only scoring the overlapping 27 questions. The two analyses yield similar results, indicating that the revision of the FCI in 1995 did not affect the results. The statistical analysis presented in this paper is based on the complete set of questions for both test versions.

In addition to the FCI, we also examined scores on the Mechanics Baseline Test²⁵ (MBT) which we administered just before the final exam. These scores are tabulated in the online supplementary materials.²⁶ The average gender gap in the MBT scores is less than or similar to the gap found on the FCI posttest for all groups surveyed. Although there is a small post-instruction gender gap for all three groups, the gap in the MBT scores is smallest for the IE2 group (4.8%). In addition, the IE2 group shows the highest MBT scores of the three groups.

We analyzed data only from students who completed the FCI pretest (for IE courses), the FCI posttest, and the MBT, for a total of 1048 students. We focused our analysis on the FCI rather than the MBT, because the FCI directly probes the effects of instruction, was given as a pretest and a posttest, while the MBT was given only as a posttest. Because the grading scheme for the course and the nature of the final examinations changed from year to year, the data from examinations and final grades are not comparable from year to year and were therefore omitted from this study.

III. RESULTS AND DISCUSSION

Table II provides the average FCI pretest score S_i and posttest score S_f for males and females, with the corresponding gender gap ($S^M - S^F$) and p values. The average pretest scores for both males and females in the IE courses remained approximately constant over the duration of the study; the year-to-year variations are not statistically significant. Therefore, the differences between groups observed on the posttest scores can be attributed to differences in instruction rather than different student backgrounds. In addition, the pretest gender gap, $S_i^M - S_i^F$, is statistically significant for each course.

Table II shows that the scores for both genders increase after instruction. In addition, the posttest scores for both genders increase as the level of interactive engagement increases from T to IE1 and then IE2. Both male and female students in the T course score lower than the students in the IE groups on the posttest. According to a study by Henderson *et al.*,²⁷ administering the pretest does not bias posttest results and so the low posttest scores for the T course cannot be attributed to the absence of a pretest.

Table II. Force Concept Inventory data for the introductory calculus-based physics course at Harvard University; N^M (N^F) is the number of males (females); S_i^M (S_i^F) equals the average pretest score for males (females); and S_f^M (S_f^F) equals the average posttest score for males (females). The numbers in parentheses represent the standard deviation in the reported values. $S_i^M - S_i^F$ ($S_f^M - S_f^F$) equals the gender gap before (after) instruction. All data reported to two significant figures. Values marked with an asterisk are not statistically significant.

Group	Year	N^M	N^F	FCI pretest score (%)				FCI posttest score (%)			
				S_i^M	S_i^F	$S_i^M - S_i^F$	p value	S_f^M	S_f^F	$S_f^M - S_f^F$	p value
T	1990	61	44	82 (13)	71 (16)	10	0.0004
IE1	1991	105	61	74 (15)	62 (16)	12	<0.0001	86 (8.6)	78 (11)	7.9	<0.0001
	1993	91	52	72 (14)	61 (14)	11	<0.0001	88 (7.0)	80 (11)	8.2	<0.0001
	1994	121	77	75 (15)	60 (16)	15	<0.0001	89 (8.1)	81 (12)	7.6	<0.0001
	1995	115	61	72 (18)	60 (17)	13	<0.0001	90 (9.4)	83 (14)	7.4	<0.0001
IE2	1996	94	52	71 (19)	61 (19)	9.8	0.0039	90 (11)	87 (10)	3.3*	0.0828
	1997	67	47	71 (19)	62 (20)	8.5	0.0205	92 (11)	91 (8.3)	1.5*	0.4290

Figure 1 shows how the level of interactive engagement affects the gender gap. The average gender gap for each of the three groups (T, IE1, and IE2) is shown for both the pretest and posttest. No pretest data are available for the T course. However, the pretest scores stay essentially constant for each gender over the following six years and the difference in pretest gender gap between the IE1 group (13%) and the IE2 group (9.2%) is only marginally significant (p value=0.047). We therefore compare the T posttest gender gap to the average of the IE courses' pretest gender gaps (11%).

A comparison of the pretest and posttest gender gaps within each group shows that increasing the degree of interactivity significantly reduces the gender gap from pretest to posttest. The posttest gender gap for the T course (10%) is almost the same as the average of the IE courses' pretest gender gaps. In contrast, the introduction of Peer Instruction in the lectures (IE1) cut the posttest gender gap to 7.8%, less than two-thirds of the pretest gender gap. Full interactivity (IE2) reduced the posttest gender gap even further, to 2.4%—only a quarter of the pretest gap. Most important, after instruction the gender gap for the IE2 group is no longer statistically significant, indicating that the males and females in this group perform similarly after instruction.

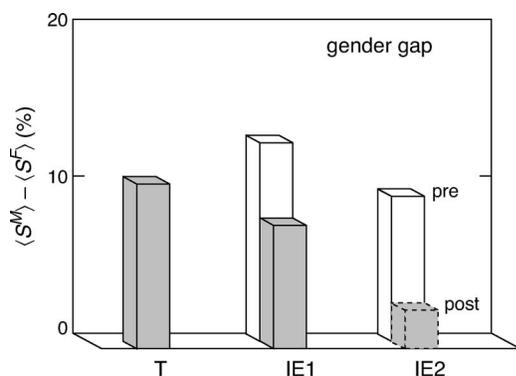


Fig. 1. Effect of instructional approach on the gender gap in student performance on the Force Concept Inventory test for introductory calculus-based physics course at Harvard University; Symbols: M (male), F (female); pre (before instruction), post (after instruction); T (traditional), IE1 (partially interactive), and IE2 (fully interactive teaching approach). The posttest gender gap in the IE2 group is not statistically significant.

These results show that the higher the level of interactive engagement in the course, the smaller the gender gap is after instruction.

To further evaluate the impact of each of the three teaching approaches on the students' performance, we also calculated the class average normalized gain $\langle g \rangle = (\langle S_f \rangle - \langle S_i \rangle) / (100 - \langle S_i \rangle)$ from the FCI scores. The normalized gain accounts for differences in the pretest score by determining what fraction of the total possible gain from pretest to posttest is achieved. The average of pretest scores for the IE courses was used as the T course pretest. Figure 2 shows that $\langle g \rangle$ increases for both genders as the level of interactive engagement in the course is increased. For the IE1 and IE2 groups, the difference in the normalized gain between males and females is not statistically significant; in the IE2 group, both genders achieve the same high normalized gain (0.70 for males and 0.71 for females). The data in Fig. 2 are tabulated in the online supplement along with the p values.²⁶ We were unable to determine if the gender gap for the T group is statistically significant because the T group did not take the pretest.

According to Hake's normalized gain classification,²⁴ both male and female students in the IE2 group fall into the high-gain category, $\langle g \rangle \geq 0.7$. In contrast, females in the T course

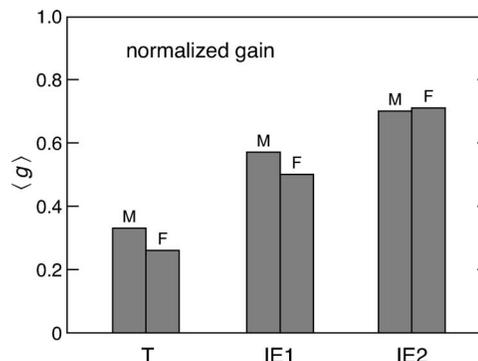


Fig. 2. Effect of the instructional approach on the normalized pre- to post-test gain $\langle g \rangle$ on the Force Concept Inventory test for introductory calculus-based physics course at Harvard University. Because no pretest was administered in the T course, the average pretest from all IE courses is used as the T pretest.

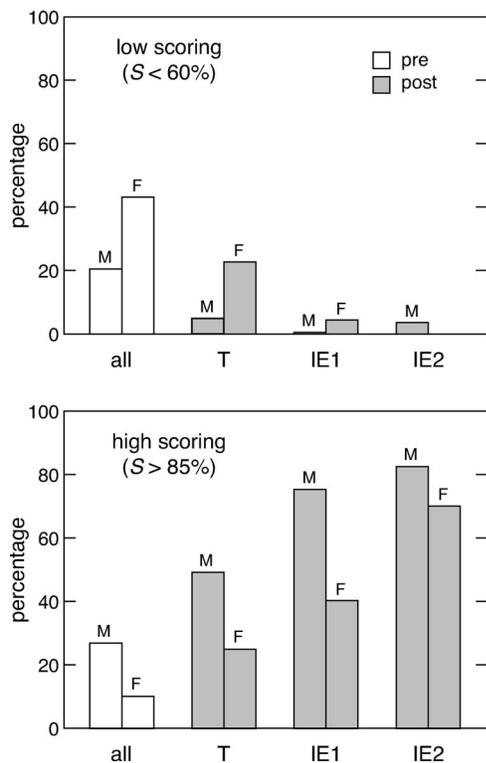


Fig. 3. Effect of instructional approach on the percentage of students with (upper panel) low ($<60\%$) and (lower panel) high ($>85\%$) scores on the Force Concept Inventory test for introductory calculus-based physics course at Harvard University. For the IE2 group the difference in percentage between high-scoring males and females is not statistically significant.

fall into the low-gain category, $\langle g \rangle < 0.3$, while the male students in the T course are at the low end of the medium-gain category, $0.3 \leq \langle g \rangle < 0.7$. Although the normalized gains in the IE groups are almost the same for both genders, females achieve higher absolute gains than males because their pretest scores are lower than those of the males. The data in Fig. 2 clearly show that increasing the level of interactive engagement during instruction not only benefits all students but also helps equalize the normalized gains of the two genders.

We now analyze how different methods of instruction affect the number of students who achieve very low or very high scores on the FCI. The authors of the FCI consider a score below 60% to indicate no real understanding of Newtonian mechanics, and a score above 85% as indicating mastery of Newtonian mechanics.²⁸ We classified students scoring less than 60% on the FCI as “low scoring” and students scoring above 85% as “high scoring.” We found no significant variation in pretest percentages from year to year for each gender, so we compared the average of all IE pretest data to the posttest percentages of the three groups.

Before instruction, the percentage of low-scoring females is twice that of the males, as shown in the upper panel of Fig. 3. Almost half (43%) of the females did not have a basic understanding of mechanics at the beginning of the term. In contrast, the percentage of females within the high-scoring category is only one-third that of the males (10% in comparison to 27%).

After instruction, the situation for female students improved dramatically in the interactive engagement courses. The percentage of females remaining in the low-scoring cat-

egory after instruction was only 4.4% for the IE1 group—a tenth of those who were in the low-scoring category initially—and there were no low-scoring females in the IE2 group after instruction. Improvement in the T course was less dramatic; 23% of female students were still in the low-scoring category, half as many as earned low scores on the pretest. In all three groups, the number of low-scoring males dropped to only a few percent (on average 2%), with no significant differences between groups.

Three features of the data for high-scoring students, shown in the lower panel of Fig. 3, are noteworthy. The percentage of high-scoring female students before instruction is small (10% for females in comparison to 27% for males). Although the percentages of high-scoring students increased somewhat after traditional instruction for both genders, it increased more after interactive instruction. In the IE2 group the female students increased their scores so much that none of them remained in the low-scoring category after instruction, and the difference between the percentages of high-scoring males and females is not statistically significant. These results show that interactive engagement courses more effectively reduce the percentage of low-scoring females and increase that of high-scoring females than traditionally taught courses.

The reduction or elimination of the gender gap in mechanics in the interactive courses is due to the remarkable improvement in the performance of female students with no observed loss of achievement among the male students. Male students also achieve at a higher level with interactive engagement instruction than with traditional instruction.

We attribute the observed reduction of the gender gap to the use of Peer Instruction, the *Tutorials*,²⁰ and cooperative quantitative problem-solving activities. These instructional methods give students opportunities to interact and explain their ideas during both lecture and section, providing frequent feedback to students on their understanding through the conceptual questions and tutorials, alternating between structured teaching and peer discussion, emphasizing conceptual reasoning, promoting collaboration among peers, and creating a less competitive classroom culture. Our results confirm that these instructional methods help reduce the gender gap in physics understanding. We hypothesize that teaching in this fashion provides a learning environment that is good for both male and female students.

IV. CONCLUSIONS

Our study demonstrates that interactive engagement effectively reduces the gender gap in physics performance. Although both genders benefit and achieve similar high normalized gains, females improve their performance most, and overcome a considerable pre-instruction gender disparity. As the attention to conceptual understanding and the level of classroom interactivity and collaboration is increased, the gender gap decreases; in the fully interactive courses, it is entirely eliminated. By creating a classroom environment that benefits both genders, the teaching approach described here improves student understanding and narrows the gender gap in physics education.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Veronica McCauley, Jessica Watkins, Dr. Jessica Rosenberg, Dr. Martin Vogt, and

Iva Maxwell for input on the manuscript; Professor Steve Wang for advice on statistical analysis; and Dr. Suwendra Nath Dutta and Vijay Salagala for technical support. ML gratefully thanks EM and the Division of Engineering and Applied Sciences at Harvard University for the opportunity to visit, and the Consejería de Educación de Castilla-La Mancha (Spain) for its generous support during her leave of absence. Her deepest gratitude goes to her brother Raimundo José for his support, encouragement, and help in all her projects and dreams. This work was supported by a grant from the National Science Foundation.

^{a)}Electronic mail: ccrouch1@swarthmore.edu

¹R. Ivie and K. Stowe, "Women in Physics, 2000," AIP Report R-430, College Park, MD, 2000 (<http://www.aip.org/statistics/trends/wmtrends.htm>).

²J. 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 (Macmillan, New York, 1994), pp. 542–557.

³I. V. S. Mullis, M. O. Martin, E. G. Fierros, A. L. Goldberg, and S. E. Stemler, *Gender Differences in Achievement. IEA'S Third International Mathematics and Science Study (TIMSS)* (Boston College, TIMSS International Study Center, Chestnut Hill, MA, 2000).

⁴R. R. Hake, "Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization," physics/0106087.

⁵A. Zohar and D. Sela, "Her physics, his physics: Gender issues in Israeli advanced placement physics classes," *Int. J. Sci. Educ.* **25**(2), 245–268 (2003).

⁶G. L. Erickson and L. J. Erickson, "Females and science achievement: Evidence, explanations, and implications," *Sci. Educ.* **68**(2), 63–89 (1984).

⁷P. Labudde, W. Herzog, M. P. Neuenschwander, E. Violi, and C. Gerber, "Girls and physics: Teaching and learning strategies tested by classroom interventions in grade," *Int. J. Sci. Educ.* **22**(2), 143–157 (2000).

⁸L. McCullough, "A gender context for the Force Concept Inventory," Paper presented at the winter meeting of the American Association of Physics Teachers, San Diego, CA, 2001 (unpublished).

⁹J. Taylor, "Sexist bias in physics textbooks," *Phys. Educ.* **14**(5), 277–280 (1979).

¹⁰L. Danzl-Tauer, "The relationship between intervention, equity, and excellence in rural high school biology classrooms," Ph.D. dissertation, Purdue University, 1990. Results described in 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 (Macmillan, New York, 1994), pp. 542–557.

¹¹A. Kitchenham, "Vive la difference: Gender, motivation, and achievement," *School Libraries in Canada* **22**(2), 34–37, 44 (2002); **22**, 44 (2002).

¹²P. Laws, P. Rosborough, and F. Poody, "Women's responses to an

activity-based introductory physics program," *Am. J. Phys.* **67**(7), S32–S37 (1999).

¹³M. Schneider, "Encouragement of women physics majors at grinnell college: A case study," *Phys. Teach.* **39**(5), 280–282 (2001).

¹⁴B. Smail, "Organizing the curriculum to fit girls' interests," in *Science for all girls?*, edited by A. Kelly (Buckingham and Philadelphia, Open University Press, 1984).

¹⁵G. M. Hilderbrand, "Redefining achievement," in *Equity in the Classroom: Towards Effective Pedagogy for Girls and Boys*, edited by P. F. Murphy and C. V. Gipps (Falmer and UNESCO, Washington, DC, 1996), pp. 149–171.

¹⁶H. Stadler, R. Duit, and G. Benke, "Do boys and girls understand physics differently?," *Phys. Educ.* **35**(6), 417–422 (2000).

¹⁷R. Kimbell, K. Stables, T. Wheeler, A. Wosniak, and V. Kelly, *The assessment of performance in design and technology* (London, School Examinations and Assessment Authority, 1991), cited in D. Y. Yip, M. M. Chiu, and E. S. C. Ho, "Hong Kong Student Achievement in OECD-PISA Study: Gender Differences in Science Content, Literacy Skills, and Test Item Formats," *Int. J. Sci. Math. Ed.* **2**(1), 91–106 (2004).

¹⁸E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997). Additional information and resources for Peer Instruction can be found at (<http://galileo.harvard.edu>).

¹⁹C. H. Crouch and E. Mazur, "Peer instruction: Ten years of experience and results," *Am. J. Phys.* **69**(9), 970–977 (2001).

²⁰L. C. McDermott, P. S. Schaffer, and the University of Washington PERG, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 1998).

²¹Eric Mazur, *Introductory Physics* (to be published).

²²P. Heller, R. Keith, and S. Anderson, "Teaching problem solving through cooperative grouping. 1. Group versus individual problem solving," *Am. J. Phys.* **60**(7), 627–636 (1992); P. Heller and M. Hollabaugh, "Teaching problem solving through cooperative grouping. 2. Designing problems and structuring groups," *ibid.* **60**(7), 637–644 (1992).

²³D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *Phys. Teach.* **30**(3), 141–151 (1992). Revised in 1995 by I. Halloun, R. R. Hake, E. Mosca, and D. Hestenes. The revised version appears in Ref. 18 and can be obtained from Professor Hestenes at Arizona State University.

²⁴R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *Am. J. Phys.* **66**(1), 64–74 (1998).

²⁵D. Hestenes and M. Wells, "A mechanics baseline test," *Phys. Teach.* **30**(3), 159–166 (1992).

²⁶See EPAPS Document No. E-AJPIAS-74-003603 for tables of the Mechanics Baseline Test scores and Force Concept Inventory gains and normalized gains. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>).

²⁷C. Henderson, "Common concerns about the Force Concept Inventory," *Phys. Teach.* **40**(9), 542–547 (2002).

²⁸I. Halloun and D. Hestenes, "Interpreting the Force Concept Inventory: A response to Huffman and Heller," *Phys. Teach.* **33**(8), 502–506 (1995).