“Assessing and Enhancing the Introductory Science Course in Physics and Biology: Peer Instruction, Classroom Demonstrations, and Genetics Vocabulary”

A thesis presented

by

Adam Paul Fagen

to

The *Ad Hoc* Committee on Molecular Biology and Education

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Molecular Biology and Education

Harvard University
Cambridge, Massachusetts

May 2003

Advisor: Professor Eric Mazur Adam Paul Fagen

"Assessing and Enhancing the Introductory Science Course in Physics and Biology: Peer Instruction, Classroom Demonstrations, and Genetics Vocabulary"

Abstract

Most introductory college science courses in the United States are taught in large lectures with students rarely having the opportunity to think critically about the material being presented nor to participate actively. Further, many classes focus on teaching rather than learning, that is, the transfer of information as opposed to actual student understanding. This thesis focuses on three studies about the assessment and enhancement of learning in undergraduate science courses.

We describe the results of an international survey on the implementation of Peer Instruction (PI), a collaborative learning pedagogy in which lectures are interspersed with short conceptual questions designed to challenge students to think about the material as it is being presented. We present a portrait of the many instructors teaching with PI and the settings in which it is being used as well as data on the effectiveness of PI in enhancing student learning in diverse settings. The wide variety of implementations suggests that PI is a highly adaptable strategy that can work successfully in almost any environment. We also provide recommendations for those considering adopting PI in their classes.

Classroom demonstrations are an important aspect of many introductory science courses, but there is little evidence supporting their educational effectiveness. We explore the effect of different modes of presentation on enhancing student learning from demonstrations. Our results show that students who actively engage with a demonstration by predicting the outcome before it is conducted are better able to recall and explain the scenario posed by that demonstration.

As preliminary work for the creation of an inventory of conceptual understanding in introductory biology, we discuss results from a survey of vocabulary familiarity and understanding in an undergraduate genetics course. Students begin introductory classes with significant gaps in their understanding, some of which are retained beyond instruction. Further, they overstate their knowledge, and the degree to which they exhibit overconfidence increases over the period of instruction.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I confess that this is the first section of my thesis that I started writing, and it is the last as well. But there are a lot of people who helped me make it through nearly ten years of graduate school. It actually seems a little strange for the dissertation to be a single-author paper, since so many others contribute to the effort and dedication needed to get through a Ph.D. program.

First, I thank those that allowed me to pursue the Ph.D. program that met my needs and interests. Most vitally important are the members of my committee who have gone out of their way to accommodate me in doing something "non-traditional." The support, advice and encouragement from Eric Mazur, Philip Sadler, Everett Mendelsohn, Matthew Meselson, and Daniel Hartl have been above and beyond what is normally expected of a thesis committee. I would especially like to single out the chair of my committee, Eric Mazur. After a single 15-minute meeting on Commencement Day 1998, somehow this physicist with more than enough on his plate already was willing to help guide a biologist through a program in science education. The welcome and openness from Eric and Catherine Crouch made me feel like I could have a home in Gordon McKay Laboratory.

Several of the projects described in this thesis have been carried out with colleagues who were (obviously) integral to their success. Catherine Crouch was with me nearly every step of the way of the projects described in Chapters 2 and 3. In many ways, she should get a co-authorship on these chapters as the work described there was very much a collaboration.

For the Peer Instruction survey described in Chapter 2, I also acknowledge the help of Tun-Kai (T.-K.) Yang and Katy Gladysheva in helping to identify and categorize the responses, Richard Hake and several pilot-testers for their comments on the content of the survey, and Michael Tucker and Brady Richards for setting up the technology to receive survey responses via the Web.

The demo study described in Chapter 3 was initiated by Paul Callan before I joined the Mazur group. Although Paul's results were never published, they set the stage for the work described here and helped us anticipate and avoid many problems; Paul has also been a continuing source of good advice on the project. The demonstrations themselves could not have been conducted without the support of the Lecture Demonstration Services in the Harvard Science Center, especially Nils Sorensen and Wolfgang Rueckner. I am especially grateful to the TFs for the Fall 2000 and Spring 2002 semesters who played such a critical role in the data collection.

The genetics vocabulary survey described in Chapter 4 could not have been conducted without the help of the Spring 2001 Teaching Fellows for Bio Sci 50 and especially Head TFs David Mark Welch and Colin Meiklejohn. I also thank Daniel Hartl for his support and interest in the project and its results.

I have always felt welcome in the Mazur group, despite my almost complete cluelessness with respect to the “something with lasers” that goes on behind the doors with the blinking “laser on” signs. Eric has managed to bring together a group of individuals who not only do excellent work, but who are also genuinely nice people. I am pleased that I can call my labmates friends as well as colleagues.

Many wonderful people have come through the Mazur Group during my time there: graduate students, postdocs, visiting scientists, undergraduate students, high school teachers and high school students. All have helped make the group such a welcoming place. Paul Callan gets credit for bringing me to the lab in the first place by letting me know about Eric’s interest in education and encouraging me to speak with him. Rebecca Younkin brings an incredible sense of integrity to whatever she does and provided the conscience of the Mazur group in addition to much helpful advice. Jon Ashcom is a careful scientist and a genuinely nice guy; his sense of perspective in finishing his own thesis was very helpful in finally finishing mine a month later. Nan Shen and I were even closer in finishing, with Nan defending only a week before I did; I know that I could not have finished without her support, including the New Years Eve spent in the lab thesis-writing. Jim Carey really goes out of his way to make people feel comfortable and welcome and is always available to help out. Chris Roeser always seems to know what’s going on, but in a very unassuming way. Rafael Gattass and Iva Maxwell have continued to make the lab a fun place to be, have provided good fodder for kidding around, and are some of the friendliest people I’ve ever known. Maria Kandyla provides the lab with a sense of style even if she never does eat lunch with us. Although her time at Harvard was too short, Arantza Mendioroz had it covered with style, good nature, and lunch. Chris

Schaffer always seemed like he was already a faculty member, uniting a real passion for science with deep subject knowledge and the a natural ability to mentor undergraduates, young graduate students and even a few high school students. Albert Kim was a lot of fun to have in the lab and helped me not to feel too old. It's nice that even though Mike Sheehy is officially in the Chemistry Department, he spends most of his time in McKay.

My education colleagues within the Mazur Group have preventing my independent journey through an *ad hoc* program from being a lonely one. Catherine Crouch is always generous with her time and her willingness to help out whenever I needed it. Suvendra Nath Dutta is a great asset to the group and I have enjoyed talking to him about education, science, and anything else. His phenomenal technical knowledge keeps the group running and his adorable son keeps us in good spirits. I am glad that I have had the chance to work with Veronica McCauley this past spring, and I know that I leave the education research mission of the Mazur Group in good hands.

To every other current and past member of the Mazur Group, thank you for making my experience so rewarding and so much fun.

Several people were very gracious to read or otherwise comment on various parts of this thesis and their comments and suggestions have helped to improve it significantly. Thanks especially to Catherine Crouch for a Herculean task at the last minute, and sincere appreciation to Veronica McCauley, Suvendra Nath Dutta, Paul Callan, Jon Ashcom, Iva Maxwell, Nan Shen, Jim Carey, and of course, my thesis committee of Eric Mazur, Phil Sadler, and Dan Hartl.

An *ad hoc* Ph.D. program is not easy to put together and the input on the proposed plan of study and proposal for the Administrative Board was much improved by conversations and courses with Robert Brain, Catherine Crouch, Bruce Gregory, Alison Grinthal, David Hart, John Holdren, Everett Mendelsohn, Irwin Shapiro, Rise Shepsle, and Natalie Taback.

Even though I decided that a biology lab career is not the thing for me, I am forever grateful for my experience in the Meselson lab. David and Jessica Mark Welch provided such a welcoming environment and I never felt silly asking a simple question or getting their advice. Barbara Ring kept everything running smoothly, and Jingjing Li, Irina Arkhipova, Sandy Ropper, and everyone else made it such a friendly place.

Teaching was not only a big part of my time at Harvard but also of my research. I know that my teaching strategies and philosophies were influenced by conversations with and observation of good teachers. I am especially grateful to Robert Lue, Dan Hartl and, of course, Eric Mazur, for allowing me to use their classrooms as laboratories. I also owe a debt of gratitude to the students I have had the honor and pleasure of interacting with in Bio Sci 1, 10, 14, and 50; Biology S-110; and Physics 1a. They have provided me, not only with data, but also with honest and open feedback on their educational experiences and motivation to continue working to improve their education.

My teaching colleagues and others committed to undergraduate education were generous with their interest, support, and good ideas; I would specifically like to acknowledge Paul Callan, Catherine Crouch, John Girash, Cassandra Volpe Horii, David Mark Welch, Jessica Mark Welch, Eric Mazur, Emily Patek, Dmitri Petrov, Rebecca Younkin, Sarah Evans, Emily Humphrey, Tara Harmer, Katherine Hannon, Mandana Sassanfar, and Nina Reiniger. Those from outside of Harvard include Gina Andrighetto, Kristy Beauvais, Mercedes Lorenzo, Sanjoy Mahajan, Sheila Tobias, Ceanne Tzimopolous, Meghan Walbran, Bonnie Wood, and Michael Zielik.

The continued support from my Swarthmore College professors, especially during times of indecision and uncertainty in graduate school, helped to remind me why education and caring are so important. Lois Weisman, biochemistry professor and cousin, helped to provide grounding and support at family gatherings. A special acknowledgement also goes to Robley Williams who gave me my first chance to actually *do* science after only my first year or college, put my name on my first scientific paper, and remained my advocate through the three summers I worked in his lab and beyond.

One of my major activities during graduate school was my involvement with the Graduate Student Council, including two years as president. In addition to working on issues to help improve the graduate student experience at Harvard, I was fortunate to work with a number of similarly committed individuals that made my time on GSC especially valuable; thanks to Paul Callan, Elgin Eckert, Robert Krikorian, Lisa Laskin, Ian Richmond, Rebecca Spencer, Shaun Rein, Laurie Littlepage, Christine Murasaki Millett, Kyriell Muhammad, Mike Westover, Stephanie Abundo, Matthew Leingang, Sharrona Pearl, Julie Rohwein, Wendy Roth, and Greg White. We were also lucky to work with a caring and committed administration within the GSAS and Dudley House, including Garth McCavana, Susan Zawalich, Peter Ellison, and Everett Mendelsohn.

Beyond Harvard, I also got involved with graduate student activism on a national scale, especially through the National Association of Graduate-Professional Students (NAGPS), where I served as chair of The National Doctoral Program Survey and chair of the *ad hoc* Committee on Faculty-Student Relations. Again, I had the good fortune to work with a number of caring and committed individuals including Malaina Brown, Susan Mahan Niebur, and Kimberly Suedkamp Wells, and I know our friendships will extend well beyond NAGPS.

Finally, I must credit my friends who stood by me, gave me a push when needed, listened to me whine and complain or ramble on about some topic which mattered to me. Alison Grinthal and I met during orientation at Swarthmore in September 1993 and we have remained close friends ever since. Since she and I came to Harvard together, she was always around and available to understand what we were going through, and she helped give me needed perspective during this long graduate odyssey. Hopefully now that this thesis is done, I will actually keep in better touch with other friends from Swarthmore, especially Cindy and Andrew Merrill and Anna and Marc Syvertsen.

To go along with fellow Swatties from the 1990s, I have been glad to make some new friends at Harvard and elsewhere, who have helped me stay sane. Malaina Brown has been one of the people who pushed me the most while always having my best interests at heart, trying to be sure that I actually did finish this degree. Her encouragement, support, and compassion really helped me make it to the end. Natalie Taback, in addition to being a caring friend, helped convince that finishing a thesis could be done, even in short order. Leah Kronenberg always provided an encouraging word and a "you can do it" when I most needed it. I must also acknowledge a number of people who I may not have interacted with as much as I would have liked, but who helped make my non-academic life fulfilling, through personal friendships, Dudley House dinners, or any number of other ways. To the Dudley table including Ann Bragg, Bayliss Camp, Alison Grinthal, Jeff Johnson, Bernie Perley, Ken Rines, Cynthia Soroos (ha, I did beat you to your law degree), Natalie Taback, and Wendy Zhang. For those who I haven't already mentioned but whose friendship and support meant a lot, to friends and colleagues from the dorms, activities, and elsewhere who include (but are surely not limited to): Avril Alba, Seth Berger, Tamar Gonen Brown, Scott Damrauer, Kari Davis, Shasa Dobrow, Heather Gray, Wendy Beth Hyman, Adam Juda, Avi Kogan, Amy Lakin, Dan Levenson, Judith Leone, Jia Liu, Katherine Lu, Eric Marler, the Nose, Lil Pabón, Melissa Prosky, Lisa Schechter, Gabi Soble, Diana Stiefbold, Tsuyoshi "Hashi" Takahashi, Alexandra Vega, Philip Webb, and Rhetta Wiley.

And finally (!), thank you to my family for their unwavering support and encouragement over the long years of graduate school.

Acknowledgement of financial support

Research funds for much of the work described in this thesis were provided by several grants from the National Science Foundation (NSF) Division of Undergraduate Education (including DUE-9554870 and DUE-9980802). I was personally supported by a Graduate Research Fellowship and later a Graduate Teaching Fellowship in K-12 Education from the NSF as well as several Teaching Fellowships from Harvard University and the Research Experience for Teachers program through the Harvard MRSEC.

Chapter 1:

The introductory science course

Introductory science courses at most colleges and universities have one thing in common: they are commonly quite large. Maximizing teaching resources has driven the tendency to put several hundred students into a lecture hall and to provide a lecturer to stand at the front and deliver a prepared presentation to the hundreds—or even thousands—in attendance [Stokstad 2001]. While lectures may be an effective pedagogy for *some* students in *some* classes, it is probably not the most effective way to teach most classes. It is hard to get around the efficiencies of lectures: one lecturer can transmit information to hundreds or thousands of students at one time, even if those students may not actually be learning what is taught.

There have, however, been a number of new pedagogies and suggestions for how to implement more active learning experiences, even in large introductory lecture courses [*e.g.*, Ebert-May *et al*. 1997; Lawson *et al*. 1990; MacGregor *et al*. 2000; Mazur 1997; McNeal and D'Avanzo 1997; National Research Council 1997; Udovic *et al*. 2002; Wyckoff 2001]. Many of these innovations are targeted toward introductory science courses that generally enroll non-majors, the students who might not ever take another course in that discipline. In general, the strategies have been designed to directly confront student misconceptions [*e.g.*, Driver *et al.* 1985; Novak 1987; Wandersee 1985] and to use an inquiry-based approach in order to provide students with a more genuine learning experience than a passive lecture affords. These pedagogies instill science as a process and way of knowing rather than as a body of static knowledge.

Distinction between teaching and learning

All too often we conflate and confuse two distinct processes: teaching and learning. This is especially true in the introductory science course where the overwhelming pedagogy is the lecture. An instructor may think that he or she has "covered" a topic by presenting a lecture on the subject, regardless of whether the students have actually understood the material. But education is not only about the *process* of instruction, but also about learning. It thus begs increased attention to assessment and evaluation. In particular, this suggests carefully defining course objectives, refocusing classroom practice upon enhancing student understanding, clarifying student learning goals, engaging students in their own learning, offering opportunities for them to give regular feedback on their learning, and designing assessment tools to see if the learning goals are being met [Seymour 2002].

Assessment

No feature of a course provides more incentive for students than examinations and other graded assignments. Research indicates that assessment tools emphasize to students what they should be focusing on [Mintzes *et al.* 2001; Tobias 1998, 2000]. A sound pedagogy will not be successful unless student assessment tools complement the goals and provide justification for the pedagogical strategy. This underscores the importance of genuine assessment and evaluation tools that do not merely tests recall of facts or mastery of a “plug and chug” type procedure for solving problems without understanding. Assessment tools should help achieve formative assessment, helping to diagnose difficulties in understanding so that they can be addressed at the time of instruction and thereby help to improve student understanding on the topics currently being discussed. Otherwise, students fail to get appropriately challenged and instructors fail to get feedback on their students’ understanding until weeks after that material is discussed. As with instructional material, assessment tools must also take into account our current understanding of cognitive development [Sadler 2000].

Content of the thesis

This thesis includes three main chapters, each of which describes a major research project on introductory science education. The topics of the three chapters are quite different from each other and may largely be considered independently from one another. However, they all relate to the introductory science course and ways to make that course the most effective educational experience for students.

Chapter 2: Assessing the effectiveness of Peer Instruction: Results of a Survey

The first research chapter describes Peer Instruction (PI), a research-based pedagogy for teaching large introductory science courses developed by Eric Mazur. This teaching strategy maintains the efficiencies of large lectures, but also involves students in their own learning by turning the lecture into a seminar or, more accurately, into many seminars. The chapter reports upon the results of a worldwide survey of PI users, discussing the range of settings and implementations in which PI has been implemented. It also provides recommendations for those looking to implement PI in their own courses.

Chapter 3: Enhancing the effectiveness of classroom demonstrations

The next chapter describes a study of classroom demonstrations conducted in introductory physics at Harvard University. In particular, we investigated whether altering the pedagogy by which the demonstration was presented affecting student learning of the concepts involved in that demonstration. We compared demonstrations presented traditionally to those in which students were asked to predict the outcome prior to observing the demonstration to those in which students were explicitly asked to evaluate their predictions upon observing the demonstration and to discuss their observations with their classmates.

Chapter 4: Informing conceptual biology assessment by examining genetics vocabulary

The final research chapter describes preparative work for the creation of a test of conceptual understanding in introductory genetics. After providing a context for such a conceptual inventory and discussing some of the relevant considerations for constructing one, the chapter describes an analysis of students' familiarity with and understanding of key vocabulary at the beginning and end of an introductory genetics course at Harvard University.

Chapter 2:

Assessing the effectiveness of Peer Instruction: Results of a survey

Peer Instruction (PI) is a widely-used pedagogy in which lectures are interspersed with short conceptual questions—called ConcepTests*—designed to reveal common misunderstandings and to engage students actively in lecture courses [Mazur 1997 Crouch 1998; Crouch and Mazur 2001]. Students are given a minute or two to consider the ConcepTest before committing to an answer. They are then asked to turn to classmates seated nearby in order to convince each other of the correct response. After a few minutes of discussion, students are asked to record a second answer, which has been informed by this peer interaction. Only at the end of this process does the instructor provide the correct answer and explanation. The basic idea behind PI is to get students actively engaged in their own learning, to think critically about the material during class, and to learn and teach each other in the controlled setting of the classroom; PI also provides immediate feedback to the instructor on students' understanding of the material.

* The word ConcepTest is a contraction of "concept test" and was coined by Eric Mazur to refer specifically to questions asked in a Peer Instruction setting [Mazur 1997].

PI is mainly used in introductory science courses though there is no reason it cannot be used for a wide range of subjects. The active learning incorporated in PI are consistent with research in psychology that shows learning is improved when students are actively involved with the material, instead of being passive listeners [Biggs 1996]. Students not only understand course material better when they learn it actively, but they also retain it longer and enjoy their courses more [Bonwell and Eison 1991; Murray and Brightman 1996]. PI works by breaking up the lecture into a series of smaller chunks that may make the course seem more manageable to students. The 10-15 minute chunks of lecture are more consistent with student attention span than the 50- or 90-minute monologues by the professor common in many lecture courses [Liebman 1996; Stuart and Rutherford 1978].

The hallmark of PI is the ConcepTest. It is easy to insert one or many ConcepTests into an existing lecture to make any pedagogy more interactive. As one professor was quoted as saying about PI, "it doesn't require a radical revision of the curriculum" [Mackenzie 1997]. Because of the flexibility for supplementing other strategies with ConcepTests, PI can be easily adapted to the wishes of the instructor, the characteristics of the class, and the content of the material. As such, PI differs from some other research-based pedagogies that involve a well-defined series of procedures or complex sets of materials that must be used in a proscribed fashion. Many instructors have taken full advantage of the flexibility of PI by adapting it to their local setting.

Individual correspondence and informal discussions has indicated a user base of hundreds of instructors around the world who teach with PI, yet to date, there has been no systematic study of the implementation and effectiveness of PI in the variety of settings in which it is used. As a step toward such a systematic study, we polled current and former PI users via a Web-based survey to learn about their implementation of and experience with PI. The survey collected data about how instructors learned about PI, courses in which PI was used, implementation details, course assessment, effectiveness, instructor evaluation, and the community of PI users.

Review of Peer Instruction

The canonical description and exposition of PI is in Eric Mazur's *Peer Instruction: A User's Manual* [Mazur 1997], which also includes some early data on the effectiveness of PI in introductory physics at Harvard (Physics 11, mentioned in Chapter 3). Early assessment includes student performance on a standardized conceptual assessment instrument for introductory mechanics (the Force Concept Inventory, discussed in more detail below) as well as comparative data on students' performance on conceptual and traditional quantitative problem-solving [Mazur 1997]. Since then, the Mazur Group has collected over 10 years of data on the use of PI at Harvard, showing further improvement in student learning as we have refined our own use of PI and supplemented PI with other enhancements to the course including tutorials. More recent data on our experience with PI at Harvard have been recently reported by Crouch and Mazur [2001].

In addition to the original *Peer Instruction: A User's Manual* [Mazur 1997] which describes the use of PI in physics and includes a library of physics ConcepTests, several other collections of ConcepTests have been published or soon will be. Landis *et al.* [2001] have published a recent collection of chemistry ConcepTests, drawn from an online database maintained at the University of Wisconsin. Paul Green [2003] has done likewise for astronomy. And a book compiling ConcepTests in mathematics is forthcoming from David Lomen and colleagues (to be published by Wiley). These books supplement the existing ConcepTest online collections in physics[†], chemistry[‡], and astronomy[§].

Literature review

One of the earliest publications to discuss Peer Instruction is Sheila Tobias' *Revitalizing Undergraduate Science: Why Some Things Work and Most Don't* [Tobias 1992], which highlights a number of promising advancements in undergraduate science education including PI. PI has been featured as a promising practice for undergraduate science education by the National Research Council [1997, 1999] and the Boyer Commission on Educating Undergraduates in the Research University [1998].

PI was also one of the pedagogies included in Richard Hake's comprehensive survey of introductory physics courses [Hake 1998]. Hake's study compared "traditional" or "conventional" instruction—passive lectures, algorithmic problem-solving, and "cookbook" laboratory exercises—with various "interactive engagement" pedagogies. Hake described what he meant by interactive engagement as methods

> designed in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors [Hake 1998].

[†] http://galileo.harvard.edu/

[‡] http://www.chem.wisc.edu/~concept/; http://people.brandeis.edu/~herzfeld/conceptests.html

[§] http://hea-www.harvard.edu/~pgreen/educ/ConcepTests.html

Hake found a clear difference between traditional and interactive engagement classes, making the claim than interactive pedagogies—including PI—are more effective at enhancing student learning than traditional passive lectures.

van Dijk *et al*. [2001] assessed the difference between merely posing ConcepTests in class to which students would respond, and also requiring the students to discuss their answers. They performed an experimental study in which they varied the way the same lecturer presented the same topic in an engineering course at the Delft University of Technology in The Netherlands. They randomly assigned first-year students to one of three groups: an experimental group in which they posed questions to the students to which they responding using an Interactive Voting System (IVS; similar to the PRS or ClassTalk system used elsewhere), but without discussion; a second experimental group which employed the IVS and a full implementation of PI, including student discussion; and a "traditional lecture" control group in which the same questions were posed rhetorically with no opportunity to respond. They found that use of the IVS without PI discussion produced significantly lower scores on a content post-test that both the IVS+PI group and, surprisingly, the lecture control group. This is somewhat consistent with the observations of outside observers they solicited, in which students in the IVS-only group were more passive than the control group in terms of "using opportunities to initiate student involvement," even when such opportunities were not provided by the lecturer. Even though the lecturer for the control group did not explicitly include any interactive activities, van Dijk *et al*. [2001] argue that the lecturer did activate students implicitly, such as by employing humor and non-verbal behavior in the traditional lecture class to a greater degree than in the IVS and IVS+PI classes. Students did not report significant differences between the three methods when evaluating the amount of student activation or student involvement; interestingly, outside observers rated the difference in student involvement to be highly significant when observing both students and lecturer. Students who were activated in lecture did have a positive reaction to it, found it useful, and believed that it would contribute to their learning [van Dijk *et al*. 2001]. Thus, even in the absence of a clear effect in learning itself, active pedagogies may have an important *motivational* effect on students and their willingness to be involved in the subject material.

Murray [1999] has described the use of PI in a civil engineering course at Queensland University of Technology in Australia. He constructed a course with a new model of student-centered learning, where many different activities are used to enhance the student's learning experience and allow the students to interact with each other and with the material. His model builds on the work of Gedalof [1998], who suggested that it is more valuable for students to learn from one another than it is for them to be taught directly through a passive lecture. Murray develops ConcepTests based on 3-4 key concepts for each topic. The discussion of these ConcepTests makes the topics come alive for the students, generating interest in which of the possible options is the correct one and why. Further, the discussion helps to reveal common student misconceptions, which are quickly exposed as students try to convince each other that their answer is correct. This collaborative spirit is further enhanced by "Supplemental Instruction" [Healy 1994; Martin and Arendale 1994; Taylor *et al*. 1994], in which students are given the opportunity to work collaboratively with the guidance of a trained second-year student, in a manner very similar to the way we implement tutorials in introductory physics at Harvard. Murray [1999] also sees evidence that the conversations in class often carry over to networks outside of class, so that the discussions can continue, even outside of the lecture.

Meltzer and Manivannan [2002] describe an implementation of PI at Iowa State University, Southeastern Louisiana University, the University of Virginia, and Southwest Missouri State University. They have essentially abandoned the use of classtime for presenting detailed explanations of physics principles and, instead, spend class guiding students through consideration of questions and answers in much the same way as one-on-one tutoring. They sometimes use a very high rate of questioning, as much as several ConcepTests per minute, possibly including questions created "on-the-fly" by the instructor.

Bullock *et al*. [2002] describe a procedure for enhancing the frequency of student-interaction frequency in an introductory physics course at the University of Arkansas, tracking attendance, pre-class preparation, in-class participation, homework completion, and exam scores. They employed in-class quizzes analogous to PI using a wireless classroom network in the second-semester electricity and magnetism course; the first-semester mechanics course did not use such questions and provides a control to the experimental group as the same students were enrolled in both courses. For some questions in the experimental group, the instructor employed PI by having the students discuss their answers with each other, after the class distribution from an initial poll was displayed. For questions to which many students initially responded incorrectly, showing the class histogram communicated to students that the majority of the class had an erroneous thought process and should critically reconsider their answers. Bullock *et al*.

[2002] also implemented Web-based utilities for students to answer chapter quizzes and homework. The class saw significant improvements in student participation in all areas: attendance increased 130%, class participation increased 1700%, pre-class preparation increased 1100%, and homework completion increased 650%. Student performance improved significantly with the addition of the technological components: the final exam score increased from a score of 45% to a score of 75%, a gain of 70% [Bullock *et al.* 2002].

Rao and DiCarlo [2000] describe the use of Peer Instruction during one section of a medical physiology course. They found that the percentage of correct answers to the multiple-choice questions asked increased significantly ($p < 0.05$) following discussion among students. The improvement was especially pronounced for higher-level intellectual questions, as compared with simple recall questions, although they saw improvement for all types of questions asked.

Piepmeier [1998] describes the use of ConcepTests in a large lecture course in pharmacy. He presented 2-3 ConcepTests in a one-hour lecture, but did so with a twist. Instead of asking a ConcepTest *following* discussion of a topic, Piepmeier asked the question, solicited responses, and *then* gave a 15-20 lecture on the topic. At the end of this mini-lecture, he again asked for students to respond with their answer. Only then did he have students discuss their answer with their classmates, which was followed by having students submit a *third* answer to the same ConcepTest.

Savinainen and Scott [2002b] employ PI as one aspect of what they term "Interactive Conceptual Instruction" (ICI) in a Finnish upper secondary school. Additional features of ICI are a conceptual focus to the course in which topics are motivated at a conceptual level before applying mathematics and quantitative problem-solving; the use of research-based materials; pre-class reading; and other activities such as concept maps. They observed a normalized gain[**], *g*, of 0.57 [Savinainen and Scott 2002a, 2002b], consistent with other interactive engagement courses, such as those reported by Hake [1998].

Following the work of Wright [1996], Kovac [1999] describes the use of PI in a pair of introductory chemistry courses at the University of Tennessee. He used up to four ConcepTests per class and polled students by asking them to raise their hands for the correct answer. Students viewed both ConcepTests and the cooperative learning workshops used as well to be positive aspects to the course: more than 64% agreed that the two methods were helpful in learning course material, while less than 13% disagreed.

Wimpfheimer [2002] also discusses ConcepTests in chemistry, considering the differences of using them in a small class. He used PI in a fourth-semester general chemistry course taken by biology and chemistry majors after two semesters of organic chemistry; thus, in addition to being a small class, this setting is more advanced than many of those in which PI is used. Wimpfheimer [2002] mentions the increased self-consciousness of the students in such a small setting and how even 1 or 2 students changing their votes could noticeably alter the class distribution. However, the inability of students to "hide" their incorrect answer also meant that students could not hide from giving any answer; in fact, he observed participation close to 100% throughout the semester.

[**] Normalized gain, *g*, is discussed in more detail below.

Schlatter [2002] describes the development of ConcepTests to teach multivariable calculus at Centerary College of Louisiana. He most frequently used questions that he termed "comparison" that do not involve direction computation, even in a quantitative class, *e.g.*, determining the sign or relative magnitude of a quantity. Other uses of ConcepTests included three-dimensional visualization, translation between coordinate systems, theorem-using, and theorem-provoking to prime students for upcoming material. Schlatter [2002] discusses the value of learning students' strengths and weaknesses, and maintaining high levels of student interest. The Schlatter paper does not discuss any research data, but provides proof-of-concept information for the use of PI in a mathematics class and the success of using conceptual questions, even in the context of a quantitative course. Pilzer [2001] provides a similar explanation of the use of PI in another mathematics course, and the forthcoming volume on ConcepTests in calculus provides further evidence for the use of PI in mathematics.

PI and ConcepTests has also been reported used in teaching computer science [Chase and Okie 2000], statics [Danielson and Mehta 2000], pharmacology [Near *et al.* 2002], and physics in a Chinese military academy [Yang *et al.* 2001].

In addition to the canonical PI, various instructors have introduced a number of variants to PI. For instance, Bill Junkin at Erskine College and Anne Cox of Eckerd College pair groups of students together for discussion based upon their responses to an initial question [Cox and Junkin 2002]. This way, they can assure that the discussions are as valuable as possible since the instructors choose the composition of the groups. For instance, rather than having students who all agree on the answer to a question discussing it amongst themselves—as can happen in PI—they deliberately put together groups of students who will benefit from their discussion. Cox and Junkin [2002] also report the successful use of PI in a laboratory setting as opposed to a lecture.

Survey methodology

As a result of the anecdotal reports of PI use, we chose to investigate PI users worldwide in an attempt to classify the variety of implementations used and to try to identify the important elements for the success of PI in diverse settings. The objective was to carry out a census of PI users to get a realistic picture of PI as actually used at schools and colleges around the world.

The Peer Instruction/Collaborative Learning Implementation Survey (reproduced in Appendix A) was targeted at instructors who had used PI—even if they do not refer to it by that name[††]—in one or more courses at any grade level. The survey asked respondents to report on their experience with PI for the course that was "most representative" of their experience (or most recent). The survey asked questions in a variety of areas:

- personal information (including biographical and contact information for respondents and their involvement in educational activities and education research);
- background on Peer Instruction (how they initially and subsequently have learned about PI);
- course information (subject, level, enrollment, frequency, format and activities);

[††] For instance, PI shares many features with the strategies known as "Think-Pair-Share" [Lyman 1981, 1992] and "Think-Tell-Share" [Thorton 1991].

- implementation of PI (including how often PI was used, polling method for ConcepTests, types and source of questions, and student access to ConcepTests outside of class);
- grading and assignments (whether the course was graded on a curve or on an absolute scale, credit given for ConcepTest participation, and details about pre-class reading and any associated reading assignments);
- results (both qualitative and quantitative measures of student achievement);
- evaluation (qualitative assessment by instructors on their own satisfaction with the method, student satisfaction, student participation, instructor effort, and difficulties encountered); and
- community (other PI users at their institution and elsewhere and the awareness of pedagogical issues within their department).

The survey was designed in collaboration with Catherine Crouch and Eric Mazur, based upon their experiences with PI and what we have learned to be important parameters for implementation of the pedagogy. We also solicited feedback from other experts in physics education research including pilot-testing the survey instruments with a small group of college physics instructors who use PI in their classes.

Sample population and procedure for soliciting responses

The survey was posted on the Project Galileo Web site* in late May 1999 with most responses collected by the fall of 1999. Participation was invited via e-mail in a number of ways. All registered users of Project Galileo (over 2100 people at the time)

---

* Project Galileo is an NSF-funded Web site of class-tested strategies for teaching science operated by the Mazur Group. The survey was posted at http://galileo.harvard.edu/PIsurvey.html and is also included as Appendix A of this thesis.

were e-mailed to let them know about the survey and to encourage their participation. Similar messages were also distributed to the lists of contributors and users of the ConcepTest databases for chemistry and astronomy. Anyone who had corresponded with a member of the Mazur Group about PI was added to the list.

We contacted faculty who had hosted a recent education talk by a member of the Mazur Group and asked them to pass along the names and e-mail addresses of their colleagues who were teaching using Peer Instruction. We informed several years' participants in the Workshop for New Physics Faculty sponsored by the American Association of Physics Teachers (AAPT) and the American Physical Society (APS) and the AAPT/APS Department Chairs Conference, settings in which Eric Mazur had given presentations about PI. The instructors who had participated in "Teaching Introductory Physics, Conservation Laws First: An NSF Faculty Enhancement Conference," hosted at Harvard University in June 1998, were also encouraged to respond, since many of them were known to have adopted PI. A number of names[‡‡] were also provided by Prentice Hall, the publishers of *Peer Instruction: A User's Manual* [Mazur 1997]. Finally, some PI users were identified by searching the Web for terms such as "Peer Instruction" and "ConcepTest."

We also collected additional names within the context of the survey. In addition to soliciting responses to the survey itself, the e-mail message inviting participation also asked recipients to pass along the names and e-mail addresses of any colleagues whom they knew to be using PI. This type of "viral marketing" was successful in building up

---

[‡‡] The names provided from Prentice Hall are only those instructors who received the first edition of the book (which mistakenly contained only Macintosh versions of the associated diskettes) and had requested Windows versions, so they represent only a small fraction of the thousands of instructors who have read the book.

the list of PI users with the assistance of others, many of whom were not using PI themselves. Thus, even though many of those initially invited to complete the survey were not using PI themselves, they helped identify other instructors who were. The survey instruction also asked a similar question, asking respondents to identify other PI users (see question H2 in Appendix A).

Altogether, we directly invited over 2700 instructors by e-mail to complete the survey; others were forwarded invitations about the survey, since several of those responding were not already in our contact database. Over 700 instructors completed the survey, reporting their experience with collaborative learning. The language of the survey was purposely broad in order to include instructors who had used a strategy similar to PI without being aware of our work; we therefore received responses from many instructors using other collaborative learning strategies. By looking at a number of implementation details and respondents' familiarity with PI, we identified 384 responses who were using a strategy that we identify as essentially identical to Peer Instruction [Mazur 1997; Crouch and Mazur 2001]. In short, to be classified as PI, we looked for instructions who asked questions that students thought about and answered before discussing them with each other. Several instructors used variants of PI in which they either did not have students submit an answer or did not have the student discussions; such strategies were not included in the analysis.

Response rate and characterization

As described above, over 700 instructors responded to the Peer Instruction/ Collaborative Learning Implementation Survey, of whom 384 were identified as being PI users. We have evidence, though, that not all PI users completed the survey, perhaps only a fraction of the total users. Several instructors sent e-mail messages describing certain aspects of their experience with PI but did not complete the actual survey. Many of these wrote to tell us that they did not consider their experiences with PI to be significant enough to warrant their participation in the survey; though we encouraged these instructors to complete the survey in any case, not all did so. Several instructors told us that they had not *yet* tried PI in their classes, but planned to do in the next or subsequent term. Finally, there were several known users of PI (as determined from personal communication, Web sites, and/or publications discussing PI) who did not complete the survey.

Therefore, the survey results cannot claim to represent the experience of *all* PI users. We cannot rule out a bias in the instructors who chose to respond to the survey, such as those who had an especially positive—or negative—experience. While the survey was designed to encourage instructors to provide feedback regardless of the effectiveness of their experience, we have no independent method of determining whether satisfied PI users responded at the same rate as dissatisfied PI users. We encouraged instructors to report both positive and negative experiences and especially encouraged those with disappointing experiences to respond; in fact, reports on less successful experiences would probably be the most informative in determining the essential elements of PI implementation. It is unclear to what degree these encouragements were successful at soliciting additional responses to the survey.

Demographic portrait of Peer Instruction users

Respondents represent a broad array of institution types across the U.S. and around the world. Survey respondents hail from 23 countries with the greatest number from the United States, Canada, and Australia (Table 2.1). About two-thirds of survey respondents teach at universities (Figure 2.1), though almost all of their PI classes are for undergraduates—usually introductory courses. An additional 19% of respondents teach at four-year colleges, *i.e.*, schools without graduate programs. It should be noted that there is likely to be some sampling bias towards faculty in institutions of higher education as the vast majority of presentations made on PI have been before audiences of college and university faculty members and where class sizes tend to be larger than at the primary and secondary levels.

Table 2.1. Demographic breakdown of survey respondents using PI based upon country of instructor. In almost all cases, the instructor's country corresponds to the country in which courses are being taught with PI (n = 384).

Country of instructor	Count
United States	320
Canada	20
Australia	11
Belgium	3
The Netherlands	3
Spain	3
Sweden	3
Colombia	2
Hong Kong	2
Scotland	2
Other (1 each*)	12

* Countries with one survey respondent each: Argentina, Chile, Cyprus, Israel, New Zealand, Peru, Philippines, Portugal, Slovenia, Taiwan, Thailand, and Venezuela.

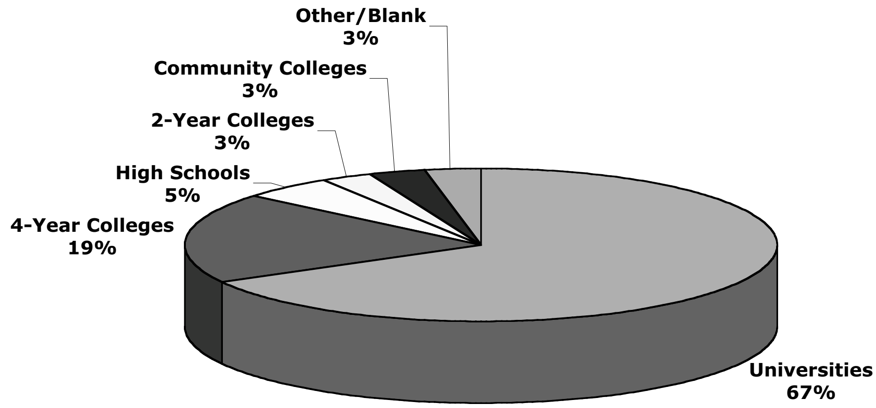


Figure 2.1. Demographic breakdown of survey respondents using PI based upon institution type as reported in question A7 in survey (n = 384).

Discipline

The vast majority of respondents use PI to teach physics, although chemistry, life sciences, engineering and astronomy courses are also represented (Figure 2.2). In particular, 82% of instructors responding to the survey teach physics, 4% chemistry, 4% life sciences, 3% engineering, 2% astronomy, 2% mathematics, and 3% other. It is not surprising that the greatest number of respondents are in physics: not only are the most materials available for physics (notably including *Peer Instruction: A User's Manual* [Mazur 1997] and the Project Galileo[†] database of ConcepTests), but the initial list of instructors contacted was also biased towards those teaching physics.

Involvement in education research

We were interested in learning how active the users of PI were in science education research and if their primary responsibilities were in teaching or research. Of those who teach at the college-level (including 2- and 4-year college, universities, and community colleges), about 10% identified themselves as primarily researchers (question

A10). Most post-secondary PI users are either equally involved in research and education (45%) or are primarily involved in education with some additional research activity (35%). Only 11% of the instructors at the college-level are solely involved in education (this number is obviously much higher at the secondary-school level). Since most instructors teaching with PI have additional responsibilities in addition to their teaching, it is important that PI be compatible with their other professional interests and responsibilities.

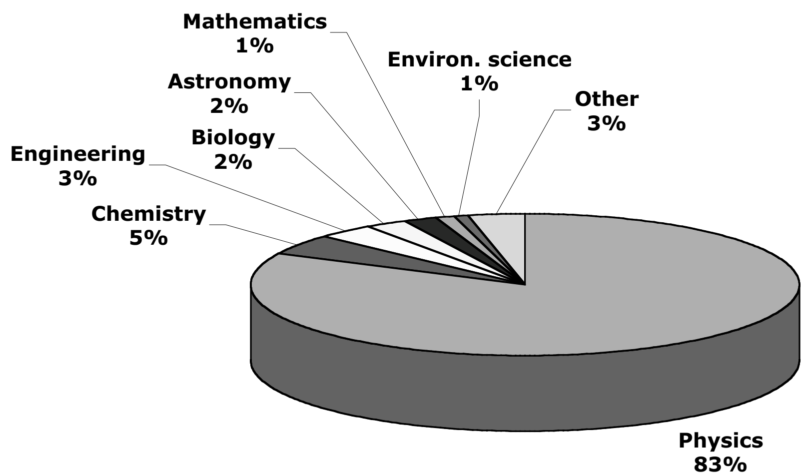


Figure 2.2. Discipline taught using PI by those responding to the survey. "Other" includes computer science, earth science, and philosophy (n = 382).

We also asked about how much these instructors kept up with the literature on science education, such as that in *The Physics Teacher*, *Journal of Chemical Education*, etc. (question A11). Four percent responded that they never read articles on science education, and 44% only do so infrequently. That leaves more than half of the respondents who keep up with the science educational literature frequently (35%) or consistently (17%). Thus, a sizable number maintain a significant interest and awareness of contemporary issues in science education, including research behind various instructional strategies.

Exposure to PI

Respondents' knowledge of PI comes from a variety of sources (Table 2.2; questions B1 and B2). Almost all users have read *Peer Instruction: A User's Manual* [Mazur 1997] at some point (93%), although this was the first exposure to PI for only 21% of instructors. The most common methods of *first* learning about PI were attending a talk or workshop (30%) or by having a conversation with colleagues (28%). It should be noted, however, that the sample population for the survey is likely to be biased on this issue since many of the respondents' names were provided by colleagues who had hosted a talk or presentation by a member of the Mazur Group or had been referred by colleagues in other ways.

Table 2.2. Respondents' familiarity with PI (or other collaborative learning strategies). Indicates how instructors first learned about PI, by source of information and the total percentage of instructors who have learned about PI from that source, initially or subsequently (*n* = 384).

	initial exposure to PI	total exposure to PI
Read *PI: A User's Manual*	21%	93%
Conversations with colleagues	28%	79%
Attended a talk or workshop	30%	62%
Read a book or article	10%	34%
Co-instructor used PI	2%	10%
Familiar with another course	< 1%	5%
TA for a course using PI	< 1%	2%
Took a course which used PI	< 1%	2%
Other	8%	19%

Course characteristics

We collected a number of pieces of information about the courses in which PI is used in addition to the discipline characterization discussed in the previous section. In colleges and universities, the majority of courses using PI are introductory undergraduate courses (90%), as has been our use of PI at Harvard. But PI is also being used (questions C2) in intermediate (5%) and advanced undergraduate courses (3%) and even in a small number of graduate-level courses (1%, or three courses). Twenty-three of the survey respondents use PI at the pre-college level, though 8 of 23 (35%) of these courses are Advanced Placement and, thus, theoretically functionally equivalent to an introductory undergraduate course.

Class sizes vary significantly with a range from 6 to 1200 students and a mean of 122 ± 178 students (question C4). Thus, it seems clear that PI is being used—and, as we will discuss below, successfully—in almost any class size, from an intimate setting to the most massive lecture. The data collected in the survey reflects the education of over 45,000 students. This significantly underestimates the number of students taught using PI, however, as respondents were asked to respond based on only their most representative or most recent *single* class. Many of the respondents had taught more than one separate course using PI and many had taught the same course with PI multiple times. It seems quite likely, therefore, that the actual number of students taught using PI is in the hundreds of thousands.

The majority of instructors were not new to their courses, and most had adapted their course to PI after teaching it in another way previously. In fact, 72% of instructors had taught the same course previously using the lecture format and 13% had used some other format for the same course (question C9). It is encouraging that even those who presumably had a set of lecture notes found the switch to PI to be worth it. It is also interesting that 20% of instructors had not taught the same course without using PI, suggesting that these were relatively new instructors who began their teaching using a research-based interactive pedagogy.

Implementation details

The survey revealed a wide range of implementations, perhaps wider than initially anticipated. At the outset, we were especially interested in investigating two characteristics that we have found especially important in the implementation of PI at Harvard—non-competitive grading in the course and pre-class reading—to see how they were implemented elsewhere. We also asked for many other details about instructors' use of PI to determine the full range of experiences with PI. As discussed below, it appears that PI is even more flexible than we had originally thought and that the necessary implementation details depend on the context of the course; characteristics we had found essential in our own classes at Harvard seem not to be so at other institutions.

Grading and competitiveness

In order to establish a non-competitive atmosphere within the classroom, we have found it necessary to adopt grading on an absolute scale—as opposed to grading on a curve—so that students do not feel as though they would negatively affect their own course performance by helping and working with their classmates. While over 59% of the respondents reported grading on an absolute scale, a significant fraction do grade on a curve (32%) or use some other grading strategy (16%) (question E1; these numbers do not add up to 100% as respondents could select more than one grading scheme).

We also asked instructors to comment upon the degree of competitiveness among the students in their class (question E2) and on how well the students worked together (question G4). Using these responses, we categorized the competitiveness in each class; among those classes for which we could make a determination (n = 263), over 65% of classes were generally non-competitive, nearly 23% were competitive, 7% had some competitive students and some non-competitive, and 11% seemed to differ from class meeting to class meeting. Regarding the degree of cooperation among the students (n = 358), about 40% of instructors said that their students worked well together throughout the entire term, an additional 17% said that student cooperation improved over time. Only 6% reported that their students did not work well together with another 7% saying that the degree of cooperation decreased over time.

Pre-class reading

In many introductory science courses, students rely upon the instructor to present *all* of the material students are responsible for during the lectures. Even basic transfer of factual material is included in the lecture so that students learn that they can be fully caught up in the class just from attending the lecture. The textbook is seen primarily as a reference that is often not consulted at all; if the textbook is read at all, it is often after the lecture on that topic has passed, such as immediately before the examination [*e.g.*, Bullock *et al.* 2002].

In our courses at Harvard, we have changed the role of both the textbook and the lecture from the way they are commonly used in a traditional lecture course. Our courses leave the basic transfer of information for students to do outside of class, freeing up the lecture period to focus on the most important and difficult elements of the readings and to provide opportunities for students to think about and discuss the ideas in class [Crouch and Mazur 2001]. As such, it is essential for students to read the textbook *before* class, both so that they receive the basic information transfer that may not happen in class itself, and so they can receive the most benefit from in-class discussions. Thus, we require pre-class reading before every class meeting, and ask students to complete a reading assignment before they come to class to provide them the appropriate incentive to actually do so. These assignments are graded based on effort—rather than the correctness of the answers—and typically account for 5-10% of a student's final course grade.

Because we found pre-class reading and related assignments to be such an essential element for success in our courses at Harvard [Crouch and Mazur 2001], we were interested to see if and how they were implemented in others' courses (questions E5-E10). In fact, over 71% of other instructors using PI report that they also expect pre-class reading from their students. In our experience, simply assigning reading may not be enough; it is also important to provide an appropriate incentive for students to actually complete the reading by way of an assignment. However, nearly 25% of those requiring reading do not have any means for assessing whether that reading had actually been completed (Table 2.3), so it is uncertain how many students are actually completing the reading before class. Along with the approximately 30% of all instructors who do not require reading at all, this means that nearly 45% of all PI users responding to the survey either do not require pre-class reading or do nothing to assess their students' compliance with any reading assignments. Because many students do not, therefore, have a requirement or incentive to read before coming to class, it does not appear that pre-class is essential for PI.

Table 2.3. Type of reading assessment for those instructors requiring pre-class reading (nearly 30% of respondents did not require such reading). Indicates the percentage of respondents who use the listed method. Respondents could choose more than one option (*n* = 274).

	% using
Multiple-choice reading quiz	52%
Free-response reading quiz	17%
Reading summary	8%
Other assessment	17%
No reading assessment	25%

Reading assessments, when presented, are most commonly in the form of a multiple-choice quiz (52% of those requiring pre-class reading) with other instructors offering free-response quizzes, reading summaries, and other types of assignments (Table 2.3; question E6). Multiple-choice quizzes obviously have the advantage that they are easy to grade so the instructor and teaching staff need not have to devote significant time to be sure that students are actually completing the assigned reading.

The reading assignments that we use at Harvard are conducted outside of class via a Web-based system in the style of Just-in-Time Teaching or JiTT [Novak *et al.* 1999]. Students generally have until sometime the night before each class to respond to three free-response questions [Crouch and Mazur 2001]. Not only does this system make it easy to administer and grade the assignments—as they are then integrated with other features of our course Web site—it allows the instructor to have access to students' answers immediately, so that the instructor can use student responses to prepare for class. Among survey respondents who present reading assignments, they are most commonly administered prior to class (54%) with a sizable number (38%) of assessments conducted during class (question E8). Only 2% of instructors who assessed student reading administered this assessment after the class for which the reading was intended. Even when performed outside of class, only 19% of reading assignments use Web technology (question E7); since the survey was conducted in 1999 when interactive course Web sites were still uncommon, it seems likely that the frequency with which such assignments are conducted online has increased.

In contrast to the way in which we view the reading assignments in our classes at Harvard, the majority of instructors using them elsewhere assign students a grade based on the content and correctness of their answers (Table 2.4; question E9). Approximately one-third of instructors follow a similar grading strategy to the one we have used at Harvard. That is, they assign a grade based on the effort shown or for simple completion of the assignment (possibly in combination with other factors since respondents could select more than one option).

Table 2.4. Credit for completing reading assignments. Indicates the percentage of respondents who use the listed method in order to assign students credit for completion of reading assignments (among instructors who administer a reading assessment). Respondents could choose more than one option ($n = 207$).

	% using
Graded for content	71%
Graded for completion	17%
Graded for effort shown	16%
Graded a subset/spot-checked	11%
Not graded	16%

ConcepTests

Almost all instructors (96%) use at least some ConcepTests that are conceptual in nature, and conceptual questions make up the most frequent type of question used (Table 2.5; question D7). For these 96% of instructors who use at least some conceptual questions in their classes, conceptual ConcepTests make up an average of 72% of all questions in those classes. Factual/mathematical questions are used by 58% of respondents, but are used much less frequently than conceptual questions even in those classes (among instructors who use factual/mathematical ConcepTests, they only make up 27% of all questions used).

Table 2.5. Types of ConcepTests used. The first column represents the percentage of respondents who use that type of question. The second column displays the percentage of time that those who use that question type at all use that particular type ± standard deviation.

	% using	average usage
Conceptual	96%	72 ± 24%
Factual/mathematical	58%	27 ± 19%
Multiple choice	80%	78 ± 31%
Open-ended/free response	33%	25 ± 4%
Demo prediction	59%	19 ± 14%
Other	5%	24 ± 16%

About 80% of instructors use at least some multiple-choice ConcepTest questions, which make up about 78% of ConcepTests by these instructors (Table 2.5; question D7). Open-ended/free response ConcepTests are used by about one-third of PI instructors but, even when they are used, they only make up about one-quarter of all ConcepTest questions asked. Questions asking students to predict the outcome of classroom demonstrations are used by more than half of all instructors, but they are used sparingly (see Chapter 3 for an investigation of the effectiveness of classroom demonstrations).

The majority of ConcepTests used by respondents to the survey are written by the instructor or adapted by the instructor from other sources (Table 2.6; question D8). Nearly 92% of the instructors write some of their own questions, with those self-authored questions making upon half of all ConcepTests used in those classes. *Peer Instruction: A User's Manual* [Mazur 1997] is also a major source of ConcepTests, being used by about two-thirds of all instructors, for about half the questions used by those instructors; this number increases to over 78% when restricted to instructors who teach physics[§§]. Thirty-six percent of responding instructors obtain at least some of the ConcepTests they use from textbooks, with those questions accounting for about one-fourth of the questions they use. Other sources include colleagues (15%), the Project Galileo Web site (12%) and other Web sites (4%). (Respondents could select more than one option; see Table 2.6).

Table 2.6. Source of ConcepTests used. The first column represents the percentage of respondents who use that source of question. The second column displays the percentage of time that those who use that question source at all use that particular source ± standard deviation.

	% using	average usage
Wrote own questions	92%	50 ± 32%
PI: A User's Manual	67%	51 ± 28%
Textbooks	36%	23 ± 18%
From colleague	15%	18 ± 15%
Project Galileo	12%	25 ± 23%
Other Web site	4%	27 ± 29%
Other	10%	32 ± 26%

[§§] It might seem strange to think that *Peer Instruction: A User's Manual* [Mazur 1997] is used by instructors in other disciplines besides physics, but it is not as surprising when one considers that instructors in related fields—such as engineering, calculus, or chemistry—may be able to get *some* content from the *User's Manual* even if much of the material in the book is not applicable for their discipline.

The majority of survey respondents (59%) do not assign students credit for their participation in ConcepTests in class (Table 2.7; question E4). This disagrees with what we have found to be most effective at Harvard, where providing students an incentive to participate is useful, a finding that is also observed by Bullock *et al.* [2002]. Only about one-fourth grade students based on their participation, as we have done at Harvard. Eighteen percent grade students on the content and correctness of their answers, presumably to give students an incentive to consider the questions careful and provide thoughtful answers.

Table 2.7. Credit for participation in ConcepTests in class. Indicates the percentage of respondents who use the listed method in order to assign students credit for participation in ConcepTests.

	% using
Graded for participation	25%
Graded for content	18%
Graded a subset/spot-checked	8%
No credit given	59%

The difficulty of the questions is also quite important. As one university physics professor responded, “with my students it is very dependent upon the questions I choose to administer, and it will take me some practice to become skilled at choosing the questions.” If the questions are too easy, students will not find them challenging enough and will likely lose interest in them. If they are too difficult, not enough students will have a good enough understand that they will be able to help convince their classmates of the correct answer. In our experience, we have found that a ConcepTest is appropriately targeted if approximately 35-70% of the class gets the correct answer before discussion [Crouch and Mazur 2001]. Others do cite benefits of easier questions, though, as they can help build students' confidence and verify that the instructor's view of student understanding is correct [Meltzer and Manivannan 2002]. Of course, if students respond correctly in high numbers, the instructor may choose to skip the discussion and move directly on to the explanation or the next topic, confident of students' understanding.

We asked whether students had access to ConcepTests outside of the lecture period (questions D10-D12) as we have found that students appreciate the opportunity to refer back to ConcepTests—and explanations of the correct answers. While 61% of survey respondents do not provide outside-of-class access to ConcepTests, 14% do include the questions on a course Web site, 14% distribute them on handouts, and others place them on reserve in the library. Although we do not have quantitative data on the degree to which students take advantage of this access to ConcepTests, 31% of instructors who do provide access thought their students took advantage of this access "often" and 47% thought "sometimes" (question D12). Of the 153 respondents who do provide access to ConcepTests outside of class, 50% offer access after each class and 17% at the beginning of each class (question D11). Fourteen percent compile all the ConcepTests and make them available to students at the beginning of the term. An additional 13% percent provide access only before exams; in addition to exam preparation for content, students like being able to consult the ConcepTests "to go back in time to the day [they] learned the material" [Schlatter 2002].

Polling for student responses

As discussed in Crouch and Mazur [2001], there are several methods for polling student responses to ConcepTests, each of which has some associated advantages and disadvantages. The optimal solution is a classroom network—either wired or wireless—because it allows the instructor to gain instant accurate feedback on student responses and to have a record of individual student responses. Additionally, such a system can generally be anonymous so that students' responses are kept confidential from their classmates; this way, students are not influenced by others' responses in the initial polling before discussion. The major drawback of a classroom network is cost, as it requires a certain capital investment in order to acquire and install the necessary equipment (see Burnstein and Lederman [2003] for a review of several commercial wireless polling systems). Only 8% of survey respondents report using an electronic polling method to receive student responses (Table 2.8; question D6), though these classes tend to be a bit larger than the overall average class size (average enrollment of 208 students for those using electronic polling vs. 122 for all classes represented in the survey). It is not surprising that the larger classes are those most likely to use an electronic polling system; not only is it more difficult to use another method in such a large class, but it is also more likely that large courses would have the resources to acquire an electronic polling system.

Table 2.8. Method of polling students for answers to ConcepTests. The first column represents the percentage of respondents who use that method. The second column displays the percentage of time that those who use the method at all use that particular method ± standard deviation.

	% using	average usage
Raising hands	65%	78 ± 73%
Flashcards	32%	79 ± 30%
Paper record	14%	68 ± 35%
Electronic polling	8%	65 ± 39%
Other	20%	55 ± 33%

The most common method of collecting student response is through the raising of hands, used by 65% of respondents (Table 2.8; question D6). This is the simplest mechanism and provides instant feedback without any technological or logistical complications. The main disadvantage to this method is that there is the potential for loss of accuracy if students are subject to peer pressure effects: hesitating to raise their hand for an "unpopular" choice or likely to jump on the bandwagon for a "popular" response. These effects can be somewhat mitigated by asking students to close their eyes so that they are not as influenced by the overall class vote and likely to side with the majority, even if they would not independently choose that response [Crouch and Mazur 2001]. Even with eyes closed, though, it is likely that students would have some idea if one choice was especially popular, from the sound of many students shifting to raise their hands.

The next most common method is the use of flashcards, used by 32% of survey respondents (Table 2.8). Often, these take the form of cards that are color-coded or have a number or letter on them. This way, all students can simultaneously raise the flashcard corresponding to the multiple-choice answer they select, so that the effects of peer influence for individual responses are diminished. The instructor can quickly scan the room to get a sense of the distribution of student responses [Meltzer and Manivannan 1996]. One interesting benefit to this method is that it allows the instructor to observe the students' body language as they raise their flashcards, providing an informal sense of their confidence in the answer they select [Meltzer and Manivannan 2002]. As with hand-raising, however, flashcards do not provide a permanent record unless the instructor counts the responses. One idea that has been suggested is using a digital camera to take a quick snapshot of the classroom when the flashcards are raised so that an instructor can later go back and count the number of students selecting each response.

Scanning forms or another type of paper record can be used to supplement the flashcard or hand-raising method. In fact, 14% of survey respondents use a paper record, often in conjunction with raising hands. Using scanning forms in isolation does not provide immediate feedback to the instructor on student responses, but it can be a very useful supplement, providing data on student attendance and understanding, as well as short-term effects of the ConcepTests, tied to individual students [Crouch and Mazur 2001].

In general, the survey results show that PI can be used successfully regardless of the feedback method and, therefore, independent of resources. The fact that most users collect student responses by the most low-technology solution—raised hands—suggests that even this method is sufficient for effective use of PI.

Student mastery

Over 108 PI users responding to the survey reported collecting quantitative data on the effectiveness of PI, of whom 81% administered the Force Concept Inventory (FCI) [Hestenes *et al.* 1992; Mazur 1997] to their students. The FCI is a multiple-choice test that assesses student understanding of introductory mechanics, written using non-technical languages so that can be given as a pretest before instruction as well as a post-test at the end of a semester. See Chapter 4 for additional discussion of the FCI and the usefulness of such an instrument for assessing student conceptual understanding.

Instructors at 11 colleges and universities provided us with matched sets of pre- and post-test FCI data, to assess the gain for individual students. In order to compare FCI scores across classes with different pretest scores, we determined the average normalized gain for each course

$$g = \frac{S_f - S_i}{1 - S_i} \quad \text{where } S_i, S_f = \text{initial, final score}$$

that has been shown to have only a very low correlation with pretest scores [Hake 1998]. The 30 courses taught with PI for which we received FCI data have a class average gain of 0.39 ± 0.09 (Figure 2.3). In his survey of FCI data, Hake [1998] defines a "medium-*g*" range from $g = 0.3$ to 0.7 and finds that 85% of the interactive engagement courses included in his survey—and none of the traditionally taught courses—show gains in this range. We find that 27 of 30 (90%) PI courses in our survey fall in the medium-*g* range, with only three below $g = 0.3$ (Figure 2.3). Data from Harvard courses are not included, as they have been reported separately elsewhere [Crouch and Mazur 2001; Mazur 1997], and the aim of this study was to investigate implementations of PI by instructors other than the developers.

More qualitatively, instructors have the sense that student command of the material has improved with PI—or at least it is not any worse than using a traditional pedagogy. Based upon instructors' responses from which it was possible to determine the instructor's assessment of student mastery, 60% state that student mastery has improved with PI (question F3). An additional 20% claim that student mastery improves somewhat and 19% think it is no different from traditional lectures. Only 2% of instructors report that their students' mastery decreases with PI. While it is disturbing to think that student learning suffers from PI, even in a small number of cases, it should be noted that none of these instructors provided nor collects any data.

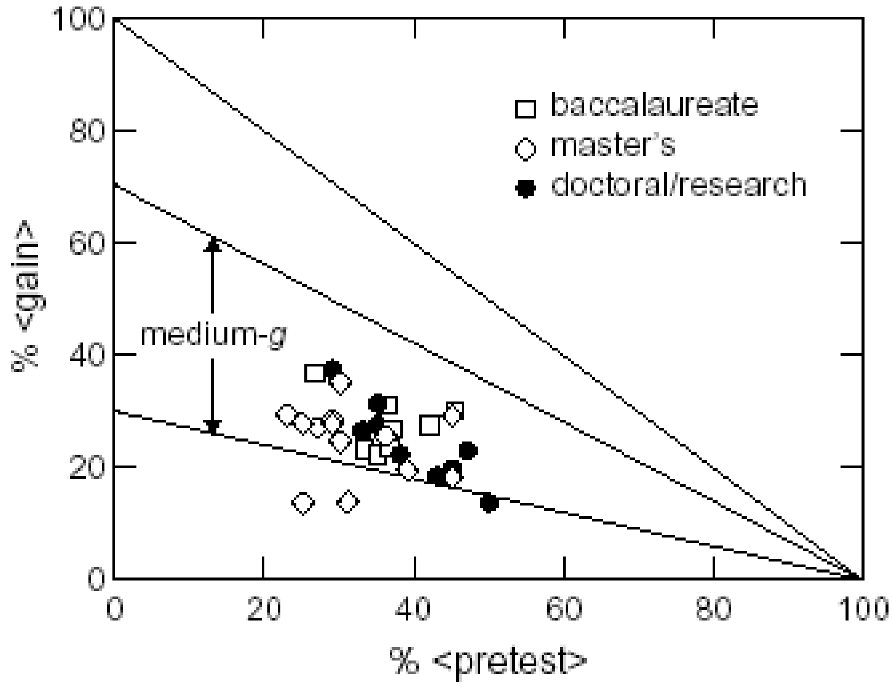


Figure 2.3. Class-averaged normalized gain of introductory physics students in college courses taught using Peer Instruction. Symbols denote institution type according to the 2000 Carnegie Classification of Institutes of Higher Education. Data from Harvard courses is *not* included and has been reported separately [Crouch 1998; Crouch and Mazur 2001; Mazur 1997].

Survey respondents felt that PI especially helped students who were not at the top of the class. One high school teacher said that there was "little change for the best students, [but] significant improvement for the weaker students." Another respondent thought that "I think it helped the weakest students most. The good students do well regardless. I have found fewer complete disasters." Similarly, "the best students did about as well as usual. It was the students of average and below average ability who showed the most improvement with Peer Instruction." Or, more quantitative—but still anecdotal—from a community college professor: "in General Physics, the only 'before and after' I have checked (once): the A and B categories showed little change but the number of students who had received F or D as a course grade decreased substantially, many moving into the C category."

It should be noted that a small number of instructors did find PI to be detrimental to strong students, even if they found it beneficial for weaker ones. A university physics professor said that "strong students learned less, weaker students understood more."

Instructor satisfaction

We asked instructors if they found their use of PI to be valuable and/or enjoyable (question G2). The vast majority of respondents reported a positive experience (93%) with less than 1% reporting a negative experience (an additional 5% had a mixed reaction). One chemist summarized his experience with PI by saying: "I'm convinced more is learned and retained. Student attendance is way up, attrition way down (almost zero), and attitude far more positive." A community college mathematics professor said that "students were more successful in the Peer Instruction class. They did more work and complained less and the average grade was higher than the traditional class."

To determine whether instructors consider the use of PI in their classes to be successful, we asked them if they were likely to use PI again in the future (Figure 2.4; question G1). Of the 384 identified PI users responding to the survey, 303 (79%) definitely planned to use PI again and 29 (8%) probably would. Only 7 respondents (2%) expressed no plans to use PI again. Thus it seems clear that the vast majority of instructors completing our survey consider their experiences with PI to be successful. It is worth repeating the possibility that these responses may not accurately represent the relative incidence of positive and negative experiences, as there may be a selection bias in who chose to complete the survey (see above). However, the responses do indicate that PI has been successfully implemented in a large number of classrooms and with a wide range of implementation details.

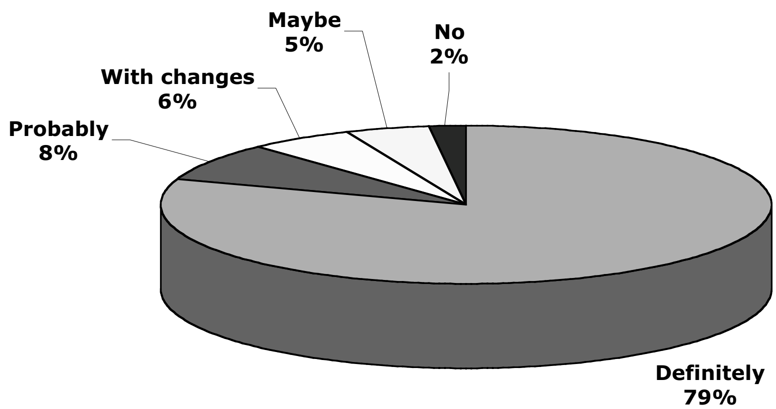


Figure 2.4. Indication of instructor's plans to use PI in a future class (n = 384).

Student satisfaction

The survey respondents indicate that students are generally satisfied with PI as well, with 70% of instructors reporting positive student evaluation and an additional 3% finding no difference from a traditional course. One high school instructor said that students became "much more enthusiastic about the material. I had some who begged for more questions!" A college physics professor said that "clearly, enthusiasm for learning physics had improved" when he adopted PI. Another reported that her teaching rating rose a full point (on an 8-point scale) when she introduced ConcepTests: "some students specifically noted on their course evaluations that they enjoyed the conceptual questions, that they made lecture more enjoyable and interesting." Or this comment from an astronomy professor at an elite university:

> I have considerable data concerning students ATTITUDES, from surveys performed by a science educator. Their attitudes typically dramatically improved. My students, who are bright, often commented on the pre (first week of class) survey that they "could do science, but it was boring; mostly memorization." They found the way science was presented, including the peer discussions and questions, much more open ended and thought-provoking. This increased their interest level substantially…. (It is worth stating that in this class for future lawyers, politicians, poets, etc. affecting the students' concept of science, understanding of it, and attitude towards it is one of my top goals.)

Another university physics professor reported that while he observed a similar performance on exams, student satisfaction was much higher: "the students almost universally enjoyed the class more and felt they understood more of what was being taught and, importantly, WHY it was being taught, *i.e.*, they saw the relevance more. The enjoyment level has been significantly higher (for them and for me!)."

Seventeen percent of instructors report a mixed student response: cases in which a significant number of students report positive reactions while others report negative experiences. Approximately 5% of instructors reported that their students had a negative response to PI, although several of these faculty members mentioned that this negative response was actually quite positive from an educational point-of-view. As one university physics respondent reported, students explained that they thought PI "was very useful, thought they did not necessarily enjoy it." That is, students complained about having to think and actively participate in class that they are not accustomed to doing. From an instructional point-of-view, this type of reaction is actually quite positive and not something to be avoided. A 4-year college physics professor reported that "there was a lot of resistance to the idea [of PI]. They just wanted me to tell them the answer. Even at the end of the semester, after a test where groups that were working well together leaped way up on test performance, they still resisted trying to think for themselves." A university biochemistry professor described student comments in a required course:

> My favorite negative comment is: "I had to do all the learning myself." This comment is both amusing and depressing because it is a sign of success in pedagogy but failure in conveying a philosophy of learning. I do have a series of orientation exercises I carry out on the first day of class to demonstrate why I have adopted this teaching technique. Nevertheless, there still are some students (fewer, as I get better at this!) who feel cheated; they feel like they have paid tuition for a PhD professor to teach. They look at me and ask, "Why aren't you teaching?" or: "Why do I have to learn from this jerk who doesn't know anything, when you are paid to teach and know it all?"

This speaks to the importance of the instructor's properly motivating the method to students at the beginning of the course and throughout the semester. This includes explaining the reasons for teaching with such a pedagogy [Crouch and Mazur 2001], possibly including presenting data on the effectiveness of PI, as compared to a traditional pedagogy. Instructors must help prime students for their active participation; since students' expectations about their education influence the degree of student involvement and, consequently, student learning [van Dijk *et al*. 2001], selling students on the method of instruction will also likely contribute to enhanced learning. It is also important to start the use of PI at the very beginning of the semester, before students have come to expect that the course will be taught as a traditional lecture class [Meltzer and Manivannan 2002].

Along these lines, about 4% of instructors report that their students had an overall positive view of PI but had been dissatisfied with the method initially; these students were skeptical of PI at the beginning of the term, but eventually warmed up to it. This is also not surprising, as one would expect students to be resistant to *any* different pedagogy because it differs from the way they are used to being taught [Felder and Brent 1996;

Kloss 1994; Woods 1994]. Resistance to change is not unusual, but as students grew more accustomed to the method, they came to realize how it was helping them to understand the material better than a traditional passive lecture pedagogy. This suggests the need for instructors to stick with the method even if students are initially suspicious of a different pedagogy. In contrast, a mere 1% of courses had student evaluations that were initially positive, but became negative over the course of the term.

Some respondents to the survey even suggested that students may have a hard time going back to traditional lectures after taking a course taught using PI. In fact, students get so used to being active learners that they maintain the inclination to discuss things in class, even when the class is not taught interactively. One university physics professor noted a colleague "who teaches the second term comments that my students [who had been taught using PI in the first term] were much more vocal in his class." Students also have come to appreciate how use of PI enhances their own learning and wish more classes were taught using an interactive pedagogy; a physics professor from Taiwan reported that while students "expressed their appreciation to this teaching style [PI], they also strongly criticized the traditional didactic teaching. My students expressed strong reluctance to be shifted back to [a] traditional class. They even encouraged me to 'teach' other faculties to adopt this teaching method."

Student retention

One consistent message from survey respondents was that the number of students who drop their course was significantly lower in classes taught using PI than in classes taught using a more traditional pedagogy. This is consistent with data from other studies, such as that of Williamson and Rowe [2002] who report a 33.3% withdrawal rate from a traditional (control) class but only 17.3% from the treatment section in which lecture was replaced with cooperative group problem-solving. Instructors also remarked that student performance, overall, was better when using PI. One noted a "dramatic reduction in the number of students who failed the course (by a factor of three.) Perhaps PI does not allow students to get behind or to fool themselves into thinking that they understand the material better than they actually do; as such, students may take corrective action before they get too far behind.

Class environment

The class learning environment changes when PI is adopted. This is again consistent with Williamson and Rowe [2002] who report that courses taught using an interactive pedagogy display better communication between instructor and students and increased tendency for students to ask questions in class. A survey respondent teaching physics at a Southern university reported that "students talked a lot more during class and asked many more questions than they usually do, even when we were not doing the [PI] discussions." Other studies have also found that students in interactively-taught classes are more likely to avail themselves of instructor office hours [Williamson and Rowe 2002], but we did not collect data on this characteristic in our survey.

Student confidence improved along with their knowledge of physics. A faculty member from Belgium found "most remarkable was the increased self-confidence after the start-up period." One respondent reported that students' attitude "became more enthusiastic and positive toward physics and not just a terrifying subject." Another respondent described a student who he described as a "late-blooming student" who had decided to pursue a summer research program: "I think this was in part due to seeing that she did indeed know how to think, partly as a result of the ConcepTest questions." A physics professor at a 4-year college concluded that "the most important difference seemed to be the confidence the students exhibited in their own knowledge and skills. Peer Instruction seemed to provide a much greater sense of mastery."

Undergraduate science courses often have the reputation of being impersonal, passive lecture courses. Students are often motivated almost exclusively by the grade and will compete against each other in order to score well on the curved grading scale. The atmosphere in the classroom and set of expectations is an important factor. Encouraging students to work together—especially when grading on non-competitive absolute grading scale—can make a class more friendly, perhaps especially to those in underrepresented groups [Ege *et al*. 1997].

Challenges and difficulties

Many successful users of Peer Instruction indicate that they had to overcome a number of challenges, which we describe here along with solutions suggested by the respondents (question G6). Thirteen percent of instructors cite the time and energy required to develop ConcepTests as an impediment to using PI. Developing good ConcepTests that are well-suited to the level of the class, that address the key issues in the material, and that successfully engage students takes a great deal of effort. To minimize duplication of this effort, and to make PI easier to implement, we and other developers of ConcepTests have made online databases of class-tested ConcepTests for introductory physics, chemistry, and astronomy freely available and more questions are always being developed for these and other fields. Consequently, for many courses, ConcepTest development need not be a major obstacle, although instructors must still be vigilant about selection questions appropriate for their students.

Politics

Ten percent of respondents report that their colleagues are skeptical of the benefit of student discussions, especially as they take away time for lecturing. A third of instructors citing this type of difficulty report addressing this skepticism by collecting data on student learning gains. One particularly effective approach is to compare achievement of students taught with and without PI on identical exams. Others suggest inviting skeptical colleagues to sit in on a class, sharing positive students feedback with them or even giving the assessment tests to other faculty.

Coverage of material

About 9% of respondents report that the quantity of material to cover in a semester often makes it difficult to devote class time to ConcepTests. One-tenth of these instructors are able to reduce the amount of material covered by the course, but the majority do not have the freedom to do so. One option for those bound by a lengthy syllabus is to require students to learn some of the material on their own, especially by assigning reading of the text *before* class (as discussed above).

Even though instructors may not be able to cover as much material in a PI class as in a traditional lecture, many do not see this as a problem: “we covered perhaps 15% less material. That didn't bother me." According to one Canadian physics instructor, "my students covered significantly less material but achieved much greater mastery." This result is also observed by those in other disciplines. For instance, Udovic *et al.* [2002] found that students in a conceptual-based Workshop Biology curriculum perform as well as students in a traditional group, even though the conceptual group was actually exposed to less content. Interestingly, not all instructors felt that PI limited the amount of material they could cover. One university physics professor said that he "felt that students' mastery or conceptual information was much improved, as were able to cover MORE material."

There is research to suggest that students perform better when they are responsible for lesser amounts of material [Sadler 2000]. For instance, the Third International Mathematics and Science Study (TIMSS) showed that students score higher on international tests when there is reduced coverage in their classes [Tamir 2000]. Sadler and Tai [1997] similarly found evidence that college students who had taken high school courses covering fewer topics were more successful in college physics courses. In fact, one might ask what is the point behind encyclopedic courses that cover many topics briefly if students do not learn the majority of the material. Sadler [2000] also considers this type of approach, questioning the usefulness of moving on to new topics when students have not yet mastered the fundamental ideas from an earlier topic, ideas that provide the necessary background for the new topic.

Motivating students

Some respondents cited the difficulty in fully engaging students in class discussions (7% of respondents). In the words of one instructor, "some students were too cool, too alienated, or perhaps too lost to participate." Nearly half of those citing this challenge say it is important for the instructor to circulate through the classroom during the group discussion of the ConcepTest, helping to guide and encourage students in discussion. Liebman [1996] also suggests explicitly asking non-participating students to join a nearby discussion. To be sure, this instructor presence is easier to achieve if there are multiple members of the teaching staff who can circulate through the room to encourage participation. In fact, at the college or university level, 21% of survey respondents teach with a co-instructor and over one-third of classes have Teaching Assistants (TAs) as part of the course staff, especially in larger classes.

If TAs are involved in teaching the course, it is important than they be motivated and sold on the method as well as the students. For instance, Crouch and Mazur [2001] suggest that TAs have the opportunity to consider and discuss ConcepTests that *they* find challenging, so that they can experience the benefits of PI as well. TAs also see the difficulties and misunderstandings of the students by being present in lecture and helping to facilitate student discussion of the ConcepTests. And it may be that the students' enthusiasm for in-class discussion can rub off on the TA and help convince him or her of the value of such interactivity.

Because students are generally only directly interacting with those sitting nearby, the hesitancy preventing shy students from participating is diminished. Students are likely and welcome to sit with their friends or others with whom they are comfortable expressing their confusion. This helps to lessen the sense of isolation that students often feel in such large-enrollment science courses [Seymour and Hewitt 1996]. One biochemist responding to the survey said that PI "provides a way to interact with the students who normally would not speak up in a traditional classroom (the overwhelming majority of students)." Thus, it encourages *all* students to be active participants in the class, not just those whose personality encourages them to speak up in class.

It is important to convince students that the instructor values conceptual questions and considers them important parts of the course. This can be best achieved by assigning credit for participation in ConcepTests and by the presence of ConcepTest-like conceptual questions on examinations [Tobias 1998; Tobias and Raphael 1997]. As discussed in Chapter 1, the content and form of assessments define for the students what the instructor really finds important. Therefore, it is essential that examinations and other forms of assessment accurately reflect the conceptual focus of the course.

Our results from Harvard indicate that students' performance on traditional quantitative problem-solving is at least as good in a course taught with PI as one taught traditionally with many problems worked as part of the lecture [Crouch and Mazur 2001; Mazur 1997]. We believe that focusing on concepts makes it easier for students to actually understand the equations and apply then accurately, rather than merely following a "plug and chug" procedure that leads to technically correct answers without much thought. Similar results are also observed elsewhere. A physics professor at a 4-year college, for instance, observed that "students gain more secure control of the concepts, and they do just as well in the traditional number-crunching problems." It should be noted that this finding may not be universal; one university physics professor reported that, in a self-described "limited initial implementation of PI…conceptual understanding increased compared to traditional lectures, BUT…ability to solve textbook problems declined." We recommend maintaining traditional problem-solving as an important part of the course, but moving most of the problems out of the lecture; in our courses, students have ample opportunity to practice problem-solving in the weekly workshop sections and on regular problem sets.

Conclusion and future directions

The results of the Peer Instruction/Collaborative Learning Implementation Survey indicate that PI is successfully used in a wide variety of settings and with a range of implementations. In fact, the survey responses do not indicate any clear common mode of implementation among those who report success. Furthermore, there are no obvious implementation differences between those who report success and those few who report mixed or discouraging experiences.

One result of the survey, however, does stand out: instructors who have difficulty engaging their students in discussions of ConcepTests are more likely to express their PI experiences as only partly successful. This finding strongly suggests that student motivation and the quality of discussion are important factors in the success of PI. Specifically, discussion among peers leads to enhanced learning for the whole class if—and perhaps only if—most members of the class are actively engaged in the discussion by giving explanations or asking thoughtful questions.

We have already initiated a focused study to closely investigate PI at a variety of institutions to examine what is required for successful implementation of PI by assessing the specific implementation parameters and gains in student learning in a small number of PI classes. To this end we will work with a group of collaborators from diverse settings to assess the different implementation models used by the instructors. This group will include settings where PI has been successful as well as settings where PI has met with some challenges. To determine success, we propose to assess student mastery of subject material and student and instructor satisfaction with the course.

Chapter 3:

Enhancing the effectiveness of classroom demonstrations

On page 3 of a pamphlet published by the Royal Institution, London, entitled *Advice to a Lecturer*[*], Michael Faraday argues that "a lecturer should exert his utmost efforts to gain completely the mind and attention of his audience, and irresistibly make them join in his ideas to the end of the subject" (quoted in [Holton 1970]). One effective way to make science come alive and be active in a large lecture is through the classroom demonstration. Student evaluations suggest that demonstrations do serve to entertain and involve students in the lecture; one study found demonstrations to be among students' favorite elements of introductory undergraduate physics courses [Di Stefano 1996]. However, there is less evidence indicating that lecture demonstrations help students understand the scientific principles underlying the demonstration. While instructors and students alike claim that students learn from demonstrations [*e.g.*, Freier 1981; Hilton

[*] The pamphlet is an anthology largely taken from Bence Jones' 1869 *The Life and Letters of Faraday* [Holton 1970]).

1981; Schilling 1959], there is little actual data to support this claim [Gattis and Park 1997].

Indeed, there is reason to believe that in-class demonstrations of scientific phenomena do little to actually enhance understanding of underlying principles—that they do not actually *demonstrate* much at all. Halloun and Hestenes [1985] observe that traditional demonstrations do not necessarily help students to recognize and correct scientific misconceptions. Observation is, in fact, not a simple practice; as Norris [1985] described, making observations is “an extremely complex activity, indeed among the most challenging enterprises in which human beings engage.” Students observe what they *expect* to observe, which is not necessarily the same as what is *actually* demonstrated [Gunstone and White 1981; Hynd *et al*. 1994]. Tobias [1992b] noticed a similar phenomenon when she asked faculty members in non-science disciplines to sit in on introductory physics courses:

> Here is the tragedy. The professor is doing a demonstration to clarify the subject. But, in fact, it confused this student even more. From a professor at Indiana, sitting in on the regular calculus-based [physics] course came another set of observations: ‘If you don’t know what you’re supposed to be seeing, you don’t see. The block didn’t move at a constant speed along the air track. The air track was too short.’ Later, she became so shy of her own sensory impressions, that she decided she could not trust them. ‘The same was true of the so-called falling objects. We were to hear them fall at the same moment, but one of them bounced, and the professor never discussed the physics of the bounce.’ She concluded profoundly ‘physics is a process of selectively ignoring.’ [Tobias 1992b, quoted in Kraus 1997]

A number of research studies provide evidence that demonstrations do not help students to understand the phenomena that are being demonstrated. Pamela Kraus’ [1997] findings indicate that students shown a demonstration can later *describe* the events better than those who were not shown the demonstration, yet their *understanding* of the physical concepts is no better. Gattis and Park [1997] report that student misconceptions may actually be reinforced by demonstrations, rather than being corrected. Kraus [1997] investigated students' experience with the classic "shoot-the-monkey" demonstration: a monkey is suspended from an electromagnet; when a gun aimed at the monkey discharges, the circuit to the electromagnet is cut, releasing the monkey at the instant that the gun is fired. Students generally felt comfortable with the demonstration upon seeing it, but many had difficulty answering questions related to the demonstration. As one student related, "you understand the specific example but you don't understand what happens in any other case involving that example" [Kraus 1997].

Perhaps most shocking among Kraus' results is that some students who see a demonstration have retained an *incorrect* memory of the outcome of the demonstration. That is, students may remember an outcome that did *not* actually occur, especially if that inaccurate outcome is consistent with their internal model of the situation. This seems especially true in situations when the demonstration does not have the outcome expected by the student. It is likely that students may alter their memory of the outcome rather than alter their model of the phenomenon to take the new observation into account. Halloun and Hestenes [1985] describe a similar phenomenon in which students claim that the demonstration they observed is somehow a special case while their model holds true for the general scenario:

> As a rule, students held firm to mistaken beliefs even when confronted with phenomena that contradicted those beliefs. When a contradiction was recognized or pointed out, they tended at first not to question their own beliefs, but to argue that the observed instance was governed by some other law or principle and the principle they were using applied to a slightly different case. [Halloun and Hestenes 1985]

This observation would not surprise psychologists. Current understanding is that memory is reconstructed at the moment of recall, rather than being stored as a continuous visualization like a video replay [Gray 1999]. Gaps in the memory are filled in with information derived from mental models (schemas and scripts) rather that by additional factual memories. Thus, having an incorrect mental model for a situation can lead to an inaccurate memory of the scenario that is inconsistent with what actually occurred.

Objective for study

We sought to confirm the results of Kraus [1997] with a wider variety of demonstrations and to determine whether pedagogy influences student learning from classroom demonstrations. Does student learning from demonstrations depend on the pedagogy with which the demonstration was presented? In particular, does engaging students in the demonstration by asking for them to make predictions about the expected outcome beforehand enhance learning over a traditionally passive presentation? Could learning be further enhanced by asking students to discuss the demonstration in addition to making predictions?

The courses: Physics 1a and 1b

Physics 1a and 1b are the two semesters of a year-long introductory physics sequence at Harvard University covering basic topics of mechanics (1a) and electricity and magnetism (1b). The enrollment for Physics 1a for the Fall 2000 semester was approximately 125 students and, for 1b in Spring 2002, approximately 100 students. The two courses are mainly taken by undergraduates majoring in biology, history and science, or other non-science fields. In both courses, the majority of students are preparing for medical school and many are taking 1a and 1b to fulfill a physics requirement for their major. Students in the physical sciences—other than physics (including chemistry, biochemical sciences, and engineering)—and/or those with Advanced Placement credit are required to take the Physics 11 series instead of the Physics 1 series. Prospective physics majors are required to take a separate three-semester introductory sequence, the Physics 15 series.

During the Fall 2000 and Spring 2002 semesters discussed in this thesis, the instructor was Professor Eric Mazur. Instruction involved two 90-minute lecture periods per week taught using the interactive pedagogy of Peer Instruction: Mazur interspersed short conceptual questions, called ConcepTests, throughout the lecture in order to actively engage students with the material (see Chapter 2 for a thorough discussion of Peer Instruction). Some ConcepTests were associated with demonstrations conducted in the lecture, but these were not part of the study described in this chapter. Students responded to the ConcepTest using a Personal Response System (PRS) transmitter (Varitronix, Ltd., Hong Kong) to register their answer to the multiple-choice question posed.

In addition to lecture, each student participated in a weekly workshop section of about 20-25 students that were led by two graduate Teaching Fellows[†]. Physics 1a had seven different workshop sections, while Physics 1b had five. Although section attendance was not required, students were strongly encouraged to take advantage of the opportunity to work through tutorial worksheets and solve homework problems with

---

[†] Physics 1b also included an independent biweekly laboratory section.

classmates and Teaching Fellows available for discussion and consultation. Attendance was generally high at the beginning of the semester but waned as student work in other classes increased, dipping to about 50% by the end of the term.

The first hour of two-hour workshop sections was devoted to working in small groups through the applicable section of *Tutorials in Introductory Physics* [McDermott *et al*. 1998, 2002]. These research-based worksheets focus "on the development of important physical concepts and scientific reasoning skills" and are meant as a supplement to traditional textbooks and lectures. They are designed to guide students, socratically, through the development of key concepts, revealing misconceptions in the process. The tutorials guide students through the reasoning necessary to construct concepts and to apply them to real-world situations and to give students practice in moving back and forth between various representations of physical concept (*e.g.*, formulas, graphs, diagrams, verbal descriptions). Students work in small groups on the tutorials with discussion an important part of the learning process.

The second hour of the workshop section focused on problem-solving, and students were encouraged to work collaboratively on that week's assigned problem set and the homework associated with the tutorials. Between the tutorial hour and the problem-solving hour, one of the Teaching Fellows presented a demonstration and worked a problem at the blackboard. These demonstrations are the subject of this chapter.

Research methodology

This study was inspired by the work of Pamela Kraus at the University of Washington, whose dissertation focused on learning from demonstrations and tutorials [Kraus 1997]. In our research group, investigations of learning from demonstrations in an experimental sense—with multiple parallel sections—was initiated by J. Paul Callan in Physics 1a in the Fall of 1998 using an early version of the methodology described here [Callan *et al*. 2000][‡].

Throughout each semester, demonstrations were presented weekly to different workshop sections. Each week, the same demonstration was presented to the different workshop sections, but the mode of presentation varied. Four different modes of presentation were used:

(1) *no demonstration*, in which the demonstration was not presented (control);

(2) *observe*, the traditional approach to demonstrations, in which students watch the demonstration and hear the instructor's explanation;

(3) *predict*, in which students record their predictions of the demonstration outcome, observe the demonstration, and hear the explanation; and

(4) *discuss*, in which students record predictions, observe the demonstration, discuss the demonstration with fellow students, and finally hear the explanation; students were given a worksheet, which they were welcome to take with them after section, to guide them through the steps of the mode.

[‡] The research described here improved on Callan's study by recording student attendance in the workshops and, thus, making it possible to reliably ascertain which students had seen which demonstrations in which presentation methods. Even though Callan's data is not statistically significant—since incomplete attendance data diminished the experimental effects—the qualitative patterns observed by Callan *et al*. [2000] are consistent with the work discussed in this chapter.

As listed here, each mode incorporated additional activities and an increasing degree of interactivity over the ones above, but each student only experienced one of the four modes for each demonstration. Appendix B includes the instructions on each mode provided to the teaching staff as well the log form used to record details about the demonstration, and a sample viewgraph and worksheet.

The assignment of modes varied from week to week with the modes rotating through the sections. For instance, one section might feature one demonstration in *observe* mode one week and a separate demonstration in *predict* mode the next week, while another section had the first demonstration in *predict* mode and the second in *discuss* mode. This rotation allowed us to account for any differences in student ability between sections. Over the entire semester, each student participated in each mode about the same number of times.

The *discuss* mode follows the approach of *elicit, confront, resolve* espoused by McDermott [1990] in which students are directed to see and recognize their own difficulty and then helped to resolve the conflict in thinking. The *predict* and *discuss* modes also incorporate elements of Peer Instruction, in particular the use of a ConcepTest for soliciting students' predictions, but without peer discussion in the *predict* mode.

The *no demo* control group included students actually attending a *no demo* section, as well as students who did not attend a workshop section that week. Midway through each semester, we stopped scheduling an explicit *no demo* section as section attendance waned as the semester progressed (as the semester progressed, attendance dropped to approximately half of the enrollment in the course). This allowed us to maximize data from the experimental modes by having all students who attended a section be included in the *observe*, *predict*, or *discuss* modes. To determine whether students who routinely skipped section were representative of the class as a whole, we computed each student's grade on the three hourly exams and the final examination combined. The overall course performance of students in the *no demo* group was virtually indistinguishable from students in any of the other modes of presentation (in fact, for Physics 1b, the exam grade for the *no demo* group was midway between that for the *predict* and *discuss* groups).

For students in the *predict* and *discuss* modes, predictions were recorded with the use of the PRS system. The TF showed the equipment and described the demonstration to be performed. Then a ConcepTest asking a question based upon the outcome of the demonstration was displayed as a viewgraph, which also included a multiple-choice list of potential answers to the ConcepTest. (The ConcepTest generally just asked for the outcome of the demonstration, but some of the answer choices included information about the reason for the outcome.) Students chose the option that most closely corresponded to their prediction and reported that answer through the PRS system. As each PRS transmitter has a unique ID that identifies the student supplying the answer, collecting responses through the PRS system also allowed us to collect attendance data simultaneously.

For the *observe* mode, the PRS was not used (as there were no questions to which the students were asked to respond). In this case, the Teaching Fellows discreetly recorded attendance for the section (as section attendance was not mandatory, we could not take explicit attendance).

Assessment of demonstrations

We assessed student learning from the demonstrations at the end of the semester with a free-response test (Physics 1a test in Appendix C; 1b test in Appendix D). The test was administered as a Web-based assignment during the reading period between the end of classes and the start of the final examination period. The test was worth as much credit as a problem set and was graded based on the effort displayed, rather than on the correctness of the answers. Any answer of more than a few words received full credit and was similar to our grading of reading assignments.

The test presented physical situations identical to the demonstrations, without reference to the fact that they had been shown in class. We asked students to predict the outcome of the situation and explain the reason for their prediction. Several follow-up questions were designed to help reveal whether students understood the underlying physics. We classified the responses separately by outcome (correct or incorrect) and by explanation (correct or incorrect). The complete end-of-semester tests can be found in Appendices C and D.

Students completed the assignment at their convenience during a period of several days. They were instructed not to discuss the test with others or to use reference materials such as the text, their notes, and other printed and online sources. In truth, most of these materials would have been unlikely to be of much help; we were primarily concerned that students not use the *discuss* mode worksheets as these would provide a written record of the demonstration outcome as well as the students' comments about the explanation for the observation. Since the test questions were primarily conceptual, computation was generally not necessary, although we did permit the use of a calculator.

For Physics 1a, we provided students with an electronic copy of the chapter summaries for each of the relevant chapters of the draft text we were using; this allowed them to consult any equations or definitions that they might need (especially relevant since we allowed students to bring the text and an equation sheet when taking in-class exams). Chapter summaries were not provided for Physics 1b.

Caveats

The main interest in classroom demonstrations comes from their use in the lectures, where they are most commonly used. We made use of the pre-existing workshop sections in order to provide internal control groups within the class and because Harvard does not offer multiple lecture sections of its introductory physics courses. However, the sections do not provide an ideal model of the lecture setting. In fact, an environment where the same course includes multiple lectures, ideally all taught by the same instructor, would provide a more appropriate experimental environment for conducting such a study.

There are several differences between conducting demonstrations in a discussion section and in a lecture that should be mentioned. One is the size difference: attendance in section was generally in the range of 8-15 students (varying from as few as 3 students to as many as 25), as compared with lecture sections up to many hundreds of students. In a small section, students are much more likely to be engaged than a large lecture, just by virtue of the more intimate setting and such factors as room size, eye contact with the instructor, etc. Second—and perhaps less obvious—is the integration of the demonstration with the rest of the material being discussed. In most lecture courses, demonstrations are done as part of the presentation of material; that is, they are well-integrated with the topic being discussed. For instance, a demonstration might be presented just after an exposition of that topic is given in words. Therefore, there is little confusion on the part of students how the phenomenon being demonstrated fits in to the overall goals of the course or the general organization of material. In section, however, the demonstrations were largely presented in isolation, without a strong connection to the tutorial just completed. Although the demonstrations were picked to mirror the general topic of the workshop as closely as possible, they were conducted without any explicit or implicit connection to course concepts. Thus there were fewer contextual clues as to which concepts were most relevant to understanding the phenomenon observed.

Demonstrations used in Physics 1a

The seven demonstrations used in the mechanics course, Physics 1a, are listed below. The end-of-semester test used to assess student understanding describes each demonstration in more detail and may be found in Appendix C.

1. *Roadbed*: Driving a radio-controlled model car on a lightweight roadbed with little friction beneath roadbed and the floor, so that the roadbed moves when the car starts;
2. *Collisions*: Colliding a rubber ball and a putty ball with a bowling pin to see which ball knocks the pin over;
3. *Tension puzzler*: Comparing tension in a string when fixed to a wall at one end vs. when attached to weights at both ends;

4. *High/low road*: Comparing time of travel for balls on two tracks which have the same starting and ending points, but one track dips lower than the other in the middle;
5. *Loop-the-loop*: Model car on a track with complete vertical loop; minimum height to car to make it around loop without falling;
6. *Orbiter*: Puck revolving in a plane at the end of a string; effect of string length on the angular speed of the puck;
7. *Loaded beam*: Beam supported at each end by platform scales; effect of position of load on scale readings.

These demonstrations were selected to be relatively simple to conduct and to involve a single physical concept. They were also chosen not to overlap with demonstrations presented in lecture or tutorial activities. The loop-the-loop demonstration closely resembled a question on one of the problem sets; further discussion of this particular demonstration is provided later in this chapter

Demonstrations used in Physics 1b

The nine demonstrations used in the electricity and magnetism course, Physics 1b, are listed below. The end-of-semester test used to assess student understanding describes each demonstration in more detail and may be found in Appendix D.

1. *Two spheres*: A charged object is brought near two metal spheres that are in contact; if the spheres are then separated before the charged object is removed, what is the charge on each of the spheres?

2. *Charged cup*: A charged object is touched to the inside of a metal cup and is then removed; what is the charge on the inside and outside surfaces of the cup?
3. *Capacitor*: A dielectric material is placed between the plates of a parallel plate capacitor; how does the capacitance of the capacitor change?
4. *Series capacitor*: Two capacitors are connected in series; how does the capacitance of the two connected capacitors compare to the capacitance of just one capacitor?
5. *B-field*: Iron filings are sprinkled on a glass plate surrounding a metal wire; what happens to the filings when a current is induced in the wire?
6. *Two coils*: Two conducting coils, each containing a switch and one also containing a battery, are placed adjacent to each other; does the current induced in the coil without the battery depend upon which switch is closed first?
7. *Eddy current*: Identical magnetic rings are placed around the tops of a metal and a wooden rod of the same diameter and length and dropped simultaneously; which ring reaches the bottom of the rod first?
8. *RC circuit*: A circuit containing a charged parallel plate capacitor, a bulb, and a switch; when the switch is closed, what happens to the light bulb?
9. *Half-lens*: An object is placed in front of a converging lens; what happens to the image of the object when the top half of the lens is blocked?

These demonstrations were selected to be relatively simple to conduct and to involve a single physical concept. They were also chosen not to overlap with demonstrations presented in lecture or tutorial activities. However, the eddy current demonstration happened to be quite similar to a demonstration included in the lecture; a discussion of this demonstration is included below.

All of these demonstrations were tested at the end of the semester except for the third ("Capacitor"). Attendance and prediction data were not collected from students in the *predict* and *discuss* modes that week, due to technical problems recording from the PRS transmitters. Data and analysis for Physics 1b is, therefore, only for the eight remaining demonstrations.

Results

We analyzed student responses on the end-of-semester tests for correctness of both outcomes and explanations. "Outcome" refers to the student's ability to correctly predict the factual outcome of the demonstration when presented with the identical physical situation; this essentially corresponds to the answer to the multiple-choice ConcepTest that was posed to students in the *predict* and *discuss* sections. "Explanation" is the student's ability to correctly explain the underlying physics that produces the correct outcome; looking at explanations—in addition to outcomes—indicates that students not only remember (or can guess) what would happen in the specific scenario, but that they also understand why that result occurs.

We find that students perform better on an open-ended assessment for both predicting the outcome of analogous physical situation (Figure 3.1) and providing an appropriate explanation that uses the correct underlying physics (Figure 3.2) when the pedagogy with which the demonstration is presented in more interactive. Students' ability to provide the correct outcome is similar between the two semesters (Figure 3.1). However, students were much better at providing the correct explanations in Physics 1b than they were in Physics 1a (Figure 3.2). I will comment briefly on this difference at the end of the chapter.

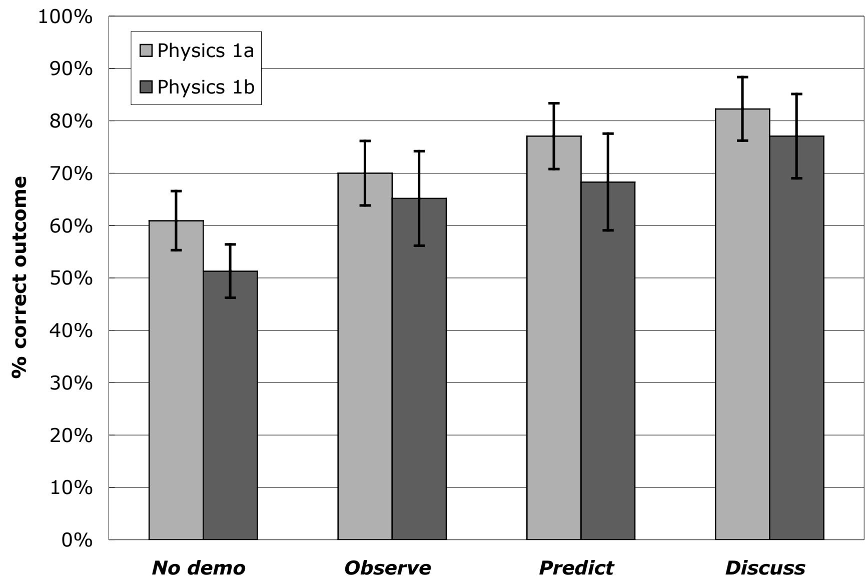


Figure 3.1. Proportion of students predicting the correct outcome by mode of presentation, for all demonstrations by semester. Error bars indicate the 95% confidence interval of the standard error of a proportion, $\pm 1.96\sqrt{\frac{p(1-p)}{n}}$.

In the following sections, results from Physics 1a are treated separately from those of Physics 1b. As the students enrolled in the two courses are not the same, and students in Physics 1a were enrolled in their first semester of college-level physics while nearly all of the students in Physics 1b had taken 1a the previous semester, it seems more appropriate to consider the two populations independently. I will revisit this issue after presentation of the results.

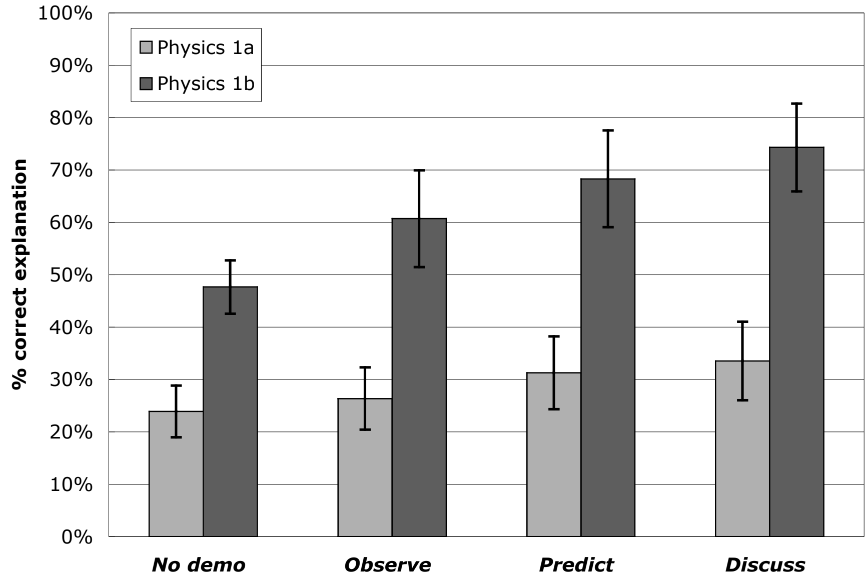


Figure 3.2. Proportion of students providing the correct explanation by mode of presentation, for all demonstrations by semester. Error bars indicate the 95% confidence interval of the standard error of a proportion, $\pm 1.96\sqrt{\frac{p(1-p)}{n}}$.

Results from Physics 1a: Mechanics

Rates of correct outcomes (Table 3.1) and correct explanations (Table 3.2) for each mode and for the *no demonstration* (control) group in the mechanics class are shown in Tables 3.1 and 3.2. Statistical significance* and normalized gain† are given in Table 3.3 for outcomes and Table 3.4 for explanations. Figure 3.3 displays the improvement in rates of correct outcomes and explanations for each mode over those rates for the *no demonstration* group, normalized to the rates for the *no demonstration* group. This graphical representation makes it easy to see the relative difference between the different

---

* The *p*-values given are the probability that the difference $D^{mode} = R^{mode} - R^{no\ demo} = 0$ (the hypothesis being tested is $R^{mode} = R^{no\ demo}$) and follow the approach outlined in Moore and McCabe [1998]. A difference is considered statistically significant if $p < 0.05$.

† I have calculated a type of normalized gain for each mode of presentation as compared to students' performance in the *no demo* group: $(R^{mode} - R^{no\ demo}) / (1 - R^{no\ demo})$ where R^{mode} represents the proportion of students providing the correct outcome or explanation for that mode.

modes, especially showing how much more effective the *predict* and *discuss* modes are, relative to the *observe* mode. Especially noteworthy is that students in the *observe* group score only marginally better on both outcomes and explanations than students in the *no demonstration* group, and the difference for explanations is not statistically significant ($p = 0.32$).

Table 3.1. Outcome responses by mode and correctness for all seven Physics 1a demonstrations, combined.

	Correct	Incorrect	Unclear	*n*
No demo	60.9%	35.7%	3.4%	297
Observe	70.0%	27.3%	2.7%	220
Predict	77.1%	20.1%	2.8%	179
Discuss	82.3%	15.2%	2.5%	158
Overall	70.6%	26.5%	2.9%	854

Table 3.2. Explanations by mode and correctness for all seven Physics 1a demonstrations, combined.

	Correct	Incorrect	Unclear	*n*
No demo	23.9%	71.7%	4.4%	297
Observe	26.4%	64.5%	9.1%	220
Predict	31.3%	60.9%	7.8%	179
Discuss	33.5%	63.9%	2.5%	158
Overall	27.9%	66.2%	6.0%	854

Table 3.3. Rates of correct responses for outcomes (R_{outcome}) by mode for all Physics 1a demonstrations, combined for all seven demonstrations; statistical significance (p_{outcome}) compared to the *no demonstration* group ($p < 0.05$ is considered statistically significant; and normalized gain (g_{outcome}) relative to the *no demonstration* group†.

	R_{outcome}	p_{outcome}	g_{outcome}	*n*
No demo	60.9%	—	—	297
Observe	70.0%	0.02	23.3%	220
Predict	77.1%	< 0.0001	41.4%	179
Discuss	82.3%	< 0.0001	54.7%	158
Overall	70.6%	—	—	854

Table 3.4. Rates of correct responses for explanations (R_{explan}) by mode for all Physics 1a demonstrations, combined for all seven demonstrations; statistical significance (p_{explan}) compared to the *no demonstration* group ($p < 0.05$ is considered statistically significant; and normalized gain ($g_{outcome}$) relative to the *no demonstration* group†.

	R_{explan}	p_{explan}	g_{explan}	n
No demo	23.9%	—	—	297
Observe	26.4%	0.32	3.3%	220
Predict	31.3%	0.02	9.7%	179
Discuss	33.5%	0.01	12.6%	158
Overall	27.9%	—	—	854

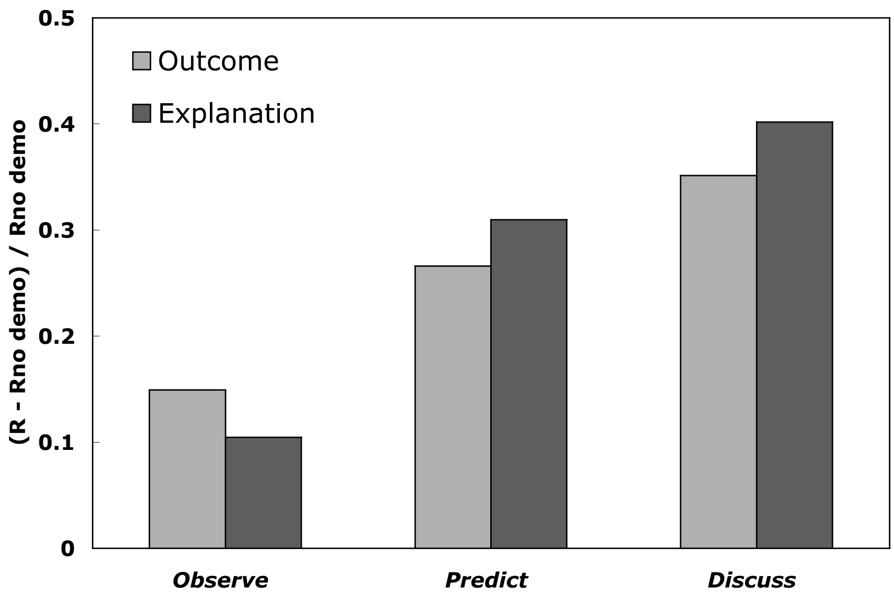


Figure 3.3. Displays the improvement in rates of correct outcomes and explanations for each mode over those rates for the *no demonstration* group, normalized to the rates for the *no demonstration* group.

Even in the *discuss* group, only one-third of students provide fully correct explanations. This low rate of correct explanations can be partly attributed to our binary scoring system and the difficulty of the questions. It may also indicate that there are limits on what students can learn from single demonstrations; many successful research-based strategies for teaching physics involve a carefully designed sequence of activities (such as the Interactive Lecture Demonstrations discussed below).

The demonstrations and related questions on the end-of-semester test differed widely in their difficulty. Table 3.5 and Figure 3.4 show the rate of predicting the correct outcome by mode of presentation for each demonstration, while Table 3.6 and Figure 3.5 show the same for providing the correct explanations.

Table 3.5. Proportion of students predicting the correct outcome, by demonstration and mode of presentation for Physics 1a demonstrations.

Demonstration	*No demo*	*Observe*	*Predict*	*Discuss*
Roadbed	52.0%	63.6%	59.3%	56.6%
Collisions	69.8%	59.5%	75.0%	73.3%
Tension puzzler	76.0%	80.0%	88.9%	100.0%
High/Low road	41.9%	62.5%	88.9%	94.7%
Loop-the-loop	25.8%	38.5%	59.3%	52.0%
Orbiter	27.9%	37.0%	48.4%	57.1%
Loaded beam	48.6%	52.8%	78.6%	76.2%
Overall	48.9%	56.3%	71.2%	72.8%

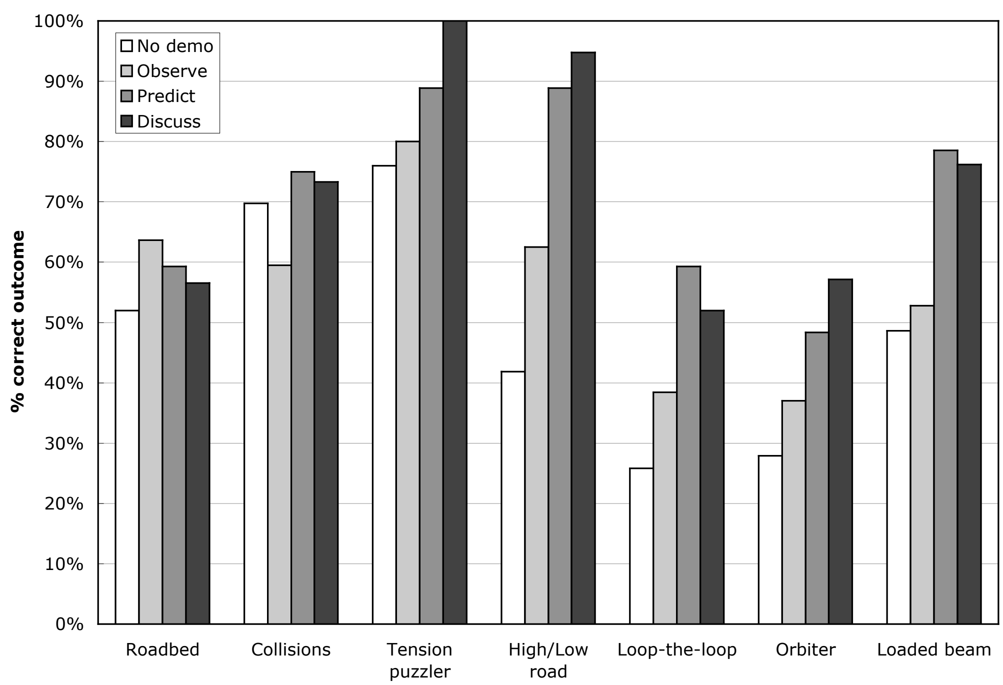


Figure 3.4. Proportion of students predicting the correct outcome by demonstration and mode of presentation for Physics 1a demonstrations.

Table 3.6. Proportion of students providing the correct explanation, by demonstration and mode of presentation for Physics 1a demonstrations.

Demonstration	*No demo*	*Observe*	*Predict*	*Discuss*
Roadbed	42.0%	36.4%	44.4%	47.8%
Collisions	39.5%	37.8%	58.3%	36.7%
Tension puzzler	28.0%	34.3%	38.9%	36.8%
High/Low road	18.6%	25.0%	38.9%	31.6%
Loop-the-loop	16.1%	28.2%	14.8%	28.0%
Orbiter	11.6%	14.8%	19.4%	42.9%
Loaded beam	2.7%	8.3%	21.4%	9.5%
Overall	22.7%	26.4%	33.7%	33.3%

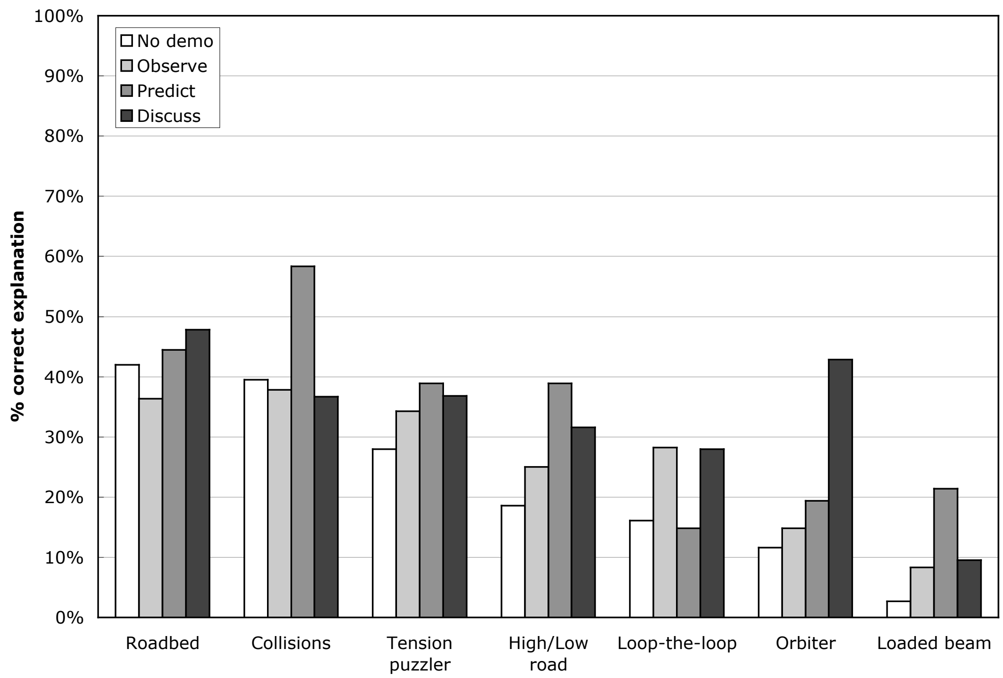


Figure 3.5. Proportion of students providing the correct explanation by demonstration and mode of presentation for Physics 1a demonstrations.

Results from Physics 1b: Electricity and magnetism

Rates of correct outcomes (Table 3.7) and correct explanations (Table 3.8) for each mode and for the *no demonstration* (control) group in the electricity and magnetism class are shown in Tables 3.7 and 3.8. Students in the *no demonstration* group provided correct outcomes 51.3% of the time and correct explanations 47.7% of the time. Those in the *observe* group correctly predict the outcome 65.2% of the time and properly explain the result 60.7% of the time. This difference is significant for both outcomes (Table 3.9) and explanations (Table 3.10). Students in the *predict* mode provide correct outcomes and explanations 68.3% of the time, which is not statistically different from those in the *observe* group. Students who saw the demonstration in the *discuss* mode significantly outperform both the *observe* and *predict* groups: 77.1% provide the correct outcome and 74.3% the correct explanation.

Table 3.7. Outcome responses by mode and correctness for all Physics 1b demonstrations, combined.

	Correct	Incorrect	Unclear	*n*
No demo	51.3%	46.9%	1.8%	384
Observe	65.2%	33.9%	0.9%	112
Predict	68.3%	31.7%	0.0%	101
Discuss	77.1%	22.9%	0.0%	109
Overall	59.9%	39.0%	1.1%	706

Table 3.8. Explanations by mode and correctness for all Physics 1b demonstrations, combined.

	Correct	Incorrect	Unclear	*n*
No demo	47.7%	48.7%	3.6%	386
Observe	60.7%	35.7%	3.6%	112
Predict	68.3%	29.7%	2.0%	101
Discuss	74.3%	22.9%	2.8%	109
Overall	56.8%	40.0%	3.3%	708

Statistical significance[*] and normalized gain[†] is given in Table 3.9 for outcomes and Table 3.10 for explanations. Figure 3.6 displays the improvement in rates of correct outcomes and explanations for each mode over those rates for the *no demonstration* group, normalized to the rates for the *no demonstration* group.

Table 3.9. Rates of correct responses for outcomes ($R_{outcome}$) by mode for all Physics 1b demonstrations, combined; statistical significance ($p_{outcome}$) compared to the *no demonstration* group (p < 0.05 is considered statistically significant; and normalized gain ($g_{outcome}$) relative to the *no demonstration* group[†].

	$R_{outcome}$	$p_{outcome}$	$g_{outcome}$	**n**
No demo	51.3%	—	—	384
Observe	65.2%	0.005	28.5%	112
Predict	68.3%	0.001	34.9%	101
Discuss	77.1%	< 0.0001	53.0%	109
Overall	59.9%	—	—	706

Table 3.10. Rates of correct responses for explanations (R_{explan}) by mode for all Physics 1b demonstrations, combined; statistical significance (p_{explan}) compared to the *no demonstration* group (p < 0.05 is considered statistically significant; and normalized gain ($g_{outcome}$) relative to the *no demonstration* group[†].

	R_{explan}	p_{explan}	g_{explan}	**n**
No demo	47.7%	—	—	386
Observe	60.7%	0.0075	24.9%	112
Predict	68.3%	0.0001	39.4%	101
Discuss	74.3%	< 0.0001	50.9%	109
Overall	56.8%	—	—	708

The demonstrations and related questions on the end-of-semester test varied widely in difficulty, even more so than in Physics 1a. Table 3.11 and Figure 3.7 show the rate of predicting the correct outcome by mode of presentation for each demonstration, while Table 3.12 and Figure 3.8 show the same for providing the correct explanations.

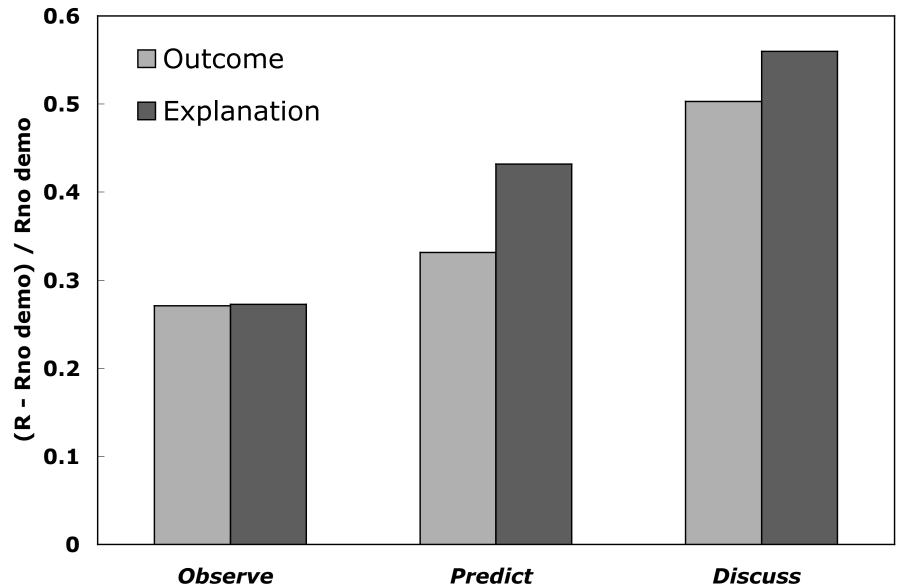


Figure 3.6. Improvement in rates of correct outcomes and explanations for each mode over those rates for the *no demonstration* group, normalized to the rates for the *no demonstration* group for Physics 1b demonstrations.

Table 3.11. Proportion of students predicting the correct outcome, by demonstration and mode of presentation for Physics 1b demonstrations.

Demonstration	*No demo*	*Observe*	*Predict*	*Discuss*
Two spheres	94.4%	100.0%	90.9%	100.0%
Charged cup	19.6%	40.0%	53.3%	41.2%
Series capacitor	46.8%	35.3%	62.5%	75.0%
B-field	38.3%	58.3%	47.4%	41.7%
Two coils	64.0%	46.7%	75.0%	81.2%
Eddy current	87.0%	90.0%	100.0%	91.3%
RC circuit	60.0%	76.9%	72.7%	88.9%
Half-lens	15.8%	61.5%	75.0%	90.0%
Overall	53.2%	63.6%	72.1%	76.2%

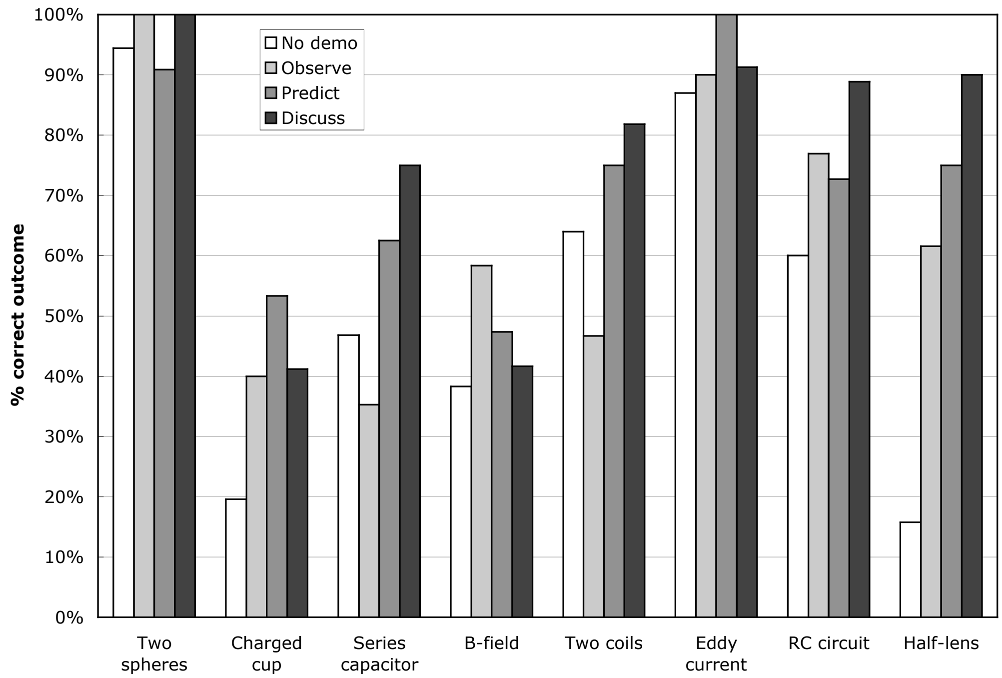


Figure 3.7. Proportion of students predicting the correct outcome by demonstration and mode of presentation for Physics 1b demonstrations.

Table 3.12. Proportion of students predicting the correct explanation, by demonstration and mode of presentation for Physics 1b demonstrations.

Demonstration	*No demo*	*Observe*	*Predict*	*Discuss*
Two spheres	88.9%	90.9%	90.9%	94.7%
Charged cup	17.4%	40.0%	53.3%	52.9%
Series capacitor	36.2%	41.2%	50.0%	50.0%
B-field	68.1%	83.3%	84.2%	91.7%
Two coils	60.0%	46.7%	58.3%	81.8%
Eddy current	78.3%	100.0%	100.0%	82.6%
RC circuit	32.7%	15.4%	36.4%	33.3%
Half-lens	18.6%	61.5%	87.5%	80.0%
Overall	50.0%	59.9%	70.1%	70.9%

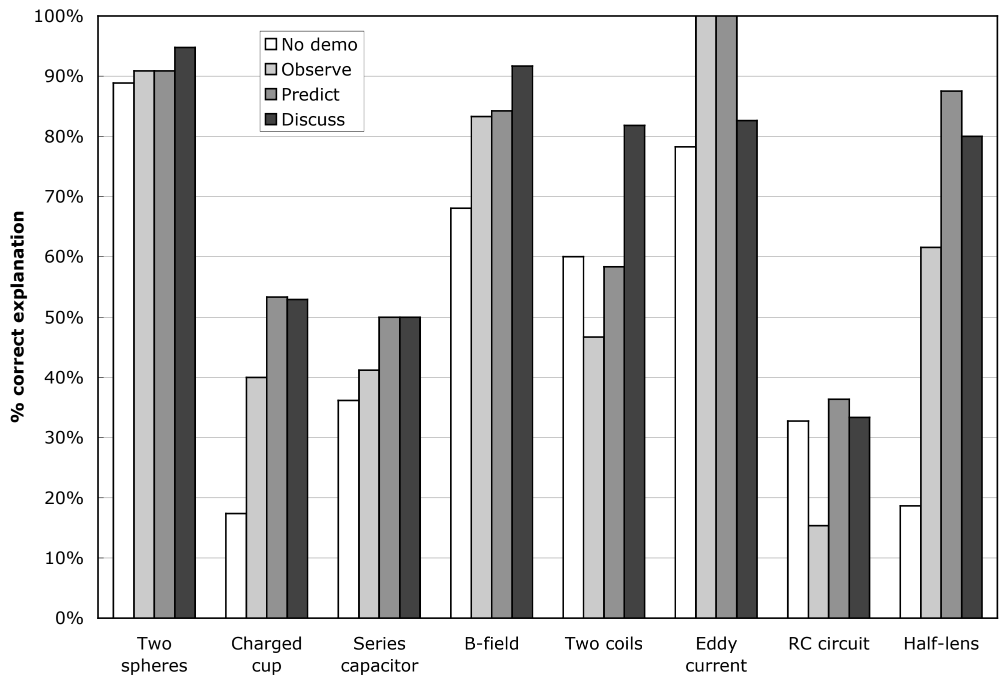


Figure 3.8. Proportion of students providing the correct explanation by demonstration and mode of presentation for Physics 1b demonstrations.

Time spent on demonstration

One might ask whether the effect of the different pedagogical modes is not due to the pedagogy itself, but to some other factor. The most obvious one is time spent on the demonstration. In particular, asking for predictions and asking students to discuss the observations and the reasons for it increases the time spent on the demonstration as well as the degree of interactivity. Can the improvement in learning be attributed solely to the increase in time spent on the demonstration?

Results from the mechanics semester argue against time being the most relevant factor. Demonstrations in Physics 1a conducted in *discuss* mode took 21 minutes on average while those in *predict* mode took 13 minutes on average (Table 3.13). Despite this 8-minute increase in time spent on the demonstration, performance on the end-of-semester demonstration test is virtually indistinguishable between these groups. Yet, the 2-minute addition of the prediction—as compared to the *observe* mode—does lead to a significant difference: there is a fourfold normalized improvement for both outcomes and explanations. As such, a modest addition to the time spent results in significantly greater learning.

Table 3.13. Average time spent on each demonstration, by mode of presentation.

	Physics 1a	Physics 1b
No demo	—	—
Observe	11 min	9 min
Predict	13 min	9 min
Discuss	21 min	16 min

Interactive Lecture Demonstrations

One might ask whether the actual demonstration matters—are some demonstrations more likely to contribute to student learning than others? There is some evidence that this is the case; for instance, Kraus [1997] found that demonstrations most effective for eliciting student difficulties are those in which the observations of the demonstration are closely linked to the concepts students find difficult. The demonstrations chosen for our initial study in Fall 2000 were selected from the standard repertoire of demonstrations offered by the Lecture Demonstration Services of the Science Center at Harvard University and are not necessarily designed to address particular misconceptions. So for Physics 1b in Spring 2002, we supplemented these demonstrations with demonstrations taken from a research-based curriculum, Interactive Lecture Demonstrations (ILDs).

ILDs consist of a series of demonstrations (often 7-8) that are designed to be conducted in sequence [Sokoloff and Thornton 1997]. Both the content and order of the ILDs is based on the results of research in physics learning, designed to address common student difficulties and misconceptions. Students are given a worksheet that guides them through their predictions and the demonstrations in order. Most of the ILDs incorporate Microcomputer-Based Laboratory (MBL) tools so that the phenomenon being investigated can be demonstrated and measured in class. Each set of ILDs is packaged to take approximately 40 minutes per lesson [Thornton and Sokoloff 1998]. Subsequent demonstrations are designed to build upon one another, adding detail to the scenario as different features are added.

Three demonstrations from the ILD curriculum were included among the nine demonstrations we conducted in Physics 1b [Thornton and Sokoloff 1998]: “Two coils,” “RC circuit,” and “Half-lens.” It is not possible to directly compare the effectiveness of an ILD-derived demonstration with a non-ILD demo since the scenarios differ in difficulty level. However, we can observe general trends in these research-based demonstrations.

If we compare the performance of students in the *no demonstration* group for ILD and non-ILD demonstrations, we can obtain a measure of the *relative* difficulty of the questions (Table 3.14 and Figure 3.9). The rate of correct prediction for students in the *no demo* group is 45.7% for the three ILD demonstrations but 55.4% for the non-ILD demonstrations, a performance gap of around 10%. This suggests that the ConcepTests based on the ILD demonstration may be slightly more difficult than the ConcepTests based on the demonstrations not derived from ILDs. A similar pattern is shown in the *observe* group: 61.0% correct outcomes for the ILDs and 67.6% for non-ILDs, a gap of 6.6%.

Table 3.14. Comparison of ILD with non-ILD demonstrations by rate of correct outcomes.

	ILDs			Non-ILDs		
	Correct	*p*-value	*n*	Correct	*p*-value	*n*
No demo	45.7%	—	162	55.4%	—	222
Observe	61.0%	0.040	41	67.6%	0.035	71
Predict	74.2%	0.002	31	65.7%	0.064	70
Discuss	86.7%	< 0.001	30	73.4%	0.002	79
Overall	70.6%	—	264	62.2%	—	442

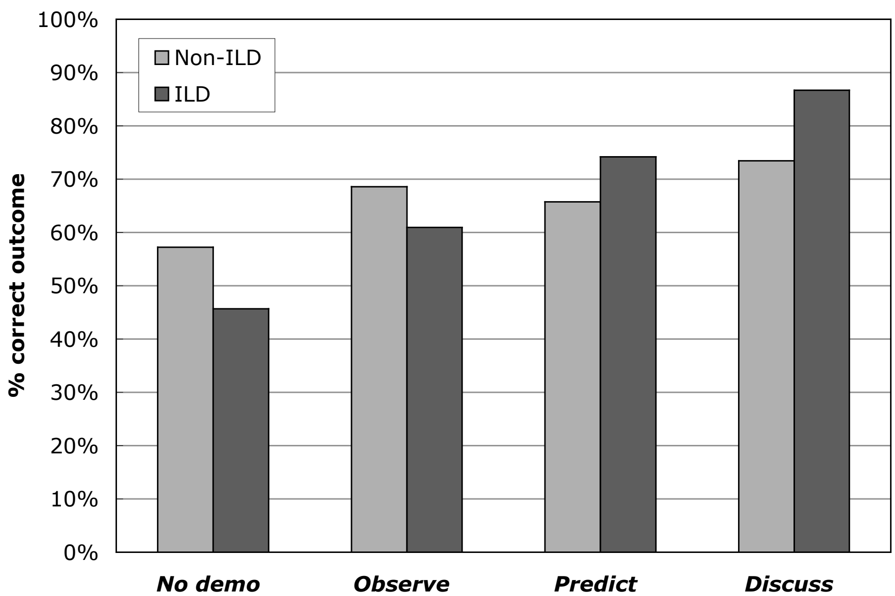


Figure 3.9. Comparison of ILD with non-ILD demonstrations by rate of correct outcomes.

However, in the more interactive modes, the improvement of students on the ILDs is greater than on the non-ILDs: the rate of correct outcomes for the *predict* mode is 74.2% for ILDs but only 65.7% for non-ILDs, and for the *discuss* mode, the rate is 86.7% for ILDs but only 73.4% for non-ILDs (Table 3.14). Students actually perform better in the *predict* and *discuss* modes in predicting the correct outcome of the scenario, in spite of the greater difficulty (Figure 3.9).

The results for explanations show a similar pattern. The rate of correct explanations for the *no demo* mode is 36.0% for the three ILD demonstrations but 56.3% for the non-ILD demonstrations, a performance gap of over 20% (Table 3.15 and Figure 3.10). This gap widens to 30% for the *observe* mode as the performance of students on the non-ILDs increases. For the more interactive modes, this performance gap narrows to approximately 15% for the *predict* mode and just over 10% for the *discuss* mode.

Table 3.15. Comparison of ILD with non-ILD demonstrations by rate of correct explanations.

	ILDs			Non-ILDs		
	Correct	*p*-value	*n*	Correct	*p*-value	*n*
No demo	36.0%	—	164	56.3%	—	222
Observe	41.5%	0.258	41	71.8%	0.010	71
Predict	58.1%	0.011	31	72.9%	0.007	70
Discuss	66.7%	< 0.001	30	77.2%	< 0.001	79
Overall	42.9%	—	266	65.2%	—	442

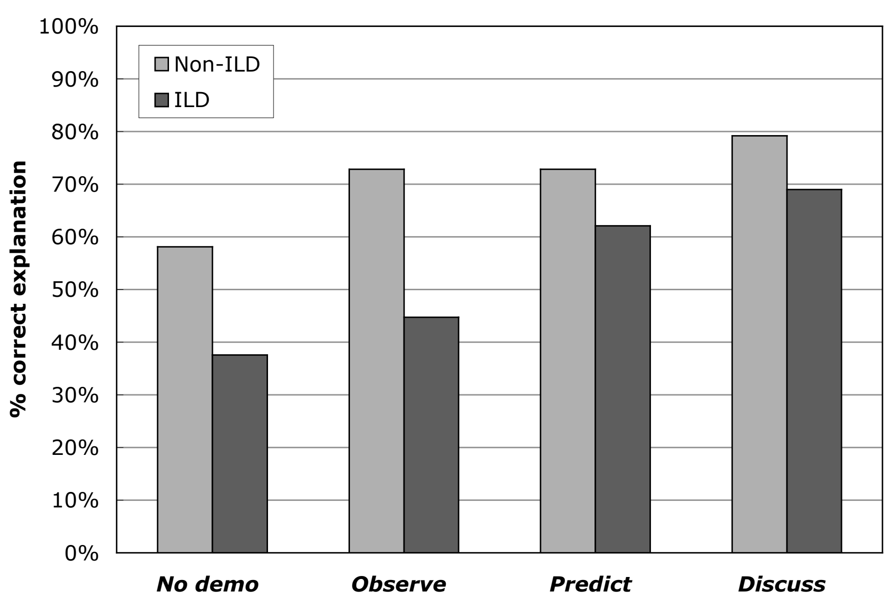


Figure 3.10. Comparison of ILD with non-ILD demonstrations by rate of correct explanations.

Thus, even though the ILD scenarios are, in some sense, more difficult than the non-ILD situations—as evidenced by student performance in the *no demo* and *observe* modes—they do appear more effective in enhancing student learning than the non-ILD demonstrations.

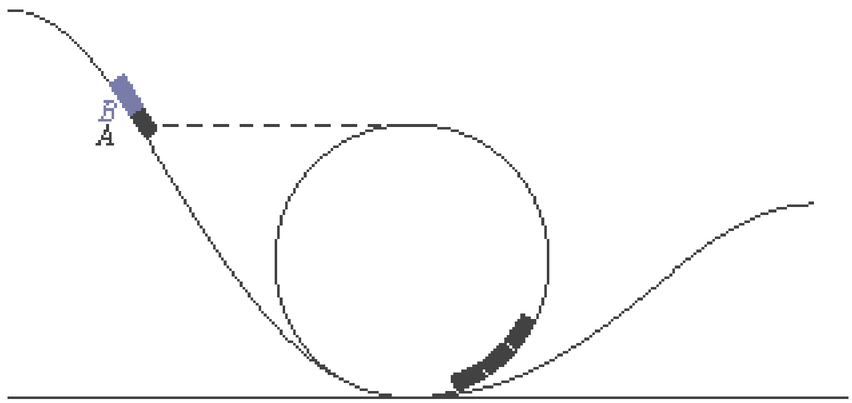
Specific demonstrations

I will now briefly discuss three of the demonstrations—one from Physics 1a and two from Physics 1b, because each of these demonstrations is associated with a unique circumstance. Although the overall results are not changed by these special cases, they bear mention in this thesis. In one case—the loop-the-loop demonstration in Physics 1a—a very similar question was asked on a homework set that all students completed. In another case—the eddy current demonstration in Physics 1b—a very similar ConcepTest was posed in the lecture component in class that almost all students attended. Finally, the half-lens demonstration in Physics 1b was performed in the week immediately prior to the end-of-semester demonstration test. Since results for that demo show such a striking effect of the *observe* mode, there is some question if student memory is the predominant factor rather than understanding of the relevant phenomena.

Loop-the-loop demonstration (Physics 1a)

A problem very similar to one of the demonstrations used in Physics 1b was assigned as homework. Consequently, all students spent time considering that situation even if they were in the *no demonstration* group for that week. Although the questions are not identical, they involve essentially the same physics. Question 7 on the end-of-semester test for Physics 1a explained the demonstration as presented during section (Figure 3.11).

A roller-coaster track includes a complete loop as shown in the diagram. Cars move along the track with negligible friction and are not held in place on the track by rails.



If a car is held on the initial part of the track at exactly the same height as the top of the loop (position A in the diagram), does it travel all the way around the loop successfully? If the car is held at position B, slightly higher than the top of the loop, does it travel all the way around the loop successfully? Explain your answer briefly.

Figure 3.11. Question regarding loop-the-loop demonstration as presented in the end-of-semester test in Physics 1a.

The corresponding homework problem (which also appears as question 33 from the *Practice Volume* accompanying the text) is shown in Figure 3.12.

1. (Ch. 12, Q33) Your uncle owns an amusement park, and he wants you to add a circular loop to an existing roller coaster ride to make it more fun. The first "hill" for the existing roller coaster is 55 m tall, and your uncle wants you the build the tallest loop possible after it without the car falling out of the tracks or the passengers out of the cars. You think for a minute and realize what the minimum speed at the top of the loop will have to be and this gives you what you need to calculate the planned height.

Figure 3.12. Question very similar to the loop-the-loop demonstration posed on Problem Set 5 of Physics 1b. This question was taken from the *Practice Volume* accompanying the textbook under development by Eric Mazur.

Table 3.16 shows the data for just the loop-the-loop demonstration.

Table 3.16. Proportion of students predicting the correct outcome and providing the correct explanation for the loop-the-loop demonstration.

	Outcome	Explanation	*n*
No demo	25.8%	16.1%	31
Observe	38.5%	28.2%	39
Predict	59.3%	14.8%	27
Discuss	52.0%	28.0%	25
Overall	42.6%	22.1%	122

Perhaps surprisingly, students did not appear to perform any better on the question on this demonstration; the pattern of correct outcomes and explanations was quite similar for the loop-the-loop and several other demonstrations (see, for instance, Figures 3.4 and 3.5). This is even though all students should have solved an essentially identical scenario in the homework (Figure 3.12). In fact, it appears to be more difficult than the other Physics 1a demonstration—as evidenced by the low rate of students predicting the correct outcome for the *no demo* mode even though, as mentioned, there should not have been any students who did not confront essentially the same question during the semester. This seems to indicate that learning from the problem set question did not necessarily sink in to the degree that students in all modes could call upon it at the end of the semester.

Eddy current demonstration (Physics 1b)

The eddy current demonstration used in the workshop section of Physics 1b was very similar to a demonstration and ConcepTest conducted in the lecture. Therefore, it is interesting to compare student performance on this question to their performance on the ConcepTest. The question as posed in the end-of-semester demonstration test is shown in Figure 3.13.

Two rods of the same diameter, one aluminum (conducting, non-magnetic) and one wooden, are held vertically. If identical magnetic rings are placed around the top of each rod and released at the same time, which ring reaches the bottom first? Explain briefly. (Neglect any friction between the ring and the rod.)

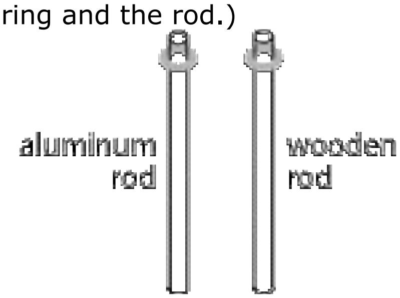
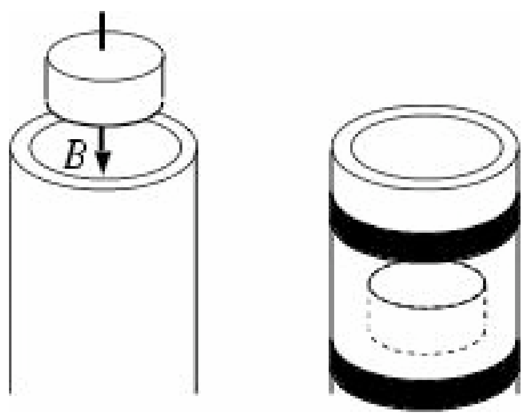


Figure 3.13. Question regarding eddy current demonstration as presented in the end-of-semester test in Physics 1b.

The ConcepTest used in class is shown in Figure 3.14.

A permanent magnet is dropped through a long aluminum tube, as shown. As the magnet drops, electric currents are induced around the tube. Compared to a freely-falling magnet, the magnet through the tube drops



1. more slowly.
2. exactly the same way.
3. faster.
4. Need more information.

Figure 3.14. ConcepTest regarding eddy currents as presented in Physics 1b on April 11, 2002. ConcepTest #3691 in the Interactive Learning Toolkit <http://www.deas.harvard.edu/galileo/>.

When the ConcepTest shown in Figure 3.14 was posed in the lecture period, 43% of the students provided the correct answer before discussion, with an increase to 57% following discussion (gain $g = 0.18$). Table 3.17 presents the data from the eddy current demonstration on the end-of-semester test. It shows that students did significantly better on this demonstration than the other Physics 1b demonstrations for both outcomes and explanations. This excellent performance is shown for all modes of presentation, including the *no demo* group. This could indicate that the demonstration is relatively easy for students to predict and explain. However, the relatively low frequency of correct predictions for the similar demonstration (Figure 3.14) in the lecture period suggest that the demonstration may not be easy. Thus, it seems likely that the observation of a nearly identical situation in lecture contributed to the high rate of correct outcomes and explanations on the end-of-semester demonstration test. A similarly good performance was also exhibited at the time the demonstration was performed: for students in the *predict* and *discuss* modes, 90% of students correctly anticipated that the ring on the wooden rod would reach the bottom first (see below for additional discussion of in-class predictions).

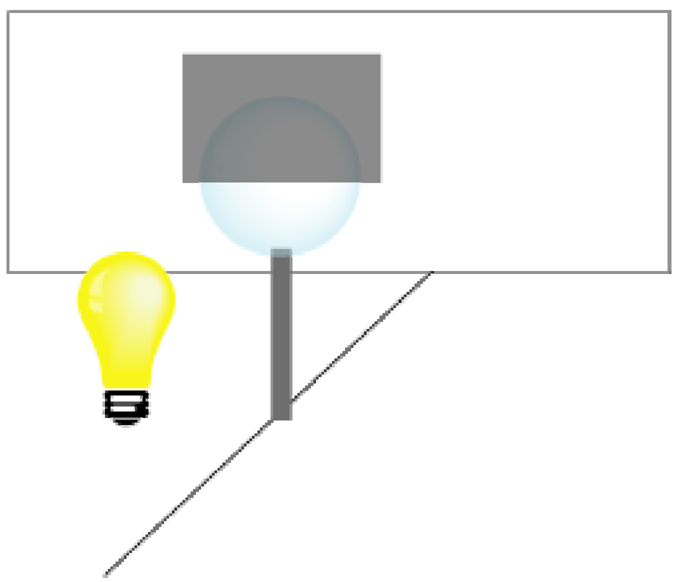
Table 3.17. Proportion of students predicting the correct outcome and providing the correct explanation for the eddy current demonstration.

	Outcome	Explanation	*n*
No demo	87.0%	78.3%	46
Observe	90.0%	100.0%	10
Predict	100.0%	100.0%	9
Discuss	91.3%	82.6%	23
Overall	89.8%	84.1%	88

Half-lens demonstration (Physics 1b)

The final demonstration to discuss separately is the half-lens demonstration from Physics 1b (Figure 3.15). Two things about this demonstration stand out: (1) it is the sole optics demonstration and might not really deserved to be included amongst the other electricity and magnetism demonstrations; and (2) it was conducted during the reading period in the week immediately prior to administration of the end-of-semester demonstration test. This demonstration was one of those from the ILD curriculum and was provided by David Sokoloff (University of Oregon).

A giant light bulb is placed to the left of a converging lens at a distance greater than the focal length of the lens. The image of the bulb is formed on a screen to the right of the lens. What will happen to the image if you block the top half of the lens with a card?



1. The top half of the image disappears.
2. The bottom half of the image disappears.
3. The entire image disappears.
4. The image becomes blurred.
5. The image becomes fainter.

Figure 3.15. Question regarding half-lens demonstration as presented in the end-of-semester test in Physics 1b.

The striking result for this demonstration is that passive observation resulted in significantly greater learning than students in the *no demonstration* group. Only 16% of students who did not see the demonstration were able to correctly predict the outcome of covering the top half of the lens—namely, that the entire image would become fainter (Table 3.18). But 60% of students who saw the demonstration in *observe* mode provide this correct outcome, the most dramatic effect of *observe* mode in the entire set of demonstrations. There is a further improvement in the rate of correct predictions to 75% in *predict* mode and 90% in *discuss* mode.

Table 3.18. Proportion of students predicting the correct outcome and providing the correct explanation for the half-lens demonstration.

	Outcome	Explanation	*n*
No demo	15.8%	18.6%	58
Observe	61.5%	61.5%	13
Predict	75.0%	87.5%	8
Discuss	90.0%	80.0%	10
Overall	36.4%	37.8%	88

Several factors about this demonstration might explain the striking effect of the *observe* mode. This is the sole optics demonstration among those conducted in both Physics 1a and 1b. Perhaps there is something about optics that differs from both mechanics and electricity and magnetism, in particular the visual nature of the subject. Optics is inherently visual in that what is observed is exactly what is happening; this is unlike the remainder of demonstrations that require a translation between observation and physical significance, especially electricity and magnetism in which one often has to rely upon a meter or instrument in order to detect an effect. In this example, it may be that the particular outcome is so surprising that merely observing the result is enough for students to dramatically shift their views. Nearly all of the students who did not see the demonstration thought that half of the object would disappear from the image when the top half of the lens was blocked—although they disagreed on whether it would be the top or the bottom half of the object. As the observed result so clearly differed from their expectation, perhaps students remembered the outcome because it was so surprising; a demonstration in which the expected outcome did not differ as radically from their expectation might not have shown the same dramatic effect.

Secondly, the half-lens demonstration was conducted so close to the end of the semester that we may be observing simple recall of the scenario, rather than actual learning or understanding. There was a gap of only one week between the time that the demonstration was carried out in section and when it was assessed on the end-of-semester test, suggesting that there may be a different process occurring here than for the other demonstrations.

Accuracy of predictions

An interesting pattern emerges when looking at the predictions students made *before* observing the demonstration, that is by examining the answers submitted via the PRS system during the section meeting. For Physics 1a, students in the *discuss* sections provided a correct prediction more frequently than their classmates in the *predict* sections: 41% vs. 26%. Despite this difference at the time the demonstration was performed, the ability of students to correctly predict the outcome of the same scenario on the end-of-semester test on the demonstration is identical for both the *discuss* and *predict* students at 69%. So, despite an initial difference in the rate of correct prediction, there is no difference at the end of instruction (Figure 3.16).

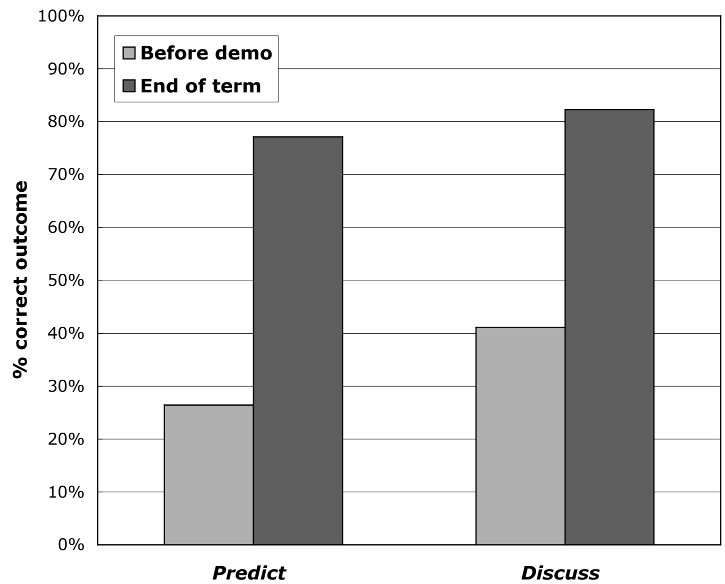


Figure 3.16. Accuracy of student predictions made in class before seeing the demonstration (in *predict* and *discuss* modes), compared with accurately predicting outcomes on the end-of semester test, Physics 1a.

I suspect the reason for the higher rate of correct predictions for students in the *discuss* section is due to the worksheet used for that mode. In particular, students are asked to write down their prediction for the demonstration before being provided with answer choices in the form of a ConcepTest. In contrast, students in the *predict* mode are not asked to record their prediction in any way until after the answer choices are provided by way of a ConcepTest. It seems reasonable that students would be more thoughtful and really think about the scenario more critically then when their consideration is free-response as opposed to merely multiple-choice. Even though students in the *predict* mode were asked to think carefully about the situation before the ConcepTest is offered, there is no written component or anything that follows-up on that request other than the multiple-choice options on the associated ConcepTest.

When this analysis was repeated with data from Physics 1b, a different pattern emerges (Figure 3.17). In particular, there is virtually no difference in the accuracy of predictions made before the demonstration was performed for the electricity and magnetism demonstrations. Students in both the *predict* and *discuss* modes predict the correct outcome of the demonstration 51% of the time, aggregated across demonstrations and sections (Figure 3.17). As mentioned above, the outcomes reported on the end-of-semester test did differ between these modes (68% for *predict* and 77% for *discuss*), although this difference is not statistically significant ($p = 0.08$).

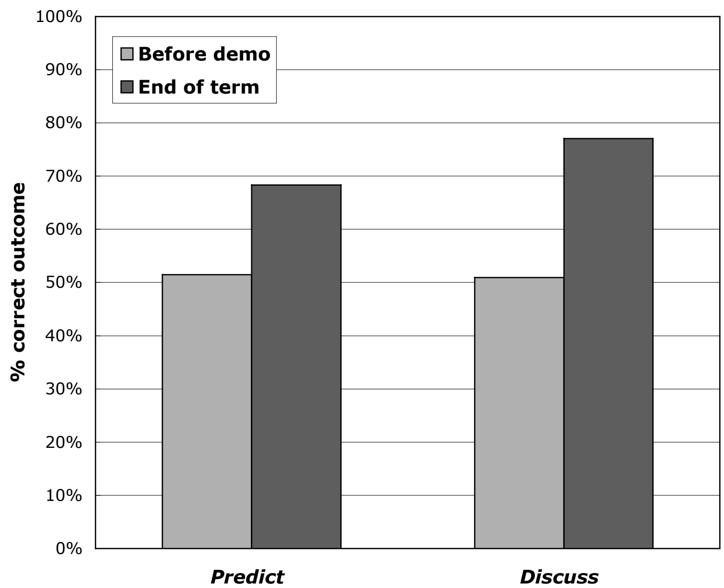


Figure 3.17. Accuracy of student predictions made in class before seeing the demonstration (in *predict* and *discuss* modes), compared with accurately predicting outcomes on the end-of semester test, Physics 1b.

Comparing Physics 1a and 1b

Before concluding, it is worth revisiting the question of whether Physics 1a and 1b are really different from each other. For instance, does it make sense to include data from the one optics demonstration with those from electricity and magnetism? There is some diversity in the difficulty of the demonstrations used, in how easy it was to observe the desired phenomenon, and in how easy it was to interpret the observation to come to an understanding of the demonstration. One area worthy of further study would to more fully consider the characteristics of the demonstrations used. This would allow appropriate aggregation of data from different demonstrations and might help in isolating characteristics of demonstrations relevant to learning.

Demonstrations in Physics 1a seem harder to understand than those in Physics 1b, as illustrated by the proportion of students able to provide the correct explanation (Figure 3.2), even though students were able to predict the correct outcomes at about the same rate (Figure 3.1). Also, students are better able to predict the outcome of Physics 1b (electricity and magnetism) demonstrations just by having seen it passively (*observe* mode) than they are for Physics 1a (mechanics) demonstrations.

Gattis [1995] observed similar results, looking at conceptual gains by students on three specific topics under three pedagogical strategies: prediction and discussion, prediction, and as problems worked during the lecture. For each of the three topic areas, he observed that a different pedagogy was most effective at preparing students for a post-test on the topic. That is, different topics may lend themselves to demonstrations in different ways [Gattis and Park 1997]:

> Demonstrations (1) may help confer belief on a concept that a student finds counterintuitive; (2) may provide visual images that are important

> components of rich and detailed concepts; (3) may help to explain concepts that have key spatial and temporal relationships; and (4) may provide especially vivid physical examples that are useful in making analogies to other examples and generalizing to a more abstract concept. [Gattis 1995]

I posit that students have more intuition about mechanics than they do about electricity and magnetism since mechanics is closer to their everyday experience. And, thus, students are more likely to use their own thinking and intuition to answer questions related to mechanics than they are to use what they learned in class [cf. Singh 2002]. Because electricity and magnetism is farther from students' normal experience, they are less likely to have a "gut" feeling than for mechanics; as such, students are perhaps more willing to call upon what they have just learned in order to explain an observed phenomenon.

Conclusion and future directions

Results from the two semesters of classroom demonstrations in introductory physics reveal some interesting observations. Increasing the degree of interactivity with which demonstrations are presented appears to also increase student learning from those demonstrations, especially understanding of underlying concepts. Although different demonstrations and topics show different individual patterns, the overall message that increased engagement—through prediction and discussion—leads to increased learning seems robust. This suggests that demonstrations should be coupled with predictions so that students are more engaged in what is going to occur. Adding a prediction to the traditionally presented demonstration adds a nominal amount of time to the presentation but seems have a significant effect on student learning, at least for some topics.

Results from Physics 1b, in particular, show that several demonstrations might be too easy for the students, *e.g.*, two spheres and eddy current demonstrations. That is, all students perform so well for outcomes and/or explanations that there is resolution lost at the top of the scale. Removing these demonstrations from the analysis presented in this chapter does not change the overall result. In the future, it would be useful to select demonstrations for these topics that are more challenging so that there is no saturation of correct responses.

Although the results from the two semesters reveal the same overall picture, there are some interesting distinctions. In particular, the four modes are more clearly distinguished from each other in Physics 1b than they are in Physics 1a (where *predict* and *observe* are virtually indistinguishable from each other and *no demo* and *observe* quite similar to each other as well). It raises the question, therefore, if the content differences between mechanics and electricity and magnetism are responsible for this different pattern. This might also shed light on the marked difference for the explanation results between the two semesters, even though the outcome results are quite similar.

Further, it would be interesting to isolate the characteristics of a demonstration that make it most effective for student learning. For example, it would be helpful to quantify the underlying difficulty of the physics involved in explaining each demonstration as well as several characteristics of the nature of the demonstration including: ease of observing the desired outcome, ease of interpretation, "tangibility" of the demonstration, surprise of the outcome, and such factors as the time in the semester. The expectation is that the inherent difficulty of the demonstration should affect performance on all modes of observation, including the *no demonstration* group. In contrast, the features associated with the nature of the presentation should only be revealed in the three modes in which the demonstration is actually observed, *i.e.*, not in the *no demo* group.

The results from Physics 1b suggest that there may be some interesting distinctions between the research-based ILDs and traditional demonstrations. We could not explore this trend more fully as there are only a small number of electricity and magnetism ILDs that do not duplicate the content in the *Tutorials in Introductory Physics* [McDermott *et al*. 2002]. It would be interesting to carry out a similar comparison of ILDs and traditional demonstrations in the mechanics course since there are more ILDs related to mechanics and since the mechanics *Tutorials* do not have as significant a demonstration/equipment component.

Discussion sections provide a good approximation to lecture but they are not the same. The setting in sections is more intimate but the demonstrations are more likely to be conducted in isolation, disconnected from the flow of the discussion, as compared to lecture. An ideal setting for such a study would be a class in which there are multiple *lecture* sections taught by the same instructor. This would allow the context of the study to truly be that of the large lecture course, the setting in which classroom demonstrations are perhaps most relevant.

Chapter 4:

Informing conceptual biology assessment by examining genetics vocabulary

I originally set out to extend the progress that the Mazur Group had demonstrated in undergraduate physics education to the biological sciences, especially in the area of Peer Instruction (see Chapter 2). What I soon discovered, however, is that the field of biology education research is less developed than physics education research. Until recently, most of the contributions to biology education have been in the form of "how to" articles that describe the use of classroom techniques or laboratory exercises used by instructors in their own classes [Sundberg 2002]. It is only recently that researchers in biology education have recognized the need for a rigorous assessment program that goes beyond simply noting the instructor's perceptions and the student's reactions.

Undergraduate education in the biological sciences too often consists of presenting an encyclopedic array of facts that students can memorize and regurgitate on an examination. This type of learning does not necessarily translate into true understanding and appreciation for the underlying concepts of biology. In fact, a number of studies have revealed that students often graduate from high school with distorted views of the biological world [Mintzes *et al*. 1998, 2000; Wandersee *et al*. 1989]. There has, however, been a growing awareness of the need to refocus biology education more on conceptual understanding [Mintzes *et al*. 1997, 1998, 2001; Wandersee *et al*. 1994]. But there is no standard tool to assess that conceptual understanding, to differentiate between biology students who excel by memorizing facts and details and those who really understand the material.

This chapter presents some preliminary work on the development on a test of conceptual understanding in introductory genetics. I will first sketch out some of the considerations necessary for developing a valid and validated assessment tool and then turn to a survey examining introductory genetics students' knowledge and comfort with basic terminology and concepts.

Biology vocabulary

Unfortunately, rote learning of vocabulary, algorithms and facts is characterizes much undergraduate science education. This is especially true in the biological sciences [Posner 1996]. The recent explosion of biological knowledge has led to a corresponding explosion of terminology—much more so than in chemistry or physics [Stanley and Stanley 1986]. The immense number of terms is a significant stumbling block for many students [Wandersee 1988]. In fact, the amount of new vocabulary students encounter in an introductory biology course often exceeds that in a beginning foreign language course [Yager 1983]! Yet biology is not taught in the same manner as a foreign language: there is rarely any opportunity for students to use the newly-learned words in biology classes, while foreign language classes devote significant time to practice conversations.

Biology vocabulary has very specialized meanings that are often must be precisely understood in order to fully appreciate the scientific issues being introduced and discussed. According to Robin Dunbar [1995], "the need to have words that refer to specific phenomena is extremely important in science: failure to do so results in unnecessary talking at cross purposes, wastes people's time and holds up the advancement of science." In order to understand many biology textbooks and lectures, one must already have a familiarity with the basic vocabulary of the science. This can be quite challenging for beginning students as different textbooks do not always agree on definitions. For instance, Wilkinson [2001] looked at 12 university-level ecology and environmental science textbooks for the meaning of 'symbiosis,' finding only three that presented both of the definitions he claimed. What is a student to do, therefore, if they cannot even trust their textbook? As complete understanding of terminology is quite difficult, many students just memorize the definition provided by their instructor, without any real attention to actual understanding.

Constructing a test for conceptual understanding

A standard instrument for assessment of students' understanding of concepts in biology would be tremendously valuable [Stokstad 2001; M. Zeilik, pers. comm.; R. Iuli, pers. comm.; S. Benson, pers. comm.]. Such a test would allow teachers to identify common misconceptions among their students and, in so doing, encourage the development and use of curricula that target these misconceptions [Hake 2002]. It would allow educators to assess the effectiveness of instructional techniques on a standardized conceptual test. And it would guide curriculum developers to design classroom activities, educational materials, and assessment instruments which are focused on common misconceptions [Sadler 1992]. Constructing such an instrument that does not depend upon knowledge of particular vocabulary is a particular challenge, so it is important to learn the terminology that introductory biology students do know.

The Force Concept Inventory (FCI)

A critical tool in recent physics education research is the Force Concept Inventory (FCI) [Hestenes *et al*. 1992; Mazur 1997]. This multiple-choice conceptual test of introductory mechanics makes it easy for instructors with an interest in education to assess their students' understanding of the basic concepts being covered in the class. The multiple-choice format of the test makes it easy to administer and score to an arbitrarily large number of students, especially with the possibility of on-line administration. Further, the FCI probes understanding of basic concepts using language that is understandable to the novice* who has not yet taken a course in the subject. Since the test has been administered so frequently, there is a significant national—and international—database of results that make it possible to interpret results from a particular class in a broader context [Henderson 2002].

The FCI is among the most widely used tests of conceptual understanding in undergraduate science and has been administered to many thousands of high school and college physics students [Hake 1998; Hestenes *et al*. 1992]. It is designed to identify a student's *common sense* notion of mechanics, as these common sense notions often get in the way of real learning in physics [Halloun and Hestenes 1985b]. The FCI is based on

* It should be noted, however, that the colloquial nature of the FCI might prompt students to consider it differently than the more formal questioning common on examinations [Steinberg and Sabella 1997].

an earlier instrument, the Mechanics Diagnostic Test (MDT). which was scrutinized and validated at its creation by experts in the field to agree on the correct answers. Interviews with pilot testers were conducted to assure that students were interpreting the questions correctly and provided the same answer to question posed orally as those on the written test. These interviews helped assure that student responses reflect "stable beliefs" [Halloun and Hestenes 1985b]. Interviewers introduced contrary information and challenged students on their beliefs, finding student answers to be quite stable [Halloun and Hestenes 1985a]. The FCI itself has not been subjected to the same rigor of validation as the MDT, although the tests do share about half of their questions [Savinainen and Scott 2002a]. The FCI was originally constructed as a 29-question multiple-choice test; it has since been revised to clarify certain ambiguities, and now includes 30 multiple-choice questions [Hestenes *et al*. 1992; Mazur 1997].

While its use is quite widespread, the FCI has also been extensively criticized. Although there is no specific evidence that the FCI is invalid, it was never directly validated [Dancy 2000]. In defense, the authors say that the MDT and FCI are essentially similar tests and cite similar scores as evidence [Dancy 2000]. More significantly, its use by many thousands of students across the country demonstrates that the test is reliable and that similar pre- and post-test scores were observed in similar courses at different institutions [Savinainen and Scott 2002a]. Yet, Steinberg and Sabella [1997] showed that the context can have an effect on students' ability to correctly answer a question. They posed examination questions that involved the identical thought process in physics as questions on the FCI but presented in a different context; over 90% of the students answered one question correctly on the Steinberg and Sabella exam, but only 54% on the FCI, even though the physics required for both questions is identical.

Among other results, the FCI has been used to demonstrate that students can successfully solve traditional quantitative problems without having a basic understanding of the underlying concepts involved; in fact, it is the FCI that prompted Mazur to revise the way he teaches [Mazur 1997]. Giving this seemingly simple test to his introductory physics course revealed that the students had not, in fact, mastered the concepts of mechanics even though they were able to perform difficult calculations correctly. Some students were even aware of the disconnect between their intuition about the physical world and what they were learning in class; Mazur [1997] quotes one student (p. 4) who, when presented with the FCI, asked if he/she should answer the questions "according to what you taught us, or by the way I *think* about these things." Whatever limitations the FCI may have, it makes it easy to test students on the concepts of introductory mechanics. The usefulness of the FCI comes from the fact that it is written in plain English[*]; it can be understood by someone who has not taken a physics course in this subject area. Even if the answers require a knowledge of physics, understanding what the questions are asking does not. As such, it can be used as an effective pretest to address students' notions of mechanics even before instruction.

Similar tests have been developed in chemistry [BouJaoude 1992; Bowen and Bunce 1997; Mulford 1996; Mulford and Robinson 2002; Russell 1994][†] and astronomy

[†] http://jchemed.chem.wisc.edu/JCEWWW/Features/CQandChP/CQs/ConceptsInventory/CCIIntro.html

[Sadler 1992, 1998; Zeilik *et al*. 1997, 1999][‡]. In biology, however, no similar instrument is widely used.

NABT-NSTA High-School Biology Examination

In the mid 1980s, the National Association of Biology Teachers (NABT) and the National Science Teachers Association (NSTA) joined together to establish a "standardized test for high school biology" [Schulz 1989]. A joint committee of the two organizations drafted a series of concepts to be included in the test, as well as "processes and skills" that should be incorporated into the questions, and put out a call for comments and suggested test questions. Responses were judged by both college- and high-school-level faculty and sent out for field-testing in 1986. The original test was revised after 1987 to correct misleading language. It was scheduled to be completely redesigned and tested at the 1990 request of the memberships of both NABT and NSTA, but I could find no mention of the revision on either the NABT or NSTA Web sites and a search of the ERIC database does not turn up any recent uses of the test. However, even if it did still exist, the NABT-NSTA test was not specifically designed to address *concepts*.

Other efforts

Recently Nazario *et al*. [2002] developed a 25-item concept-oriented test[§] administered to general biology students at the University of Puerto Rico. They offered the instrument in a pre- and post-test format to a heterogeneous population of students, though one that did not include biology majors. The items on the test focused on one of

---

[‡] http://www.flaguide.org/tools/diagnostic/adt.htm; http://solar.physics.montana.edu/aae/adt/

[§] G. Nazario graciously shared a copy of the complete test with us.

10 major topics that the authors identified as "fundamental to biology" and that would be covered during the relevant class. Nazario *et al.* [2002] found that student performance improved significantly for each question on the test ($p < 0.01$ when comparing pre- and post-test responses by question). In particular, they looked for persistent incorrect answers on the post-test that would indicate misconceptions that resisted modification during the semester. Since the test is meant to accompany a general survey course in biology, its coverage is quite broad yet perhaps too dependent upon the particular setting for which it was developed. This limits its use as a general conceptual instrument for introductory biology nation- or worldwide.

In addition, there are several conceptual inventories in biology that are restricted to specific narrow topics. For example, Anderson *et al.* [2002] recently developed a 20-item multiple-choice Conceptual Inventory of Natural Selection. This complements other tests for evaluating students' conceptual understanding of evolution [Bishop and Anderson 1985, 1990; Settlage and Odom 1995], diffusion and osmosis [Odom 1995; Odom and Barrow 1995], and photosynthesis [Wandersee 1985].

Identifying misconceptions

Gathering student data on their understanding prior to instruction is a helpful way to target the lesson to the topics students most need help with and to establish the appropriate foundations for the understanding of new material. Instructor acknowledgement of students' prior knowledge is, in fact, important for instilling conceptual understanding among students. As David Ausubel [1968] is quoted as saying (p. 336), "the unlearning of preconceptions might very well prove to be the most determinative single factor in the acquisition and retention of subject-matter." Student misconceptions can be quite strongly held and it is important that students are encouraged to "unlearn" these prior ideas before they are receptive to new, correct, ones. Therefore, identifying preconceptions is an important first step in designing instruction to address them. Students may struggle with a concept that teachers think they already understand; and teachers are not always aware of their students' misconceptions. For example, when Lightman and Sadler [1993] asked elementary teachers to predict their students' beliefs about the motion of the Earth, they found that teachers always predicted a higher level of understanding for their students than was actually observed in the student population.

Assessment of preconceptions can be quite challenging and one must be especially careful with the use of terminology. In fact, judicious use of jargon can make for very tempting distractors on multiple-choice tests since they "sound right" even if they have little scientific value [Sadler 1992, 2000]. The developers of the NABT-NSTA biology test realized that students often used a strategy they term "word association" in order to answer the questions on that test. For instance, students chose an answer that contained the word "gamete" on a question about meiosis even though that answer made no sense [Schulz 1989]. Students read for key words and select the one that has the right vocabulary, even when it is not used correctly—and these association strategies for students on most tests they take. Instructors and textbook authors may even contribute to this problem by discussing the recommended problem-solving strategy: demonstrating how to recognize the type of problem, which equation or procedure to use, and how to be sure that the answer makes sense. As long as students can recognize the type of question, they may be able to solve a traditional problem without any understanding of the underlying concepts [Elby 1999; Mahajan 1999; Mazur 1998].

Validating the test against existing assessment tools

Classical test theory argues that students' performance on individual questions should mirror their performance on the test as a whole. That is, for a test that is considered valid, any new questions added to the test should show the same pattern: students performing well on the test as a whole should score well on the new questions. In this way, questions that are answered correctly by lower-performing students but incorrectly by higher-performing students are unacceptable [Sadler 2000]. This makes it especially challenging to include substantially different types of questions on a standard assessment tool because they may not track with the rest of the test. For instance, when conceptual questions were added to the Advanced Placement physics exam, they proved inconsistent with the students' performance on items from an old test [E. Mazur, pers. comm.; Sadler 2000]. The difficulty here is not that the new questions are uninformative or misleading, even though they do not agree with the picture given by the rest of the test. Rather, the very structure of the exam is at fault and that consistency between traditional and experimental modes of assessment is *not* the desired outcome; in these instances, it is important to give up the expectation that there will be a meaningful correlation between the traditional and new assessments [Mazur 1998]. That is, it may be the original test which is problematic and not the new questions.

Multiple-choice questioning

Finding out what a student really thinks may be best accomplished by engaging that student in conversation. The technique of "Interviews about Instances" has been recommended in order to get a student to discuss his or her ideas orally [Bell *et al.* 1985]. The problem is that interviewing is very labor-intensive and thus hard to conduct on a large scale. Open-ended written questions may be the next most obvious method of assessing student understanding of scientific concepts since they allow students to show what they know without guiding them too far on the thinking process. Again, the difficulty here is that it is time-consuming and difficult to score student responses on a large scale. And it runs the danger that the tests are assessing students' reading and writing abilities rather than their understanding of science [Cohen 1980].

Multiple-choice tests provide the simple solution since they can be easily and quickly scored for large numbers of students, even automatically with the help of technology. However, they are not without their faults. O'Brien Pride *et al.* [1998] compared students' performance on the multiple-choice Mechanics Diagnostic Test [Halloun and Hestenes 1985a] to their own in-depth open-ended examination of student learning and found a large discrepancy between the two assessment instruments. They suggest that it can be misleading to use a multiple-choice assessment instrument as the sole determiner of student understanding, since it may be fooled by students who choose the right answer for the wrong reason, by students who guess, and by students who use strategies other than understanding in determining which answer they will select. Steinberg and Sabella [1997] note a similar result in observing students who provided incorrect answers on their free-response examination, but that would have been considered correct on the multiple-choice FCI as well as free-response answers that did not resemble any of the multiple choices. To be sure, the ability to recognize their understanding from a set of distractors called for on a multiple-choice test is different from the ability to generate one's own view, but is much more easily scored and compared than open-ended questions [Sadler 1992, 1998, 2000].

Appropriate selection of distractors can help mitigate some of the problems. For instance, distractors can be obtained from student interviews, open-ended or free-response questions, and from the research literature on misconceptions. Drawing upon actual student responses has the advantage of using student vocabulary, without relying on jargon, caveats, or detail of choices made up by the test author. Seemingly nonsensical answers to experts might be quite appealing to students. In fact, it seems that many of the most tempting distractors "make no sense" to those experienced in the field [Sadler 2000].

Test writers tend to be "exceedingly careful" in constructing the right answer to a question, in such a way that it is often possible to identify the correct answer without even understanding the question [Sadler 2000]. Thus, it is useful to pose both the correct and incorrect answers with the same degree of jargon, same length, and same level of caveats. As long as a scientist would select the "correct answer," it is not essential to include every subtlety and nuance. In fact, scientists are often not the best judge of distractors since they reject answers which "make no sense" yet these non-sensical answers—drawn from interviews or free-response answers—may, in fact, be those selected most often by students [Sadler 1992, 2000].

Tests can also be constructed so that they separately ask for a students' *prediction* and his or her *reason* for selecting that answer, each of which is a separate multiple-choice question [Tan *et al*. 2002; Treagust 1986, 1988]. Ease of administration and standardized answers provide a balance for the inability of uncovering unexpected misconceptions, making multiple-choice testing the method most appropriate for large-scale testing of students [Sadler 1992].

Sadler [1992] examined incorrect students answers to distinguish random guessing from true misconceptions. In particular, if students were truly guessing at the correct answer, they would be expected to select each of the multiple-choice answers approximately the same percentage of the time, while misconceptions would show a different distribution of answer choices [Sadler 1992]. Nazario *et al*. [2002] formalize this idea; they suggest that one may be able to identify student misconceptions by looking not only at the correctness of student answers but also at what they term the "misconception index" (MI). The MI, in turn depends upon the percentage of students who respond with a single "most frequent incorrect answer" (MFIA), such that

$$\mathrm{MI} = 1 - \frac{(\#\ \text{students with incorrect answer}) - (\#\ \text{students with MFIA})}{(\#\ \text{students with incorrect answer})}$$

[Nazario *et al*. 2002]. The MI is a scalar with values between 0 and 1, with a higher MI indicating a greater percentage of students answering incorrectly who responded with the single MFIA. By looking at the MI, Nazario *et al*. [2002] suggest that, by considering only questions that have a MI above a given cutoff (Nazario *et al*. use MI ≥ 0.3 for the pretest and MI ≥ 0.4 for the posttest), it is possible to remove questions from analysis that may be due to memorization of factual knowledge instead of real conceptual understanding.

Experimental overview and motivation

As discussed above, vocabulary can be a significant stumbling block to novice students in a subject area. One of the difficulties in studying biology is that it is difficult to discuss the subject without relying on a great deal of specialized language.

The vocabulary survey described in the remainder of this chapter performed in order to determine what terminology is already familiar to students entering an introductory college genetics course. My original motivation for the survey was the inform the design of a test of conceptual understanding in introductory biology. In so doing, however, I discovered that there is a much interesting to be learned from studying the vocabulary itself. In particular, I found that familiarity with terms varies widely among students, and that students do not always have an accurate sense of their own understanding.

Biological Sciences 50: Genetics and Genomics

The 2000-2001 academic year saw the introduction of a new sequence of courses in the Biological Sciences at Harvard University. These courses, jointly administered by the Department of Molecular and Cellular Biology and the Department of Organismic and Evolutionary Biology within the Faculty of Arts and Sciences serve as the introductory sequence for students majoring in Biology or Biochemical Sciences [Faculty of Arts and Sciences 2000a]. In the new curriculum, all students are expected to begin with Biological Sciences 50 (Genetics and Genomics) and 51 (Integrative Biology of Organisms) before moving on to more advanced courses in either the Biology or Biochemical Sciences majors.

Biological Sciences 50 (BS 50) is envisioned as the first college course in biology for all students, regardless of preparation or previous experience in genetics. Genetics and genomics are topics that "permeate all of biology" and which provide a framework for future coursework in the biological sciences [Faculty of Arts and Sciences 2000b]. During the Spring 2001 semester described in this chapter, the course was taught by Professor Daniel L. Hartl and enrolled 329 students: 326 undergraduates, 1 student in the Graduate School of Arts and Sciences, and 2 others.

Students in the course have a broad range of previous experience in biology. Some scored 5 on the Advanced Placement Biology test in high school while others had not studied biology at all since the ninth grade or earlier. Some students in the course had never even picked up a test tube, while others had significant laboratory experience. This diversity made BS 50 an ideal setting in which to assess the range of student understanding and preparation for college biology. Because BS 50 is the first or second college biology course for many students (see below), it may also serve as a gatekeeper for students interested in the biological sciences and genetics.

The course has no official prerequisites and is essentially required for all undergraduates with concentrations in biology and biochemical sciences as well as those students fulfilling premedical requirements who have other concentrations. The entry for BS 50 in the course catalog [Faculty of Arts and Sciences 2000a] appears as:

> **Biological Sciences 50. Genetics and Genomics**
> Catalog Number: 9370
> *Daniel L. Hartl*
> *Half course (spring term). M., W., F., at 12 and three hours of laboratory/discussion each week. EXAM GROUP: 5*
> For 2000-2001 academic year, this course is only offered during the spring term. Analysis of genes and genomes with emphasis on function, transmission, mutation, and evolution, with examples from animals, plants, bacteria, and fungi. Discusses classical and current methods of gene and genome analysis, including genetic, molecular,

> quantitative, and bioinformatic approaches. For current Biology and Biochemical Sciences concentrators, this course may be taken in lieu of Biological Sciences 1. Please refer to the respective concentration notes for additional information on the new course sequence.
> *Note:* Lectures and weekly laboratory/discussion section. For 2000-2001 academic year, this course is only offered during the spring term.

In addition to three 50-minute lecture periods per week, the course students participated in a weekly discussion/laboratory section of 16-18 students each led by a Teaching Fellow, generally a graduate student or postdoctoral fellow in the biological sciences. During these sections, students conducted laboratory exercises and the instructor reviewed the content covered in lecture. The survey discussed in this chapter was administered by some of the Teaching Fellows in their sections.

Methodology

A brief survey was prepared to assess students' background and familiarity with vocabulary used in genetics and biology (Appendix E). Eighty-seven terms from the glossary of the course text—*Genetics: Analysis of Genes and Genomes* [Hartl and Jones 2001]—were selected and students asked to rate their familiarity (*f*) with the terminology on a 1-5 scale:

1 = Totally unfamiliar
2 = Word is vaguely familiar
3 = Hazy understanding
4 = Okay understanding
5 = Confidence understanding

In addition, two terms with no known meaning in the biological sciences ("clastron" and "spooling") were included to serve as a control; for instance, a student indicating a familiarity of 3 with clastron might suggest that he or she was less likely to suggest a low familiarity at all.

Students were also asked to briefly define 18 of the vocabulary terms in their own words, including the distractor "spooling." Students' understanding of their vocabulary was also graded on a 1-5 scale and is represented as u. Items left blank were given a score of 1 (discussion of whether assigning blank responses a score of 1 biases the results appears below).

The survey also included several questions designed to address concepts covered in the course in the areas of molecular and population genetics as well as background information on the students. In particular, there was one multiple-choice question designed to address a common misconception about the nature of genetic material, one free-response question, and one three-part data interpretation question (the complete survey is available in Appendix E).

Finally, on the pre-test version of the survey, students were asked to provide information about their previous experience and preparation in the sciences, such as their high school and college science coursework, actual or expected major, and various demographic variables. On the post-test, students were asked to evaluate the course and their experience through the semester. In addition to informing this research study, this feedback was shared with the course instructor.

Only some of the Teaching Fellows chose to administer the survey to their section(s). As such, only some students completed the survey. However, the choice of whether or not to complete the survey was not made by individual students but by their sections leader. One could easily imagine that an individual student's motivation and interest in the course would influence his or her participation in such a survey (such self-selection might bias the sample population). But selection by the Teaching Fellow should be essentially independent of the motivation and ability of the students in that section. Thus, as the selection was done at the level of the *section*—rather than the level of individual *student*—it is likely that any selection bias is unrelated to the ability or level of motivation of the student.

Identifying non-serious responses

Students were asked to provide their name on the survey so that we could connect their responses from the pre- and post-tests, but their performance on and participation in the survey had no effect on their grade. One might ask if students took the survey seriously even though they were not graded on it. A useful comparison can be provided by Henderson [2002] who considered the role of a voluntary administration of the FCI. He looked at the percentage of students who clearly did not take the exercise seriously—as indicated by refusing to take the test, leaving many questions blank, or answering in patterns which suggest no relation to content (*e.g.*, AAAAA, ABCDABCD, or even drawing a "picture" by filling in circle on the response form). When it was not graded, Henderson found 2.8% of students did not take the test seriously, but only 0.6% when the test was graded. To be sure, both of these numbers are quite small, so most students do seem consider the test seriously; further, it may be easy to eliminate the responses that suggest obvious non-serious completion [Henderson 2002]. Steinberg and Sabella [1997] also observed students' taking a voluntary administration of the FCI seriously by looking at how long students took to complete the test; they found that almost all students stayed for a "reasonable" length of time, suggesting thoughtful consideration of the answers.

In the vocabulary survey described here, all student responses to the familiarity questions were included in the analysis as there were no obvious non-serious responses. Even though some students rated their familiarity as 5 for each team, it is possible that those students really *do* think they have a confident understanding of each of the terms. The two control terms provide some sense on whether students were completing the questionnaire thoughtfully or not. However, students may have truthfully *thought* that they understood these terms, even though they were made up; thus, even reporting a confident understanding for made-up words may not indicate that the survey was not completed honestly.

The problem is somewhat more complicated for the free-response definitions used to assess actual understanding. Unfortunately, it is impossible to tell if items left blank were due to a lack of information about the meaning or by a student who simply chose not to complete the definition for that word. These are quite different reasons for blank responses and it is not clear how to most accurately address this distinction within the current data set. One option would be to ignore all blank responses. However, this would significantly underestimate student understanding since it only considers responses from those students who think that they do understand the term and therefore attempted to provide a definition. Especially at the beginning of the semester, many students would leave many of the definitions blank because they really do not know how to define those terms. For this analysis, I have chosen to score blank responses as 1, indicating lack of understanding, though this is surely biased in the other direction.

I did, however, only consider definitions for students who had completed the remainder of the survey. Students who left *all* free-response definitions blank might suggest that they simply chose not complete that section of the survey; again, however, it is possible that they really had no idea about the definition, especially on the pretest. More typical, however, was a situation in which a student completed some of the definitions and not only, say, the first few asked. One solution might be to compare whether students provided a definition to their self-reported familiarity; for instance, one might only expect a definition when a student rated their familiarity as 3 or above (though it is not clear what the most appropriate cut-off for this would be and to what degree it might differ between students). Perhaps the best solution would have been to ask students to explicitly respond if they did not know to provide a definition, perhaps by means of a check-box or other easy-to-respond method.

Student demographics

Of the 187 students[**] who responded to the pretest, three-quarters of them were first-year students and more than 60% were female; these data are consistent with the official enrollment for the course. Seventy-three percent expected to complete the pre-medical requirements. This course was the first college biology course for 85% of the students; an additional 8% had previously taken only Biological Sciences 51, the introductory course in organismal biology that has very little overlap with BS 50. More broadly, 17% had taken no previous college-level science course of any type (including chemistry, physics, or science-related areas of Harvard's Core curriculum); an additional 57% of students had taken only one previous science course at Harvard, of which one of several introductory chemistry courses was the most common previous preparation. In

[**] out of a total course enrollment of 329 students

total, nearly three-quarters of the respondents were taking their first or second college science course.

Most students were planning to major in a life-sciences field. Of those who named a single major (n = 158), Biology was the most common, named by about one-quarter of respondents (43 out of 158). Biochemical Sciences was also quite common, identified by 30 students (19%). Also quite popular were neurosciences; thirty-two students (20%) mentioned Psychology, Neuroscience, Mind/Brain/Behavior, or some variant. About 8% of students planned to major in History and Science, followed by Government (4%), Chemistry (3%), and Computer Science, English, and Physics (2% each).

Students had taken an average of 1.60 ± 0.67 years of biology in high school (Figure 4.1*a*). Fifty-nine percent (110 of 187) of the students had taken the Advanced Placement (AP) examination in biology with a mean score of 4.7[††] among those self-reporting their score (Figure 4.1*b*). This means that over half of the students in the class had high-school experiences that, at many colleges, would have awarded them course credit or placed them out of the introductory college biology course. To be sure, the quality of students' pre-college preparation in biology in this course far exceeds that for average beginning college students nationally. As such, instructors in other settings should be even more cautious of their students' prior understanding than those of the students described in his chapter.

[††] AP scores are integers that range from 1-5 with 5 as the top score, equivalent to the average score of college students who received grades of A in college.

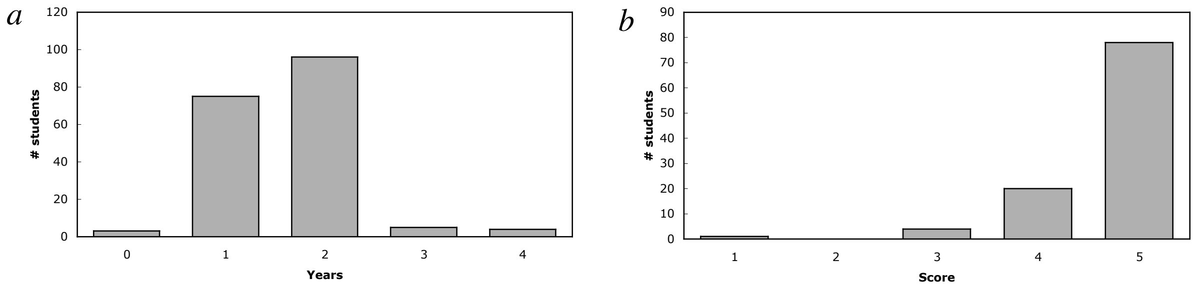


Figure 4.1. (*a*) Years of high school coursework in biology among survey respondents (*n* = 187). Mean years of high school biology is 1.60 ± 0.67. (*b*) Self-reported scores of students on Advanced Placement examination in Biology among students who reported a score (*n* = 103; 7 students took the AP examination but did not provide a score). Mean score is 4.7 (out of 5).

Students had similarly taken an average of 1.56 ± 0.69 years of high school chemistry (Figure 4.2*a*). Forty-three percent (82 of 187) of the students had taken the AP chemistry examination with a mean score of 4.5[††] among those reporting a score (Figure 4.2*b*).

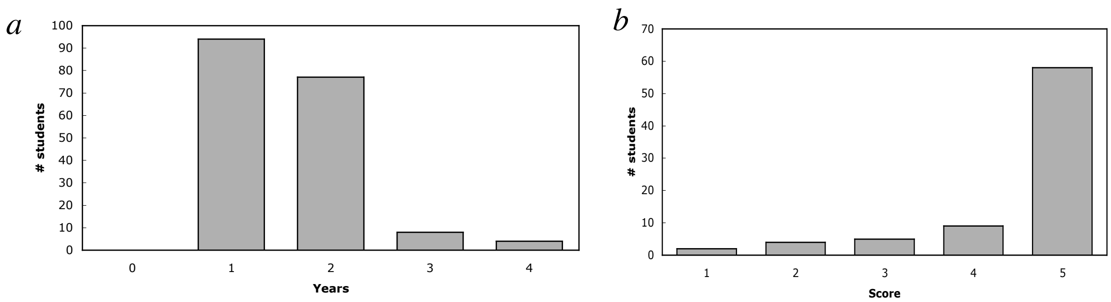


Figure 4.2. (*a*) Years of high school coursework in chemistry among survey respondents (*n* = 187). Mean years of high school chemistry is 1.56 ± 0.69. (*b*) Self-reported scores of students on Advanced Placement examination in Chemistry among students who reported a score (*n* = 78; 4 students took the AP examination but did not provide a score). Mean score is 4.5 (out of 5).

Students were a bit less likely to have had as extensive physics course work with a mean of 1.21 ± 0.71 years in high school (Figure 4.3*a*). About one-third (60 out of 187) of students took the AP physics examination with a mean score of 4.4†† among those reporting their score (Figure 4.3*b*). As with AP preparation in biology, the students in the course described here are far above the average beginning college student, both in their high school coursework and in their performance on standard assessment like the AP examination.

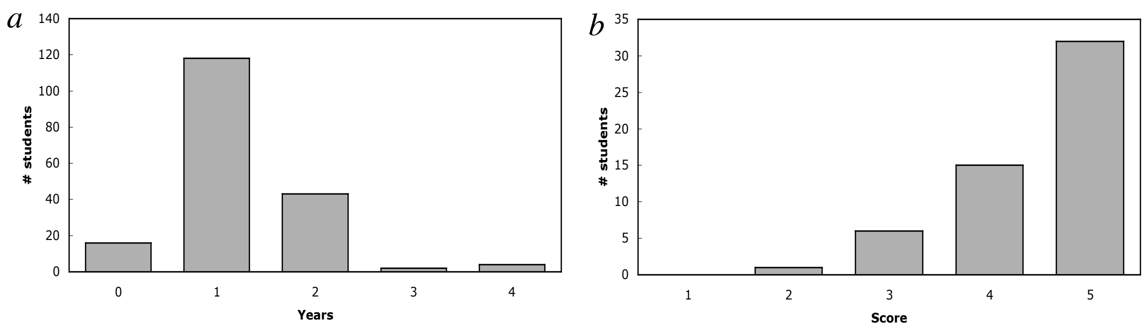


Figure 4.3. (*a*) Years of high school coursework in physics among survey respondents (n = 187). Mean years of high school physics is 1.21 ± 0.71. (*b*) Self-reported scores of students on Advanced Placement examination in Physics among students who reported a score (n = 55; 5 students took the AP examination but did not provide a score). Mean score is 4.4 (out of 5).

Pretest

The pretest immediately shows that previous familiarity with the terminology is widely distributed among the class (Figure 4.4). As described above, the survey asked students to self-identify their familiarity with various terms and to actually define a subset of those terms. Students have a pretty good idea of their understanding. That is, there is a correlation between students' familiarity (f_{pre}) and their understanding (u_{pre}). Figure 4.5 shows familiarity plotted vs. assessed understanding, for each student who completed the pretest. Fitting a line to these data reveals a $R^2 = 0.55$, suggesting a overall correlation between students' self- and actual understanding.

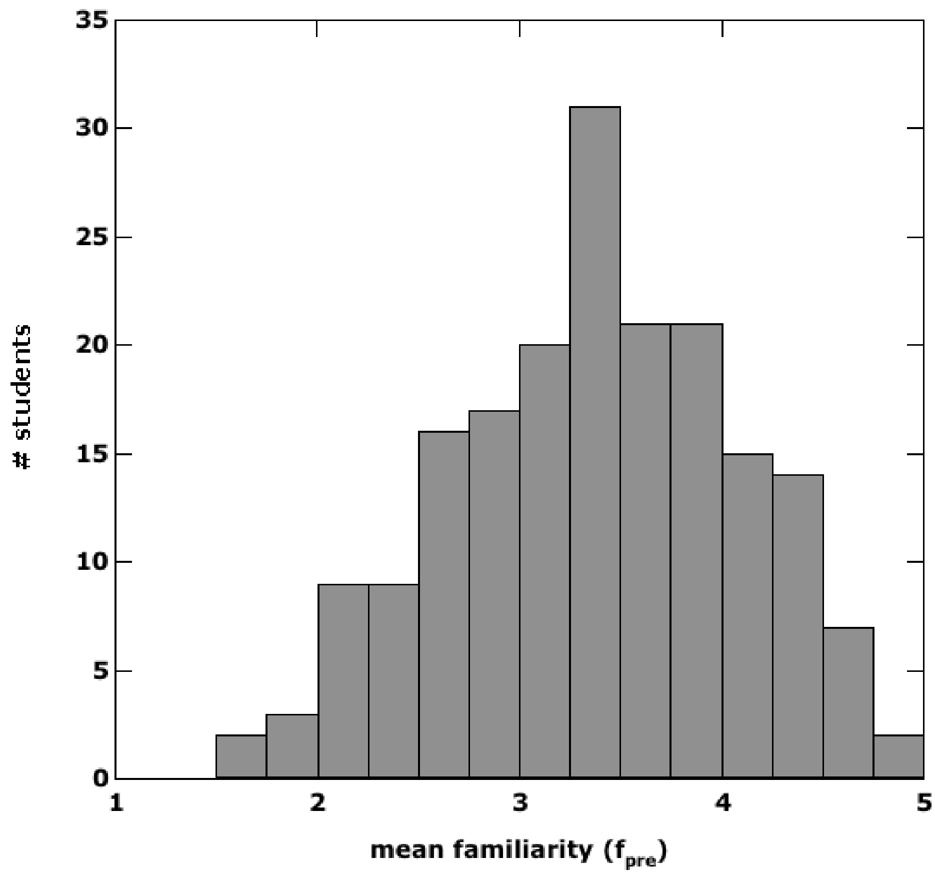


Figure 4.4. Students' mean vocabulary familiarity (f_{pre}) for all 89 terms. Scale is 1-5, with 5 indicated "confident understanding" as described in the text (n = 186).

However, there is not always a clear correspondence between students' familiarity with terminology and their actual understanding (Table 4.1). This is consistent with Evans [1978], who concluded that the ability to recognize biological terms was, to some extent, independent of the ability to understand the concepts they described. For instance, most students feel that they understand the meaning of the term "genetic code" with students rating their familiarity (f_{pre}) at a mean of 4.5 ± 0.7 (5 indicates "confident understanding," the highest on the scale). However, students' performance on the free-response definitions shows a different story (Figure 4.6). Here, students average an understanding (u_{pre}) of 2.6 ± 1.4 (on a 5-point scale, with 5 as the maximum). That is, there is a disconnect between what students *think* that they understand and what they *actually* understand. This has important consequences for instruction: even students who seem comfortable with the material may not actually be mastering it. Thus, it is important to have some mechanism for assessing whether students actually are understanding the material correctly.

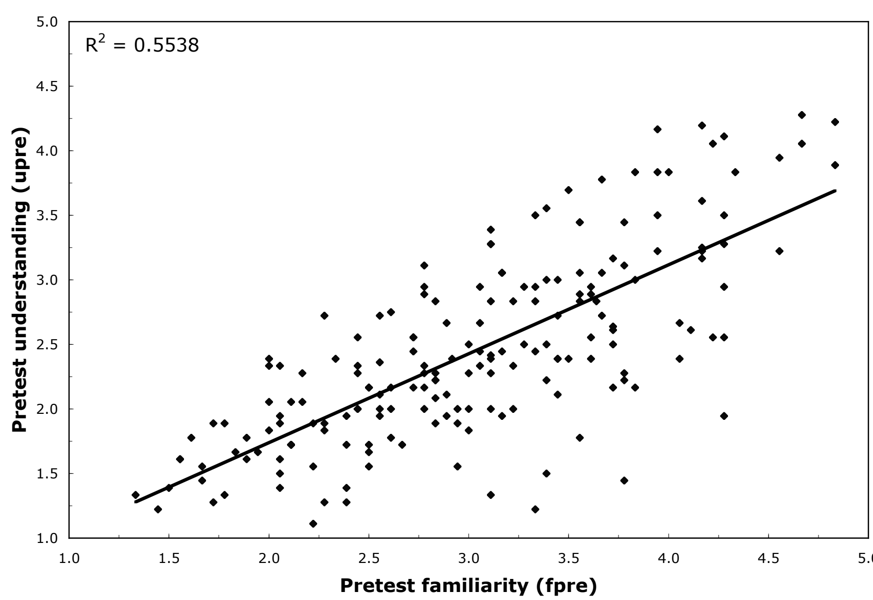


Figure 4.5. Mean vocabulary familiarity (f_{pre}) plotted against mean assessed understanding (u_{pre}) on the pretest. Each data point represents a single student who completed the pretest. The linear regression line is also plotted (n = 186).

Table 4.1. Comparison between students' self-reported familiarity (f_{pre}) and understanding (u_{pre}) evaluated via a free-response definition on the pretest. The mean ± standard deviation is given for each term.

	self-reported familiarity (f_{pre})	assessed understanding (u_{pre})	difference ($f_{pre} - u_{pre}$)
anticodon	4.0 ±□.2	3.6 ± 1.5	0.4
autosomes	2.6 ±□.3	1.9 ±□.5	0.7
cell fate	2.1 ±□.2	1.7 ± 1.3	0.4
consensus sequence	1.5 ±□.8	1.2 ± 0.7	0.3
eukaryote	4.5 ±□.8	4.3 ± 1.3	0.2
gamete	4.1 ± 1.1	4.2 ± 1.5	□0.1
genetic code	4.5 ± 0.7	2.6 ± 1.4	1.9
genome	4.5 ± 0.7	4.1 ± 1.2	0.4
intron	2.5 ± 1.5	2.1 ± 1.5	0.4
linkage	3.2 ± 1.3	2.7 ± 1.7	0.5
nonsense mutation	2.5 ±□.4	1.6 ± 1.1	0.9
operon	2.4 ±□.2	1.6 ± 1.1	0.8
primer	3.6 ± 1.2	2.7 ± 1.5	0.9
semiconservative replication	2.2 ±□.5	2.0 ± 1.6	0.2
spooling	2.5 ±□.3	1.0 ± 0.0	1.5
sticky end	3.4 ± 1.5	3.2 ± 1.7	0.2
telomere	2.6 ±□.4	1.7 ±□.4	0.9
transposable element	2.1 ± 1.2	1.9 ± 1.5	0.2
Overall			0.6 ± 0.5

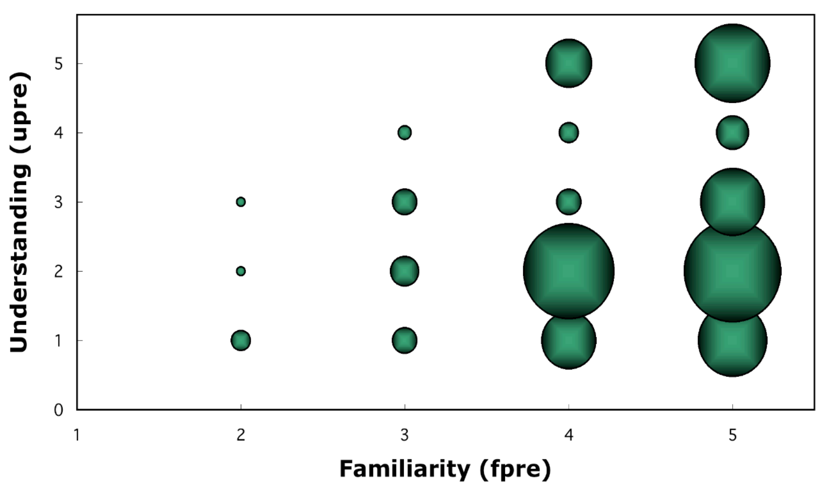


Figure 4.6. Students' self-reported familiarity with the term "genetic code" *vs*. understanding evaluated via a free-response definition on the pretest (f_{pre}). Maximum score is 5 in each case, as described in the text. Size of each point represents the number of students with that combination of familiarity and understanding (n = 187).

Post-test

The same instrument was also administered to students at the end of the semester, with 106 students completing the post-test. Sixty-one of these had also completed the pretest and, therefore, allowed matching the pre- and post-test responses. Overall, the students responding to the post-test show a much enhanced familiarity with the vocabulary as well as an enhanced ability to define those terms. If we compare the students' familiarity (f_{post}) to their assessed understanding (u_{post}), we see a much weaker correlation than we did for the pretest (Figure 4.7). Here the R^2 for the linear regression line is only 0.11, suggesting that students do not have as clear an idea of their understanding as they did on the pretest. In particular, students still assort themselves based on self-rated familiarity, but that distribution is not correlated with their actual understanding.

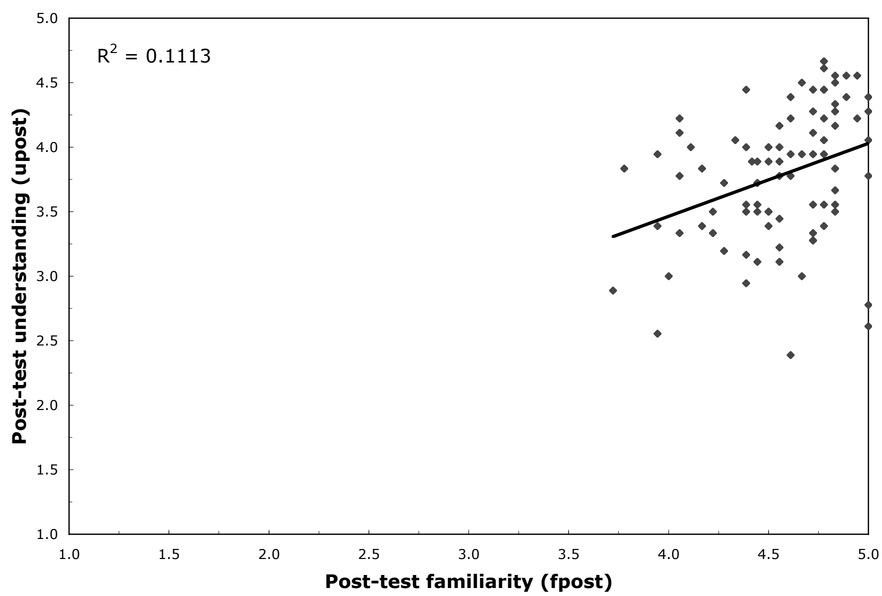


Figure 4.7. Mean vocabulary familiarity (f_{post}) plotted against mean assessed understanding (u_{post}) on the post-test. Each data point represents a single student who completed the post-test. The linear regression line is also plotted (n = 87).

As before, we can also look at the difference between familiarity (f_{post}) and understanding (u_{post}) for each term (Table 4.2).

Table 4.2. Comparison between students' self-reported familiarity (f_{post}) and understanding (u_{post}) evaluated via a free-response definition on the post-test. The mean ± standard deviation is given for each term.

	self-reported familiarity (f_{post})	assessed understanding (u_{post})	difference ($f_{post} - u_{post}$)
anticodon	4.9 ± 0.3	4.2 ± 1.0	0.7
autosomes	4.7 ±□.7	3.8 ± 1.6	0.9
cell fate	3.6 ±□.2	2.4 ± 1.3	1.2
consensus sequence	3.1 ± 1.5	2.4 ±□.7	0.7
eukaryote	4.9 ±□.2	4.5 ± 0.9	0.4
gamete	5.0 ± 0.1	4.8 ± 0.7	0.2
genetic code	4.9 ±□.3	3.7 ± 1.5	1.2
genome	5.0 ±□.2	4.4 ±□.9	0.6
intron	4.7 ±□.6	4.0 ± 1.2	0.7
linkage	4.9 ±□.3	4.1 ± 0.9	0.8
nonsense mutation	4.7 ± 0.6	3.8 ±□.4	0.9
operon	4.8 ± 0.4	3.7 ±□.1	1.1
primer	4.8 ± 0.5	3.7 ±□.0	1.1
semiconservative replication	4.6 ± 0.7	4.2 ± 1.3	0.4
spooling	3.2 ± 1.2	1.0 ±□.0	2.2
sticky end	4.8 ± 0.5	3.9 ±□.0	0.9
telomere	4.7 ±□.6	4.4 ± 1.3	0.3
transposable element	4.5 ± 0.7	4.3 ± 1.1	0.2
Overall			0.8 ± 0.5

Turning again to the term "genetic code," we find that students have a greater degree of understanding than they did at the beginning of the semester. Students are more confident than they were before, with an average familiarity (f_{post}) of 4.9 ± 0.3 (out of 5). But now, the assessed understanding (u_{post})—by way of a free-response definition—has increased as well, averaging 3.7 ± 1.5, out of a maximum of 5. To be sure, there is still room for improvement, but the gap between familiarity and understanding has decreased.

Students do still hold misconceptions about many concepts, even after instruction. For instance, more than 10% of students do not know that genetic material in one human somatic cell is identical to that in another human somatic cell. This is one of the most fundamental concepts in genetics yet a non-negligible fraction of students have not understood this, even at the end of the semester.

Comparing pre- and post-tests

We have matched pre- and post-test data for 61 students, meaning that we can compare the performance of individual students on the pre- and post-tests. Initially, we might ask if a student's initial familiarity with a term (f_{pre}) is correlated with his or her familiarity at the end of the semester (f_{post}). Figure 4.8 shows that there is such a correlation with a linear regression line showing an R^2 of 0.39.

A similar trend is shown in comparing a student's pre- and post-test understanding (u_{pre} vs. u_{post}). Figure 4.9 is a plot of understanding for each student who completed the definitions on both the pre- and post-test. The linear regression line here shows a similar correlation to that of familiarity with an R^2 of 0.42. In absolute terms, therefore, a student who enters the class with a greater degree of familiarity or understanding is likely to exit the course with a similarly elevated level of familiarity or understanding.

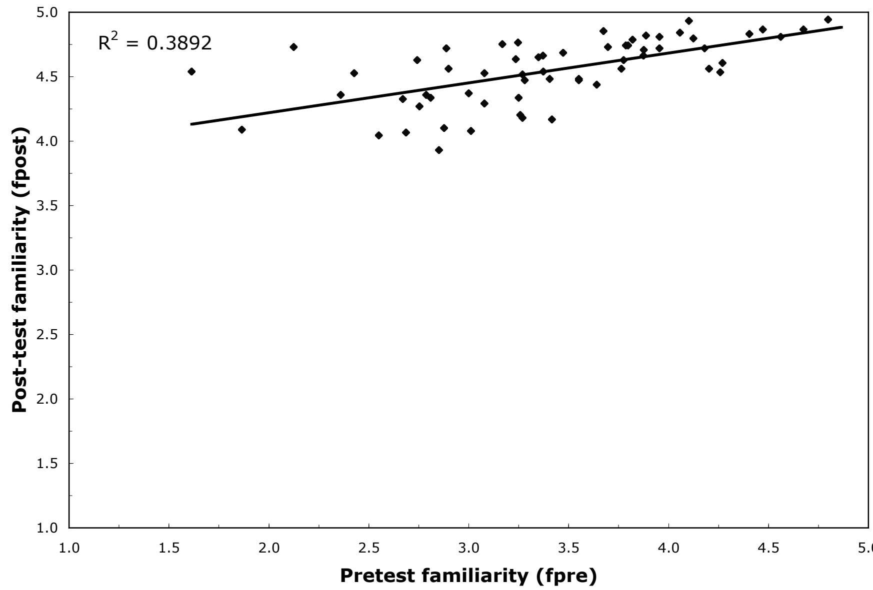


Figure 4.8. Matched pre- (f_{pre}) and post-test scores (f_{post}) for familiarity. Each data point represents a single student who completed both the pre- and post-tests. The linear regression line is also plotted ($n = 61$).

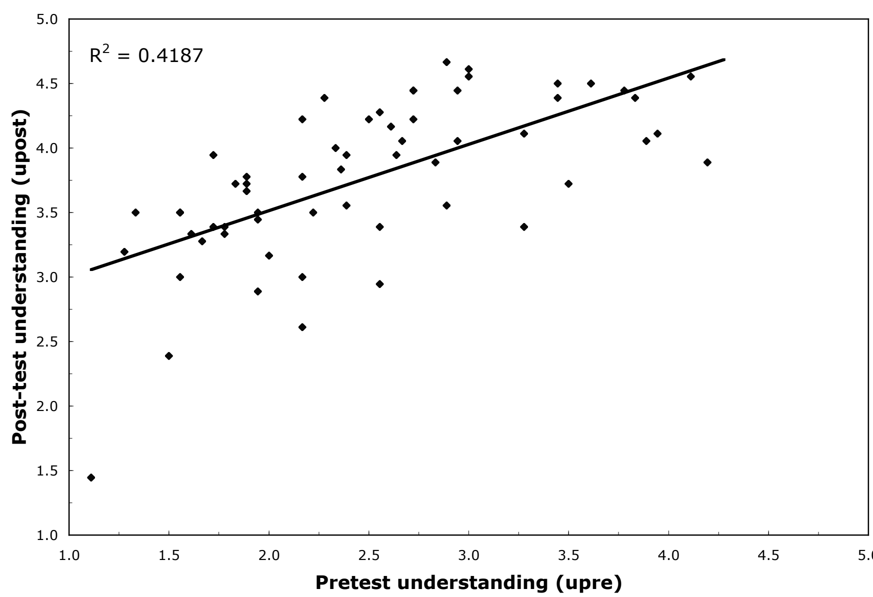


Figure 4.9. Matched pre- (u_{pre}) and post-test scores (u_{post}) for understanding. Each data point represents a single student who completed the definitions on both the pre- and post-tests. The linear regression line is also plotted ($n = 60$).

In order to remove the effects of student ability from this equation, we can consider the normalized gain, as proposed by Richard Hake [1998]. As described in Chapter 2, the normalized gain looks at how much of a student's possible gain is achieved by the end of the course. So, for instance, if a student performed poorly on a pretest, he or she has more potential to gain in absolute terms. In this case, normalized gain is computed by calculating the percentage of the possible gain that is achieved (the 5 in the denominator reflects the maximum possible score, since both familiarity and understanding were assessed on a 1-5 scale):

$$g_f = \frac{f_{\text{post}} - f_{\text{pre}}}{5 - f_{\text{pre}}} \quad \text{and} \quad g_u = \frac{u_{\text{post}} - u_{\text{pre}}}{5 - u_{\text{pre}}}$$

Figure 4.10 shows the normalized gain for familiarity plotted against the pretest familiarity f_{pre}. The linear regression line shows that there is essentially no correlation between a student's initial familiarity and the normalized gain in familiarity they achieve (R^2 = 0.03). Figure 4.11 shows the same plot for understanding. Again, a linear regression line shows no correlation between a student's initial understanding (u_{pre}) and his or her normalized gain in understanding (R^2 = 0.03). These plots demonstrate that there is no correlation between how well students think—or actually—know the vocabulary and how much they benefit over the course of the semester. That is, a student with weak understanding coming in to the class will improve by about the same degree relative to the stronger student, consistent with the results shown by Hake [1998] on the FCI.

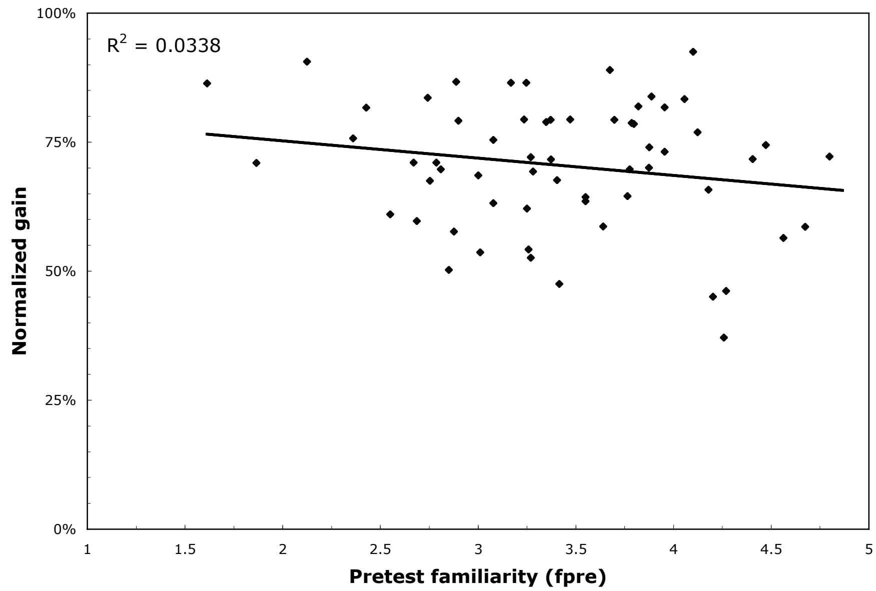


Figure 4.10. Normalized gain for familiarity, following the method of Hake [1998], as described in text. Each data point represents a single student who completed both the pre- and post-tests. Pretest familiarity (f_{pre}) is represented as a percent. The linear regression line is also plotted ($n = 61$).

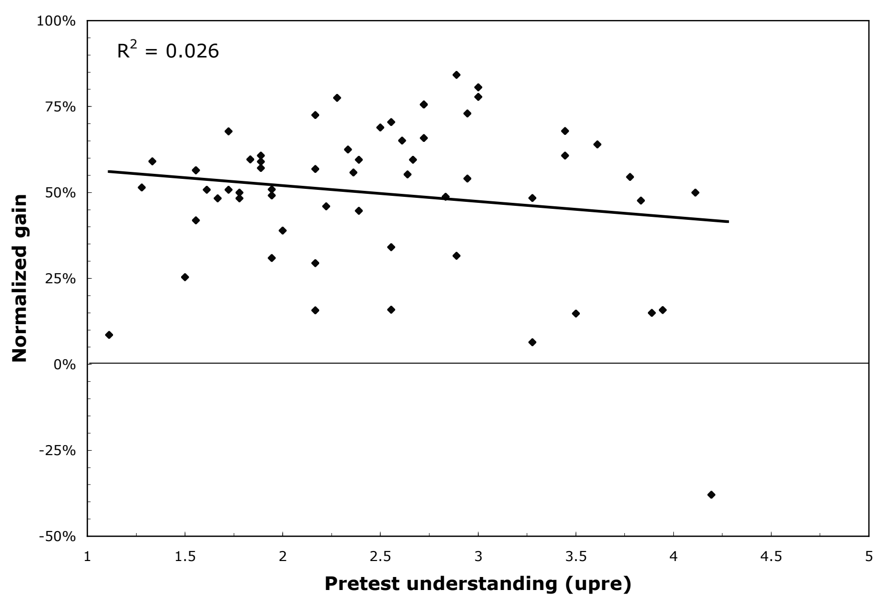


Figure 4.11. Normalized gain for understanding, following the method of Hake [1998], as described in text. Each data point represents a single student who completed the definitions on both the pre- and post-tests. Pretest understanding (u_{pre}) is represented as a percent. The linear regression line is also plotted ($n = 60$).

It should be noted that one student actually shows a negative gain, that is a *decrease* in understanding (u_{pre} = 4.2; u_{post} = 3.9). This particular example speaks to the appropriateness of considering a blank to be regarded as a lack of understanding; for example, this student provided a fully correct definition for “consensus sequence” on the pretest but left that term blank on the post-test, even though he completed each of the other definitions.

Student overconfidence

One interesting result arises if we consider the difference between students’ familiarity (*f*) and understanding (*u*) as a measure of the students’ overconfidence. This provides a measure of the degree to which their *perceived* understanding exceeds their *actual* understanding (as judged by a grader), when both are rated on a 5-point scale. Looking at just the pre- or post-test, there is a question about how the grader’s scale for understanding compares with the student’s scale for assessing familiarity. But if we compare this difference between the pre- and post-tests, this question is resolved. What we observe (Table 4.3 and Figure 4.12) is that for almost all of the 18 terms students were asked to define (“genetic code” and “telomere” are the only exceptions), there is a greater understanding gap in the post-test as there is in the pretest; in other words, students rate their gain in understanding ($f_{post} - f_{pre}$) more significantly than their actual gain in understanding ($u_{post} - u_{pre}$). So students *think* that their understanding has increased more than it actually has.

Table 4.3. Increase in students' overconfidence in their understanding of key vocabulary terms. Show the change in the difference between students' familiarity (*f*) and understanding (*u*) between pre- and post-tests.

	difference ($f_{pre} - u_{pre}$)	difference ($f_{post} - u_{post}$)	Δ difference (post – pre)
anticodon	0.4	0.7	0.3
autosomes	0.7	0.9	0.2
cell fate	0.4	1.2	0.8
consensus sequence	0.3	0.7	0.4
eukaryote	0.2	0.4	0.2
gamete	−0.1	0.2	0.3
genetic code	1.9	1.2	−0.7
genome	0.4	0.6	0.2
intron	0.4	0.7	0.3
linkage	0.5	0.8	0.3
nonsense mutation	0.9	0.9	0.1
operon	0.8	1.1	0.3
primer	0.9	1.1	0.2
semiconservative replication	0.2	0.4	0.2
spooling	1.5	2.2	0.7
sticky end	0.2	0.9	0.7
telomere	0.9	0.3	−0.6
transposable element	0.2	0.2	0.0
Overall	0.6 ± 0.5	0.8 ± 0.5	0.2 ± 0.4

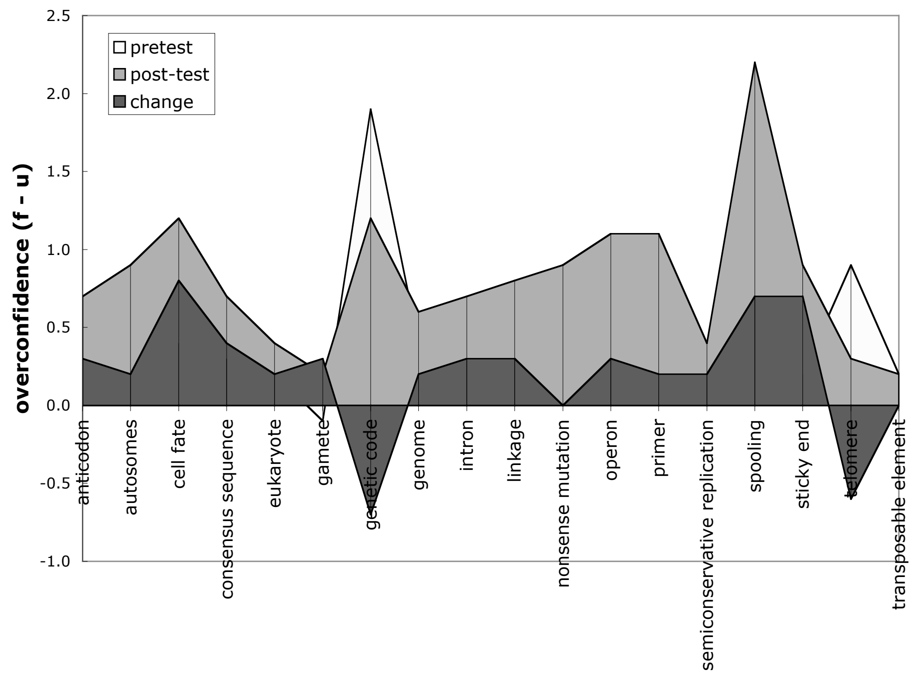


Figure 4.12. Increase in students' overconfidence in their understanding of key vocabulary terms. Show the change in the difference between students' familiarity (*f*) and understanding (*u*) between pre- and post-tests. A positive change in overconfidence means that overconfidence has increased from pre- to post-test.

Conclusion and implications for instruction

The survey shows that the majority of students began Harvard's introductory biology class with significant deficiencies in their understanding of biology terminology. And, while there was improvement by the end of the semester, gaps in understanding can remain even after instruction. Moreover, students believe that they understand more than they actually do, and this overconfidence increases over the course of the semester. That is, students are even more likely to think they understand something after they think they *should* have learned it. Some students may express uncertainty about their preparation for the course when it begins, but even these students increase their confidence in their understanding by the end of the semester. While this increase in confidence is almost certainly a positive step for most students, it can also mean that students may not recognize when they still have significant deficiencies in their understanding.

This study raises caution about assuming that students enter an introductory college genetics course with a thorough understanding of the basic terminology of genetics. Even at an elite institution like Harvard University and even when many students have taken—and scored well on—Advanced Placement examinations, there can be significant gaps in understanding of even some of the most basic genetics vocabulary. This chapter highlights the value of assessing their understanding upon entrance to the class—and continually throughout the term. The comparison with foreign language instruction also suggests the importance of providing students with opportunities to use the new terminology they are learning, perhaps by the use of an interactive pedagogy like Peer Instruction.

In addition, the results suggest the difficulty in using specific terminology in any test of conceptual understanding. Unlike the topics discussed in the FCI that are similar to students' prior life experience, much of biology is foreign to those who have not studied the subject previously. As such, it will be quite challenging to construct a test of conceptual understanding that does not rely upon a student's prior familiarity—and indeed understanding—of the terminology used in the test questions.

Plans are already underway to repeat this study. The revised survey will address concerns about scoring blank definitions and will also incorporate relevant questions from the conceptual test developed by Nazario *et al.* [2002]. It will also be important to consider the consistency of assessing student understanding through the free-response definitions, looking at both inter-rater and same-rater reliability. Ultimately, it would also be valuable to have a more robust assessment of understand (in addition to definitions), such as by coupling the conceptual questions aimed at differentiating understanding of vocabulary. With a larger dataset a repeat of study will likely achieve, it will also be interesting to investigate differences between subpopulations of students, for instance gender and previous coursework in biology.

References

Anderson, Dianne L.; Kathleen M. Fisher; and Gregory J. Norman (2002): "Development and evaluation of the conceptual inventory of natural selection." *Journal of Research in Science Teaching* 39(10), 952–978.

Ausubel, David P. (1968): *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart & Winston.

Bell, Beverley; Roger Osborne; and Ross Tasker (1985): "Finding Out What Children Think." In: *Learning in Science: The Implications of Children's Science*. (Roger Osborne and Peter Freyberg, eds.) Portsmouth, NH: Heineman, 151–165.

Biggs, John B. (1996): "Enhancing teaching through constructive alignment." *Higher Education* 32(3), 347–364.

Bishop, Beth A. and Charles W. Anderson (1985): *Evolution by Natural Selection: A Teaching Module*. (Institute for Research on Teaching Occasional Paper, 91) East Lansing, MI: Michigan State University.

Bishop, Beth A. and Charles W. Anderson (1990): "Student conceptions of natural selection and its role in evolution." *Journal of Research in Science Teaching* 27(5), 415–427.

Bonwell, Charles C. and James A. Eison (1991): *Active Learning: Creating Excitement in the Classroom*. (ASHE-ERIC Higher Education Report, 1) Washington, DC: The George Washington University, School of Education and Human Development.

BouJaoude, Saouma B. (1992): "The Relationship between Students' Learning Strategies and the Change in their Misunderstandings During a High School Chemistry Course." *Journal of Research in Science Teaching* 29(7), 687–699.

Bowen, Craig W. and Diane M. Bunce (1997): "Testing for Conceptual Understanding in General Chemistry." *The Chemical Educator* 2(2).

Boyer Commission on Educating Undergraduates in the Research University (1998): *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*. <http://naples.cc.sunysb.edu/Pres/boyer.nsf/>

Bullock, D.W.; V.P. LaBella; T. Clingan; Z. Ding; G. Stewart; and P.M. Thibado (2002): "Enhancing the Student-Instructor Interaction Frequency." *The Physics Teacher* 40(9, Dec), 535–541.

Burnstein, Ray A. and Leon M. Lederman (2003): "Comparison of Different Commercial Wireless Keypad Systems." *The Physics Teacher* 41(5, May), 272–275.

Callan, J. Paul; Catherine H. Crouch; and Eric Mazur (2000): Classroom demonstrations: education or entertainment? Contributed talk presented at 2000 AAPT Winter Meeting (American Association of Physics Teachers), Kissimmee, FL, January 15-19. <http://mazur-www.harvard.edu/Talks/pdf_files/Talk_328.pdf>

Chase, J.D. and E.G. Okie (2000): Combining cooperative learning and peer instruction in introductory computer science. Presented at Thirty-First SIGCSE Technical Symposium on Computer Science Education (ACM), Austin, TX, 8-12 March. 372–376. (*SIGCSE Bulletin* 31(1, Mar)).

Cohen, H.G. (1980): "Dilemma of the objective paper-and-pencil assessment within the Piagetian framework." *Science Education* 64, 741–745.

Cox, Anne J. and William F. Junkin III (2002): "Enhanced student learning in the introductory physics laboratory." *Physics Education* 37(1, Jan), 37–44.

Crouch, Catherine (1998): "Peer Instruction: An Interactive Approach for Large Classes." *Optics and Photonics News* 9(9, Sep), 37–41.

Crouch, Catherine H. and Eric Mazur (2001): "Peer Instruction: Ten Years of Experience and Results." *American Journal of Physics* 69(9, Sep), 970–977.

Dancy, Melissa Hayes (2000): Investigating Animations for Assessment with an Animated Version of the Force Concept Inventory. Ph.D. Dissertation, North Carolina State University, Raleigh, NC.

Danielson, S. and S. Mehta (2000): Concept questions for improved learning and outcome assessment in statics. Presented at 30th Annual Frontiers in Education Conference: Building on a Century of Progress in Engineering Education (IEEE Education Society), Kansas City, MO, 18-21 October. S3A/1 vol. 2.

Di Stefano, R. (1996): "Preliminary IUPP results: Student reactions to in-class demonstrations and to the presentation of coherent themes." *American Journal of Physics* 64(1, Jan), 58–68.

Driver, Rosalind; Edith Guesne; and Andrée Tiberghien (eds.) (1985): *Children's Ideas in Science*. Milton Keynes and Philadelphia: Open University Press.

Dunbar, Robin (1995): *The Trouble with Science*. London: Faber.

Ebert-May, Diane; Carol Brewer; and Sylvester Allred (1997): "Innovation in large lectures—teaching for active learning." *BioScience* 47(9, Oct), 601–607.

Ege, Seyhan N.; Brian P. Coppola; and Richard G. Lawton (1997): "The University of Michigan Undergraduate Chemistry Curriculum: 1. Philosophy, Curriculum, and the Nature of Change." *Journal of Chemical Education* 74(1, Jan), 74–83.

Elby, Andrew (1999): "Another reason that physics students learn by rote." *American Journal of Physics* 67(7, Phys. Educ. Res., Jul), S52–S57.

Evans, [illegible]. Darryl (1978): "Putting names to concepts in biology." *Journal of Biological Education* 12(4), 261–266.

Faculty of Arts and Sciences (2000a): *Courses of Instruction, 2000-2001* [Official Register of Harvard University, Vol. 22, No. 5]. Cambridge, MA: Harvard University.

Faculty of Arts and Sciences (2000b): *Handbook for Students, 2000-2001*. Cambridge, MA: Harvard University.

Felder, Richard M. and Rebecca Brent (1996): "Navigating the bumpy road to student-centered instruction." *College Teaching* 44, 43–47.

Freier, George (1981): "The Use of Demonstrations in Physics Teaching." *The Physics Teacher* 19(6), 384–386.

Gattis, Kenneth William (1995): An Investigation of the Conceptual Changes Resulting from the Use of Demonstrations in College Physics. Doctoral Dissertation, North Carolina State University, Raleigh, NC.

Gattis, Kenneth W. and John C. Park (1997): Effectiveness of Demonstrations in Facilitating Physics Concept Acquisitions. Paper presented at NARST Annual Meeting (National Association for Research in Science Teaching), March 23.

Gedalof, Allan [illegible]. (1998): *Teaching Large Classes*. (STLHE Green Guide, 1.) Halifax, NS, Canada: Dalhousie University. <http://www.uwo.ca/edo/tlc/theguide.html>

Gray, Peter (1999): *Psychology*. 3rd ed. New York: Worth Publishers.

Green, Paul [illegible]. (2003): *Peer Instruction for Astronomy*. (Prentice Hall Series in Educational Innovation.) Upper Saddle River, NJ: Prentice Hall.

Gunstone, Richard F. and Richard T. White (1981): "Understanding of Gravity." *Science Education* 65(3), 291–299.

Hake, Richard R. (1998): "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses." *American Journal of Physics* 66(1, Jan), 64–74.

Hake, Richard (2002): "Lessons from the physics education reform effort." *Conservation Ecology* 5(2), 28. <http://www.consecol.org/vol5/iss2/art28/>

Halloun, Ibrahim Abou and David Hestenes (1985a): "Common sense concepts about motion." *American Journal of Physics* 53(11, Nov), 1056–1065.

Halloun, Ibrahim Abou and David Hestenes (1985b): "The initial knowledge state of college physics students." *American Journal of Physics* 53(11, Nov), 1043–1055.

Hartl, Daniel L. and Elizabeth W. Jones (2001): *Genetics: Analysis of Genes and Genomes*. 5th ed. Sudbury, MA: Jones and Bartlett.

Healy, C.E. (1994): "Introducing supplemental instruction in engineering." In: *Helping Students to Learn from Each Other: Supplemental Instruction*. (C. Rust and J. Wallace, eds.) Birmingham, UK: Staff and Educational Development Association, 25–30.

Henderson, Charles (2002): "Common Concerns About the Force Concept Inventory." *The Physics Teacher* 40(9, Dec), 542–547.

Hestenes, David; M. Wells; and G. Swackhamer (1992): "Force Concept Inventory." *The Physics Teacher* 30, 141–158.

Hilton, Wallace (1981): "Demonstrations as an Aid in the Teaching of Physics." *The Physics Teacher* 19(6), 389–390.

Holton, Gerald R. (1970): "What is Conveyed by Demonstrations." Chap. 5. In: *Physics Demonstration Experiments*. 2 vols. (Harry F. Meiners, ed.) New York: Ronald Press Company, 61–82.

Hynd, Cynthia R.; J. YeVette McWhorter; Virginia L. Phares; and C. William Suttles (1994): "The Role of Instructional Variables in Conceptual Change in High School Physics Topics." *Journal of Research in Science Teaching* 31(9), 933–946.

Kloss, R.J. (1994): "A nudge is best: Helping students through the Perry scheme of intellectual development." *College Teaching* 42(4), 151–158.

Kovac, Jeffrey (1999): “Student active learning methods in general chemistry.” *Journal of Chemical Education* 76(1, Jan), 120–124.

Kraus, Pamela Ann (1997): Promoting active learning in lecture-based courses: Demonstrations, tutorials, and interactive tutorial lectures. Ph.D. Dissertation, University of Washington, Seattle, WA.

Landis, Clark R.; Arthur B. Ellis; George C. Lisensky; Julie K. Lorenz; Kathleen Meeker; and Carl C. Wamser (2001): *Chemistry ConcepTests: A Pathway to Interactive Classrooms*. Upper Saddle River, NJ: Prentice Hall.

Lawson, A.E.; S.W. Rissing; and S.H. Faeth (1990): “An inquiry approach to nonmajors biology.” *Journal of College Science Teaching* 19, 340–346.

Liebman, Judith S. (1996): “Promote Active Learning During Lectures.” *OR/MS Today* 23(6, Dec), 8.

Lightman, Alan and Philip Sadler (1993): “Teacher predictions versus actual student gains.” *The Physics Teacher* 31(3, Mar), 162–167.

Lyman, F. (1981): “The responsive classroom discussion: The inclusion of all students.” In: *Mainstreaming Digest*. (A.S. Anderson, ed.) College Park, MD: University of Maryland College of Education.

Lyman, F. (1992): “Think-pair-share, thinktrix, thinklinks, and weird facts.” In: *Enhancing Thinking Through Cooperative Learning*. (Neil Davidson and Toni Worsham, eds.) New York: Teachers College Press.

MacGregor, Jean; James L. Cooper; Karl A. Smith; and Pamela Robinson (eds.) (2000): *Strategies for Energizing Large Classes: From Small Groups to Learning Communities*. (New Directions for Teaching and Learning, 81) San Francisco: Jossey-Bass.

Mackenzie, Dana (1997): “New Physics Profs Get a Helping Hand, Too.” *Science* 277(5329, Aug 22), 1031.

Mahajan, Sanjoy (1999): Observations on teaching first-year physics. <http://www.inference.phy.cam.ac.uk/sanjoy/teaching/tc-report> Posted: 28 Sep 1999.

Martin, Deanna C. and David R. Arendale (eds.) (1994): *Supplemental Instruction: Increasing Achievement and Retention*. (New Directions in Teaching and Learning, 60) San Francisco: Jossey-Bass.

Mazur, Eric (1997): *Peer Instruction: A User’s Manual*. (Prentice Hall Series in Educational Innovation.) Upper Saddle River, NJ: Prentice Hall.

Mazur, Eric (1998): “Moving the mountain: impediments to change.” In: *Indicators of Success in Postsecondary SMET Education: Shapes of the Future. Synthesis and Proceedings of the Third Annual NISE Forum*. (Susan B. Millar, ed.) Madison, WI: Wisconsin Center for Education Research, 91–93.

McDermott, Lillian C. (1990): “Millikan Lecture 1990: What we teach and what is learned—Closing the gap.” *American Journal of Physics* 59(4), 301–315.

McDermott, Lillian C.; Peter S. Shaffer; and Physics Education Group (1998): *Tutorials in Introductory Physics*. Preliminary ed. 2 vols. (Prentice Hall Series in Educational Innovation.) Upper Saddle River, NJ: Prentice Hall.

McDermott, Lillian C.; Peter S. Shaffer; and Physics Education Group (2002): *Tutorials in Introductory Physics*. First ed. 2 vols. (Prentice Hall Series in Educational Innovation.) Upper Saddle River, NJ: Prentice Hall.

McNeal, Ann P. and Charlene D’Avanzo (eds.) (1997): *Student-Active Science: Models of Innovation in College Science Teaching*. (Proceedings on the NSF sponsored conference on Inquiry approaches to Science Teaching held at Hampshire College, June 1996) Fort Worth: Saunders College Publishing.

Meltzer, David E. and Kandiah Manivannan (1996): “Promoting interactivity in physics lecture courses.” *The Physics Teacher* 34(2), 72–76.

Meltzer, David E. and Kandiah Manivannan (2002): “Transforming the lecture-hall environment: The fully interactive physics lecture.” *American Journal of Physics* 70(6, Jun), 639–654.

Mintzes, Joel J.; James H. Wandersee; and Joseph D. Novak (1997): “Meaningful learning in science: The human constructivist view.” In: *Handbook of Academic Learning: Construction of knowledge*. (Gary D. Phye, ed.) San Diego, CA: Academic Press.

Mintzes, Joel J.; James H. Wandersee; and Joseph D. Novak (1998): *Teaching Science for Understanding: A Human Constructivist View*. San Diego, CA: Academic Press.

Mintzes, Joel J.; James H. Wandersee; and Joseph D. Novak (eds.) (2000): *Assessing Science Understanding: A Human Constructivist View*. San Diego, CA: Academic Press.

Mintzes, Joel J.; James H. Wandersee; and Joseph D. Novak (2001): “Assessing understanding in biology.” *Journal of Biological Education* 35(3, Summer), 118–124.

Moore, David S. and George P. McCabe (1998): *Introduction to the Practice of Statistics*. 3rd ed. New York: W.H. Freeman and Company.

Mulford, Douglas Robert (1996): An Inventory for Measuring College Students' Level of Misconceptions in First Semester Chemistry. M.S. Thesis, Purdue University, West Lafayette, IN. <http://faculty.pepperdine.edu/dmulford/thesis/title.html>

Mulford, Douglas R. and William R. Robinson (2002): "An Inventory for Alternate Conceptions among First-Semester General Chemistry Students." *Journal of Chemical Education* 79(6, Jun), 739–744.

Murray, Martin H. (1999): "Building Networks through Peer Instruction." *Journal of Professional Issues in Engineering Education and Practice* 125(4, Oct), 159–162.

Murray, R. and J.R. Brightman (1996): "Interactive teaching." *European Journal of Engineering Education* 21, 295–308.

National Research Council (1997): *Science Teaching Reconsidered: A Handbook*. Washington, DC: National Academy Press. (Committee on Undergraduate Science Education) <http://www.nap.edu/catalog/5287.html>

National Research Council (1999): *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology*. Washington, DC: National Academy Press. (Committee on Undergraduate Science Education) <http://www.nap.edu/catalog/6453.html>

Nazario, Gladys M.; Patricia A. Burrowes; and Julio Rodríguez (2002): "Persisting Misconceptions: Using Pre- and Post-Tests to Identify Biological Misconceptions." *Journal of College Science Teaching* 31(5, Feb), 292–296.

Near, Joseph A.; Talmage R. Bosin; and John B. Watkins III (2002): "Inclusion of a competency-based curriculum in medical pharmacology." *Naunyn-Schmiedeberg's Archives of Pharmacology* 366, 26–29.

Norris, Stephen (1985): "The Philosophical Basis of Observation in Science and Science Education." *Journal of Research in Science Teaching* 22(9), 817–833.

Novak, Gregor M.; Evelyn T. Patterson; Andrew D. Gavrin; and Wolfgang Christian (1999): *Just-in-Time Teaching: Blending Active Learning with Web Technology*. (Prentice Hall Series in Educational Innovation.) Upper Saddle River, NJ: Prentice Hall.

Novak, J.D. (ed.) (1987): *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*. Ithaca, NY: Cornell University.

O'Brien Pride, Tara; Stamatis Vokos; and Lillian C. McDermott (1998): "The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems." *American Journal of Physics* 66(2, Feb), 147–157.

Odom, Arthus Louis (1995): "Secondary & College Biology Students' Misconceptions About Diffusion & Osmosis." *The American Biology Teacher* 57(7, Oct), 409–415.

Odom, Arthus Louis and L.H. Barrow (1995): "Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction." *Journal of Research in Science Teaching* 32, 45–61.

Piepmeier, Edward, Jr. (1998): "Use of conceptests in a large lecture course to provide active student involvement and peer teaching." *American Journal of Pharmaceutical Education* 62(3, Fall), 347–351.

Pilzer, Scott (2001): "Peer Instruction in Physics and Mathematics." *PRIMUS* XI(2, Jun), 185–192.

Posner, Herbert B. (1996): "Teaching introductory cell & molecular biology: A historical and empirical approach." *The American Biology Teacher* 58(5), 272–274.

Rao, Sumangala P. and Stephen E. DiCarlo (2000): "Peer Instruction Improves Performance on Quizzes." *Advances in Physiology Education* 24(1, Dec), 51–55.

Russell, Arlene A. (1994): "A rationally designed general chemistry diagnostic test." *Journal of Chemical Education* 71(4), 314–317.

Sadler, Philip Michael (1992): The initial knowledge state of high school astronomy students. Ed.D. Dissertation, Harvard University, Cambridge, MA.

Sadler, Philip M. (1998): "Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments." *Journal of Research in Science Teaching* 35(3, Mar), 265–296.

Sadler, Philip M. (2000): "The Relevance of Multiple-Choice Tests in Assessing Science Understanding." Chap. 11. In: *Assessing Science Understanding: A Human Constructivist View*. (Joel I. Mintzes; James H. Wandersee; and Joseph D. Novak, eds.) (Series ed.: Gary D. Phye. Educational Psychology Series.) San Diego, CA: Academic Press, 249–278.

Sadler, Philip M. and Robert H. Tai (1997): "The role of high-school physics in preparing students for college physics." *The Physics Teacher* 35(5, May), 282–285.

Savinainen, Antti and Philip Scott (2002a): “The Force Concept Inventory: a tool for monitoring student learning.” *Physics Education* 37(1, Jan), 45–52.

Savinainen, Antti and Philip Scott (2002b): “Using the Force Concept Inventory to monitor student learning and to plan teaching.” *Physics Education* 37(1, Jan), 53–58.

Schilling, Harold K. (1959): “On the Rationale of Lecture Demonstrations.” In: *Wesleyan Conference on Lecture Demonstrations*. (V.E. Eaton; C.J. Overbeck; and R.P. Winch, eds.) Middletown, CT: Wesleyan University, 7–8.

Schlatter, Mark D. (2002): “Writing ConcepTests for a Multivariable Calculus Course.” *PRIMUS* XII(4, Dec).

Schulz, Barbara (1989): “The NABT-NSTA High-School Biology Examination: Its Design and Rationale.” Chap. 12. In: *High School Biology: Today and Tomorrow*. (Committee on High-School Biology Education, Papers from a conference held October 1998, in Washington, DC) (Walter G. Rosen, ed.) Washington, DC: National Academy Press, 100–109.

Settlage, J. and A.L. Odom (1995): Natural selection conceptions assessment: Development of a two-tier test ‘Understanding Biological Change’. Talk presented at Annual Meeting (National Association for Research in Science Teaching), San Francisco, CA, 22-25 April.

Seymour, Elaine (2002): “Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology.” *Science Education* 86(1, Jan), 79–105.

Seymour, Elaine and Nancy M. Hewitt (1996): *Talking about Leaving: Why Undergraduates Leave the Sciences*. Boulder, CO: Westview Press.

Singh, Chandralekha (2002): “When physical intuition fails.” *American Journal of Physics* 70(11, Nov), 1103–1109.

Sokoloff, David R. and Ronald K. Thornton (1997): “Using Interactive Lecture Demonstrations to Create an Active Learning Environment.” *The Physics Teacher* 35(Sep), 340–347.

Stanley, J.C. and B.S.K. Stanley (1986): “High-school biology, chemistry or physics learned well in three weeks.” *Journal of Research in Science Teaching* 23(3), 237–250.

Steinberg, Richard N. and Mel S. Sabella (1997): “Performance on Multiple-Choice Diagnostics and Complementary Exam Problems.” *The Physics Teacher* 35(3), 150–155.

Stokstad, Erik (2001): “Reintroducting the Intro Course.” *Science* 293(5535, Aug 31), 1608–1610.

Stuart, J. and R.J.D. Rutherford (1978): “Medical student concentration during lectures.” *The Lancet* 2, 514–516.

Sundberg, Marshall D. (2002): “Assessing Student Learning.” *Cell Biology Education* 1(1, Spring/Summer), 11–15.

Tamir, Pinchas (2000): “National and International Assessment.” Chap. 12. In: *Assessing Science Understanding: A Human Constructivist View*. (Joel J. Mintzes; James H. Wandersee; and Joseph D. Novak, eds.) San Diego, CA: Academic Press, 279–302.

Tan, Kim Chwee Daniel; Ngoh Khang Goh; Lian Sai Chia; and David F. Treagust (2002): “Development and application of a two-tier multiple choice diagnostic instrument to assess high school students’ understanding of inorganic chemistry qualitative analysis.” *Journal of Research in Science Teaching* 39(4, Apr), 283–301.

Taylor, G.T.; C.E. Healy; and M. Macdonald (1994): “Engineering educational development: Raising the quality through partnerships.” In: *Kingston University HEFCE Supplemental Instruction Project*. (J. Wallace, ed.) London: Kingston University, 225–230.

Thornton, Ronald K. and David R. Sokoloff (1998): *Interactive Lecture Demonstrations*. Beaverton, OR: Vernier Software & Technology.

Thorton, C.A. (1991): “Think, tell, share - success for students.” *Arithmetic Teacher* 38(6), 22–23.

Tobias, Sheila (1992a): *Revitalizing Undergraduate Science: Why Some Things Work and Most Don’t*. Tucson, AZ: Research Corporation.

Tobias, Sheila (1992b): “What Makes Science Hard?” In: *Lectures on Science Education, 1991-1992*. (Thomas Moore and Peggie Hollingsworth, eds.) Ann Arbor, MI: Sigma Xi, University of Michigan chapter, 44–68.

Tobias, Sheila (1998): In-class Examinations in College-level Science: What do your tests tell your students about themselves and about your subject? Talk presented at Teaching Introductory Physics, Conservation Laws First: An NSF Faculty Enhancement Conference (Harvard University), Cambridge, MA, June 15-26, 1998. <http://galileo.harvard.edu/conference/tobias_abs.html>

Tobias, Sheila (2000): “From innovation to change: Forging a physics education reform agenda for the 21st century.” *American Journal of Physics* 68(2, Feb), 103–104.

Tobias, Sheila and Jacqueline Raphael (1997): *The Hidden Curriculum: Faculty-Made Tests for Science. Part II: Upper Division Courses.* Vol. II. (Innovations in Science Education and Technology) New York: Plenum Press.

Treagust, David F. (1986): “Evaluating students’ misconceptions by means of diagnostic multiple choice items.” *Research in Science Education*, 363–369.

Treagust, D.F. (1988): “The development and use of a diagnostic instrument to evaluate students’ misconceptions in science.” *International Journal of Science Education* 10, 159–169.

Udovic, Daniel; Deborah Morries; Alan Dickman; John Postlethwait; and Peter Weatherwax (2002): “Workshop Biology: Demonstrating the Effectiveness of Active Learning in an Introductory Biology Course.” *BioScience* 52(3, Apr), 272–281.

van Dijk, Liesbet A.; Gerard C. van den Berg; and Hanno van Keulen (2001): “Interactive lectures in engineering education.” *European Journal of Engineering Education* 26(1, Mar), 15–28.

Wandersee, James H. (1985): “Can the history of science help educators anticipate students’ misconceptions?” *Journal of Research in Science Teaching* 23(7), 581–597.

Wandersee, James H. (1988): “The Terminology Problem in Biology Education: A Reconnaissance.” *The American Biology Teacher* 50(2, Feb), 97–100.

Wandersee, James H.; Joel J. Mintzes; and Mary W. Arnaudin (1989): “Biology from the Learner’s Viewpoint: A Content Analysis of the Research Literature.” *School Science and Mathematics* 89(8, Dec), 654–658.

Wandersee, James H.; Joel J. Mintzes; and Joseph D. Novak (1994): “Research on alternative conceptions in science.” Chap. 5. In: *Handbook of Research on Teaching and Learning*. (Dorothy L. Gabel, ed.) New York: Macmillan, 177–210.

Wilkinson, David M. (2001): “At cross purposes.” *Nature* 412(6846, Aug 2), 485.

Williamson, Vickie M. and Marvin W. Rowe (2002): “Group Problem-Solving versus Lecture in College-Level Quantitative Analysis: The Good, the Bad, and the Ugly.” *Journal of Chemical Education* 79(9, Sep), 1131–1134.

Wimpfheimer, Todd (2002): “Chemistry ConcepTests: Considerations for Small Class Size.” *Journal of Chemical Education* 79(5, May), 592.

Woods, Donald R. (1994): *Problem-Based Learning: How to Gain the Most from PBL.* Hamilton, ON, Canada: Self-published. (Distributed by McMaster University Bookstore)

Wright, John C. (1996): "Authentic Learning Environment in Analytical Chemistry Using Cooperative Methods and Open-Ended Laboratories in Large Lecture Courses." *Journal of Chemical Education* 73(9, Sep), 827–832.

Wyckoff, Susan (2001): "Changing the culture of undergraduate science teaching." *Journal of College Science Teaching* 30(5, Feb), 306–312.

Yager, Robert E. (1983): "The importance of terminology in teaching K-12 science." *Journal of Research in Science Teaching* 20(6), 577–588.

Yang, Tien-Hsing; Wei-Lee Lu; Jenn-Gee Lo; Chih-Hung Li; and Yun-Jang Yang (2001): A teaching strategy for physics learning in a Chinese military academy environment. Presented at 31st Annual Frontiers in Education Conference (IEEE Education Society), Reno, NV, 10-13 October. F1B–1 vol. 2.

Zeilik, Michael; Candace Schau; Nancy Mattern; Shannon Hall; Kathleen W. Teague; and Walter Bisard (1997): "Conceptual astronomy: A novel model for teaching postsecondary science courses." *American Journal of Physics* 65(10, Oct), 987–996.

Zeilik, Michael; Candace Schau; and Nancy Mattern (1999): "Conceptual astronomy. II. Replicating conceptual grains, probing attitude changes across three semesters." *American Journal of Physics* 67(10, Oct), 923–927.

Appendix A:

Peer Instruction/Collaborative Learning

Implementation Survey

PROJECT GALILEO

Peer Instruction/Collaborative Learning Implementation Survey

Thank you for your willingness to share your experiences with collaborative learning. This survey should take about thirty minutes to complete.

Although the survey below may use some language specific to Peer Instruction, please feel free to adapt the questions to other, similar strategies for engaging students, in which lectures are interspersed with questions aimed at uncovering student difficulties. If you have not used such an approach, please tell us by sending an e-mail to PIsurvey@ultrafast.eas.harvard.edu.

The results of this survey will be used to develop a Peer Instruction Transplantation Guide that will provide suggestions for implementation of active learning strategies in a variety of disciplines and settings. The guide will be freely available to survey participants. Participants will also receive unlimited free access to an exciting new ConcepTest delivery Web site that we are currently implementing. In addition to providing an expanded database of ConcepTests in several disciplines, this site will be able to deliver class-ready and Web-ready teaching materials, such as customized overhead transparencies in PDF format, Web pages and interactive study guides for students, and provide access to secure on-line testing of students.

A. Personal Information

A1. First Name ______________________________ (Required)

A2. Last Name ______________________________ (Required)

A3. Position [Select one..............]
(_) Faculty
(_) Instructor (non-Faculty)
(_) Administrator
(_) Researcher
(_) Student
(_) Other

A4. Department ______________________________

A5. Institution ______________________________ (Required)

A6. Country ______________________________

A7. Type of Institution [Select one.................]
(_) University
(_) 4-Year College
(_) 2-Year College
(_) Community College
(_) Trade or Vocational School
(_) High School

(_) Middle or Elementary School
(_) Other

A8. E-mail Address ______________________________________ (Required)

A9. Phone Number ____

A10. I consider myself to be [Select one...]
(_) Primarily a researcher
(_) Equally involved in research and education
(_) Primarily involved in education with some research activity
(_) Solely involved in education

A11. I read articles on science education (e.g., in The Physics Teacher, Journal of Chemical Education, etc.) [Select one....]
(_) Consistently
(_) Frequently
(_) Infrequently
(_) Never

B. Background on Peer Instruction (or other collaborative learning strategies)

B1. How did you FIRST learn about Peer Instruction (or other collaborative learning strategies)? [Select one...]
(_) Attended a talk or workshop
(Optional: which? ____)
(_) Read Peer Instruction: A User's Manual
(_) Read another book or article
(Optional: which? ____)
(_) Conversations with colleagues
(_) Co-instructor for a course used Peer Instruction
(_) Teaching assistant for a course which used Peer Instruction
(_) Took a class which used Peer Instruction
(_) Familiar with another course taught using Peer Instruction
(_) Other ____

B2. How have you SINCE learned about Peer Instruction (or other collaborative learning strategies)? [Select all that apply]
[_] Attended a talk or workshop
(Optional: which? ____)
[_] Read Peer Instruction: A User's Manual
[_] Read another book or article
(Optional: which? ____)
[_] Conversations with colleagues
[_] Co-instructor for a course used Peer Instruction
[_] Teaching assistant for a course which used Peer Instruction
[_] Took a class which used Peer Instruction
[_] Familiar with another course taught using Peer Instruction
[_] Other ____

B3. What prompted you to implement Peer Instruction in your own teaching? [Select all that apply]
[_] Wanted to try active learning after hearing about it

[_] Co-instructor for a course used Peer Instruction
[_] Course previously taught with Peer Instruction
[_] Encouraged by department chair
[_] Encouraged by teaching center
[_] Other ____

C. Course Information

How many courses have you taught using Peer Instruction (or other collaborative learning strategies)? _____

Please count teaching the same course multiple times or teaching multiple terms of the same integrated course (such as a year-long sequence in which all terms must be taken to receive credit) as one course.

If more than one, please complete this form for the course you have taught with Peer Instruction (or other collaborative learning strategies) MOST recently.

C1. Subject of class ____

C2. Level [Select one...........................]
(_) Pre-college
(_) Pre-college Advanced Placement
(_) Introductory Undergraduate
(_) Intermediate (second class in an area)
(_) Advanced Undergraduate
(_) Graduate

C3. Approximate student composition in course (categories are not necessarily mututally exclusive)
____% Majors in your department
____% Majors in a closely related department (e.g., majors in other sciences)
____% Taking to fulfill a departmental requirement (for any department)
____% Taking to fulfill a pre-medical requirement
____% Taking to fulfill an engineering requirement
____% Taking as an elective
____% First-year students

C4. Approximate course enrollment ____________________

C5. Approximate percent attendance ____________________ %

C6. Length of class period __________ minutes _____ times per week

C7. Institution at which course offered (if different from current institution) ____

C8. Number of times you have taught this course using Peer Instruction
____________________ times

C9. Have you taught the same course without Peer Instruction? [Select all that apply]
[_] Yes, in traditional lecture format
[_] Yes, other format(s) ______________________________
[_] No

C10. Other staff involved in teaching this course [Select all that apply]
[_] Co-instructor(s)
[_] Teaching assistant(s)
[_] Other ____

C11. Course format and activities offered [Select all that apply]
[_] Lectures
[_] Collaborative seminars
[_] Student presentations
[_] Discussion sections; format: _________________________________
[_] Laboratories
[_] Problem-solving sessions
[_] Other ____

D. Implementation of Peer Instruction

D1. Did you use Peer Instruction throughout the entire term? [Select one..............]
(_) Yes, for the entire term
(_) No, started mid-term
(_) No, stopped mid-term
(_) Other

D2. Of the course activities listed in C11, which used Peer Instruction? [Select all that apply]
[_] Lectures
[_] Collaborative seminars
[_] Student presentations
[_] Discussion sections
[_] Laboratories
[_] Problem-solving sessions
[_] Other

D3. Average percentage of lecture period spent on questions (ConcepTests)
______ % of lecture period, including discussion and explanation

D4. Average number of questions per lecture period
____

D5. If the number of questions or the amount of time varies greatly from class to class, please describe.

D6. Polling method(s) used [Select all that apply and, if possible, include approximate percentages of each type]

[_] ____% -- Flashcards
[_] ____% -- Raising of hands
[_] ____% -- Electronic polling; specify system: _______________
[_] ____% -- Scanning forms or other paper record
[_] ____% -- Other ______________________________

D7. Type(s) of questions used (categories are not necessarily mututally exclusive) [Select all that apply and, if possible, include approximate percentages of each type.]

[_] ____% -- Conceptual
[_] ____% -- Factual/Mathematical
[_] ____% -- Open-ended/Free response
[_] ____% -- Multiple-choice
[_] ____% -- Predict outcome of demonstration
[_] ____% -- Other ______________________________

D8. Source of questions [Select all that apply and, if possible, include approximate percentages of each type]

[_] ____% -- Wrote own questions (including adapting from other sources)
[_] ____% -- Received from a colleague
[_] ____% -- Peer Instruction: A User's Manual by Eric Mazur
[_] ____% -- Project Galileo Web site
[_] ____% -- Other Web site; URL: ______________________________
[_] ____% -- Textbook(s)
[_] ____% -- Other ______________________________

D9. If you have the students discuss their answers (as described on the Peer Instruction Web page <http://mazur-www.harvard.edu/education/pi.html>), what is the range of correct answer percentage before your students talk to each other? In your experience, what percentage yields the best discussion and improvement in answers?

D10. Did students have access to these questions outside of class? [Select all that apply]

[_] No [Continue with question D13]
[_] Yes, on Web site; URL: ______________________________
[_] Yes, via handouts
[_] Yes, but only upon request
[_] Yes, other ____

D11. If yes, access provided [Select one.......................]

(_) At beginning of term
(_) At beginning of each class meeting
(_) After each class meeting
(_) Before exams
(_) At end of term

D12. How much do you think students took advantage of this access? [Select one...]

(_) Often
(_) Sometimes
(_) Rarely
(_) Never
(_) No idea

D13. Please describe any additional important features of how you implemented Peer Instruction. We are particularly interested in any adaptations you have made to fit your particular environment, as well as the reasoning that led to those adaptations.

E. Grading and Assignments

E1. How are student final grades determined? [Select all that apply]
[_] On a curve
[_] On an absolute scale
[_] Other ____

E2. Please comment upon the degree of competition among the students, as well as any elements of the course structure or policy designed to reduce competition.

E3. How did the types of questions used for Peer Instruction compare with examination questions? For example, if some ConcepTest questions were conceptual, were some exam questions also conceptual?

E4. Did students receive any credit/grades based upon ConcepTests? [Select all that apply]
[_] Graded for content and correctness of answers
[_] Graded a subset/spot-checked
[_] Credit for participation
[_] No credit

E5. Did you require pre-class reading?
(_) Yes
(_) No [Continue with question F1]

E6. If yes, how did you assess if students had completed the reading? [Select all that apply]
[_] Multiple-choice reading quiz
[_] Free-response reading quiz
[_] Reading summary
[_] No assessment of reading completion [Continue with question F1]
[_] Other ____

E7. Were these reading assignments administered on the Web?
(_) Yes
(_) No

E8. When were these reading assignments administered? [Select one...]
(_) Before class
(_) During class
(_) After class

E9. Credit for completing reading assignments [Select all that apply]

[_] Graded for content and correctness of answers
[_] Graded for effort shown
[_] Graded for completion
[_] Graded a subset/spot-checked
[_] Not graded/no points awarded

E10. Please comment upon the effectiveness of these reading assignments.

F. Results

F1. Did you use any standardized instruments to assess your students' understanding and/or attitudes, such as those listed on the Physical Science Resource Center <http://www.psrc-online.org/>, ACS General Chemistry Conceptual Exam, or other assessment tool?

(_) Yes; which? ____
(_) No [Continue with question F3]

F2. How did the students perform on these standard instruments? Please compare with any assessment data you have collected for the same course using other pedagogies.

F3. How did students' mastery of the material, both in quantity and in quality, compare to students' mastery in the same course taught another way?

G. Evaluation

G1. Overall, was your experience with Peer Instruction successful? Would you be likely to use it in the future?

G2. Did you find using Peer Instruction valuable and/or enjoyable?

G3. Did students find Peer Instruction valuable and/or enjoyable? How did their reactions and course evaluations compare to those received under different pegagogies?

G4. How eager or relucant were students to work together on Peer Instruction questions? Did they all participate in discussions with each other? Did this level of participation change throughout the term?

G5. How did the effort you expended in using Peer Instruction compare to other ways you have taught? Please consider the effort you expended the first time you taught this class and compare the effort involved in teaching a class the first time using Peer Instruction with subsequent semesters.

G6. What difficulties did you encounter in implementing Peer Instruction (e.g., logistical, political)? If you were able to overcome these obstacles, please say how.

H. Community

H1. How many of the other instructors in your department also teach using Peer Instruction (or other collaborative learning strategies)?

____

H2. If you know of any other instructors at your institution or elsewhere who have used Peer Instruction (or other collaborative learning strategies), please provide any contact information for them so we can include them in our survey.

H3. How many of the other instructors in your department know that you teach in a nontraditional fashion?

____

H4. How many of the other instructors in your department have visited your class or discussed your specific teaching strategies with you?

____

H5. How many other instructors at your institution, but outside of your department, have visited your class or discussed your specific teaching strategies with you?

____

Do you have any additional remarks (including suggestions, recommendations, or warnings for others considering or using Peer Instruction)?

Thanks again for your time! We will inform you as soon as an analysis of these survey results is available.

Please e-mail PIsurvey@ultrafast.eas.harvard.edu with any questions or problems.

SUBMIT   RESET

Appendix B:

Demonstration Study Documents

Introduction to classroom demonstration study for Teaching Fellows

Instructions for Teaching Fellows on *observe* mode

Instructions for Teaching Fellows on *predict* mode

Instructions for Teaching Fellows on *discuss* mode

Demonstration log form

Sample viewgraph for *predict* and *discuss* modes

Sample worksheet for *discuss* mode

Physics 1b – Spring 2002
Demonstration Study

We are delighted that you will be a Teaching Fellow for Physics 1b this semester. Not only will you be helping your students learn physics, but you will also be helping to conduct a research study of the effectiveness of classroom demonstrations. With your help, we can help improve the pedagogy of demonstrations so they can maximally helpful for students.

We'll have a series of demonstrations that you will be presenting in several different "modes" (see below), according to a pre-determined schedule (this allows students in each section to experience the different modes for different demonstrations). We will be asking you to record certain data about the presentation of the demonstration in your section. It is important that the students not be made aware of this study (since it may bias their participation), so try to be as natural and equally effective as you can be.

The purpose of the study is to see if student understanding — and indeed memory — of demonstrations can be improved by a mode of presentation which engages the minds of students and forces them to address their misconceptions, rather than allowing them to be merely passive observers. During the semester, we'll have several demonstrations, one per week, that we will ask Teaching Fellows to present in one of several different modes, according to a schedule to be supplied later. The different modes reflect different degrees of student engagement:

1. **no demonstration** (control group);
2. **observe**: TFs present and explain the demo as in a standard lecture (no student discussion);
3. **predict**: students are asked to individually predict the outcome of the demonstration with a ConcepTest before the demo;
4. **discuss**: students follow a worksheet which not only asks them to predict the outcome, but also explicitly asks them to record the observed result of the demo and compare it with their prediction; they are also given the opportunity to discuss the outcome of the demonstration *after* it is conducted.

No matter which mode you are doing in a given week, **you should do your best to make sure students understand the demonstration and the underlying physics**. You should also try to spend approximately the same total amount of time each week on the demo (obviously unless you are in the no demo mode). **Try to be as natural as possible in implementing your assigned mode** for that week so that the students do not think one or modes to be artificial or less important. If your section is in the no demo mode for that week, you should try to put the apparatus out of sight.

Towards the end of the semester, students will be asked to complete a Web-based test (which counts as a homework set) as part of their preparation for the final exam. This test will be based around the concepts involved in the demonstrations, but will not specifically refer to the demonstrations (after all, some students will not have seen them). With this test, we will assess their memory of the outcome of the demonstrations and their ability to explain the underlying physics.

Data Collection

In order to record as much information as possible about how the demonstration went in your section, we will be distributing a brief form for you to submit each week. Since there are two TFs in each section, we suggest that one of you conduct the demo each week, and the other sit at the back and record the requested data.

The form will ask how much time was spent in various phases of the demo presentation, the results of any polling data from the students (for "predict" and "discuss" modes), and any other variables. Students will be asked to bring their PRS transmitter to section so that we can accurately record their responses to predictions. The PRS system will also provide a record of student attendance. (Although student attendance in section is not required, it is important that we know who saw which demo and in which mode for data analysis.) In order to record attendance for the "observe" mode, we'll either have to include an unrelated question for them to respond via PRS, or ask you to unobtrusively record who is there in a given week.

There are four loaner PRS transmitters that will stay in Science Center 309A for students who may forget their own. Please be sure to record the name and Harvard ID of students who borrow a transmitter, along with the ID of the loaner PRS and section date and time on the loaner log.

Background (thanks to Paul Callan)

Demonstrations form an integral part of most introductory physics courses. They serve many different functions, of which the two most important are: (1) to assist students in learning physics by showing them physical principles in action; and (2) to entertain and motivate student interest. But how well do demonstrations fulfill these functions? Can the manner of presentation of a demonstration make it more or less effective? It is normally evident when students are entertained by a demonstration, and some professors pull off flashy demonstrations more effectively than others. Less clear is the effectiveness of demonstrations in achieving the first of these goals, namely assisting in student learning. The goal of our project described here is to find out how best to present demonstration to enhance student understanding of the physical concepts being demonstrated.

Previous studies, by Pamela Kraus of the Physics Education Group at the University of Washington and others, suggest that demonstrations do not contribute much to student understanding, *i.e*., that they don't actually *demonstrate* much at all. Pamela Kraus's results [1] indicate that students shown a demonstration can later describe what happens better than those not shown the demonstration, but that their understanding of the physical concepts is no better. Perhaps most shocking is that many students who see a demonstration make incorrect observations and/or alter their memory of the outcome. In other words, students may remember something that did not actually occur, especially when the observed phenomenon does not match their expectation. While shocking to us as scientists, this conclusion would not surprise psychologists; there is much evidence that observations and memory are affected by previously held beliefs and understanding. [2, 3]

[1] P. Kraus, Ph.D. thesis, University of Washington (1997).
[2] P. Gray, *Psychology* , 2nd ed., (Worth, New York, 1994).
[3] R. Gunstone and R. White, *Science Education* **65**, 291 (1981).

Physics 1b Demonstrations
Mode: **Observe**

With the Observe mode, you just introduce the demo, do it, then explain why it worked You will need:

- ✓ Demo apparatus

which should be provided for you in Science Center 309A. If you can't find anything you need, please let me know.

What to do

Below, I've sketched out a general procedure to follow when doing a demonstration in the Observe mode. As we suggested earlier, we would like one TF to take the responsibility for recording info about the demo (time spent in various parts, attendance, etc.) while the other TF actually does it.

1. Begin by **setting up the demo** (if it's not already ready to go).
2. **Explain the setup** so that students understand the apparatus and what you will be doing.
3. Now you actually **do the demo**. Make sure it's in a place that all the students can see it.
4. After you've done the demo, one of the TFs should then **explain the result and the reasoning** behind the observation.

Since we don't use the PRS system with the Observe mode, we won't automatically know who saw which demo in which mode. For this reason, it's important that you keep an accurate attendance list. Section attendance is still up to the student, so this should be done unobtrusively. For instance, the recording TF can check off names on a section enrollment list when the other TF is presenting the demo.

Physics 1b Demonstrations
Mode: **Predict**

With the Predict mode, students will be responding via the PRS system so that we can have a record of their understanding. You will need:

- ✓ Demo apparatus
- ✓ PRS system setup
- ✓ Overhead with demo-related ConcepTest

All of these materials should be provided for you in Science Center 309A. If you can't find anything you need, please let me know.

PRS Transmitters

Students have been asked to bring their PRS transmitters with them to section. If they remember them, all they have to do is respond at the appropriate time. For students who forget their transmitters, we have several loaners which students can borrow for the section (these loaners should not leave the room). Please be sure to record the name and Harvard ID of students who borrow a transmitter, along with the ID of the loaner PRS and section date and time on the loaner log.

Since students will generally not be able to see the computer screen to know that their response has been recorded, you can call out the numbers as responses are recorded by the computer until the number of responses equals the number of students present.

PRS Software

Please see the separate page explaining the operation of the computer and PRS software. For the predict mode, you will be collecting one data point for each students: the initial answer to the ConcepTest.

What to do

Below, I've sketched out a general procedure to follow when doing a demonstration in the Predict mode. As we suggested earlier, we would like one TF to take the responsibility for recording info about the demo (time spent in various parts, attendance, etc.) while the other TF actually does it.

1. Begin by **setting up the demo** (if it's not already ready to go).
2. **Explain the setup** so that students understand the apparatus and what you will be doing. Display the ConcepTest without showing the answers.
3. **Display the ConcepTest** associated with the demo, and read it to the students.
4. Then **ask the students to record their response** to the ConcepTest using the PRS system. Since the students won't be able to see if their responses were recorded as in lecture, you should use the counter in the software to

make sure that everyone has responded. There's no harm in students responding more than once so, rather than trying to track down the one missing person by PRS number, you can ask everyone to respond again.

5. Now you actually **do the demo**. Make sure it's in a place that all the students can see it.
6. After you've done the demo, one of the TFs should then **explain the result and the reasoning** behind the observation.

Physics 1b Demonstrations
Mode: **Discuss**

With the Discuss mode, students will both be completing a worksheet for them to take and responding via the PRS system so that we can have a record of their understanding. You will need:

- ✓ Demo apparatus
- ✓ PRS system setup
- ✓ Worksheets for students to complete and take
- ✓ Overhead with demo-related ConcepTest
- ✓ Overhead for students to respond on prediction vs. observation

All of these materials should be provided for you in Science Center 309A. If you can't find anything you need, please let me know.

PRS Transmitters

Students have been asked to bring their PRS transmitters with them to section. If they remember them, all they have to do is respond at the appropriate time. For students who forget their transmitters, we have several loaners which students can borrow for the section (these loaners should not leave the room). Please be sure to record the name and Harvard ID of students who borrow a transmitter, along with the ID of the loaner PRS and section date and time on the loaner log.

Since students will generally not be able to see the computer screen to know that their response has been recorded, you can call out the numbers as responses are recorded by the computer until the number of responses equals the number of students present.

PRS Software

Please see the separate page explaining the operation of the computer and PRS software. For the discuss mode, you will actually be collecting *two* sets of data with the PRS devices: students' initial answers to the ConcepTest and how closely their prediction matches the actual observation.

What to do

Below, I've sketched out a general procedure to follow when doing a demonstration in the Discuss mode. As we suggested earlier, we would like one TF to take the responsibility for recording info about the demo (time spent in various parts, attendance, etc.) while the other TF actually does it.

1. One of you could begin by **setting up the demo** (if it's not already ready to go). Meanwhile, the other TF can **hand out the worksheets** for that demo.
2. Once everyone has a copy, you should **explain the setup** so that students understand the apparatus and what you will be doing to initiate the demo.

3. Then **give students 1-2 minutes** to *individually* answer part A of the workshop in which they should predict what they think will happen and why. They'll have time to discuss it collectively later in the process.
4. After it seems that students have completed this part, you should **display and read the ConcepTest** associated with the demo and introduce it as "these are some common predictions."
5. Then **ask the students to record their response** to the ConcepTest using the PRS system. Since the students won't be able to see if their responses were recorded as in lecture, you should use the counter in the software to make sure that everyone has responded. There's no harm in students responding more than once so, rather than trying to track down the one missing person by PRS number, you can ask everyone to respond again.
6. Now you actually **do the demo**. Make sure it's in a place that all the students can see it.
7. Immediately after performing the demo, **ask students to report their observation** in section B of the worksheet.
8. Then **have them complete section C** and use the choices listed there to **respond using their PRS device** (this records whether their prediction was correct or not). You should have an overhead of the choices for you to display (it's also listed on their worksheet)
9. Now have **students discuss** their predictions and observations which each other, trying to convince each other.
10. After a few minutes of discussion (in which the TFs can float around the room), one of the TFs can **explain the reasoning** behind the correct answer.

DEMO LOG

Date: __________________ Section: T2 / T4 / T7 / W2 / W7 Students present: ____

Demo: _____________________________________ Mode: observe* / predict / discuss

TF Presenting: _________________________ TF Recording: _______________________

Please be sure to complete the column for the assigned mode of presentation fully. If you did not engage in one of the listed activities (such as if there were no student questions), please say so explicitly. Please also be sure to note the end time of the demo.

Record starting times for each activity:	**observe***	**predict**	**discuss**
1. distribute recording sheets			
2. explain demo setup			
3. students complete block 1 on sheets			
4. put up and present demo transparency			
5. students think and vote using PRS system			
6. carry out demonstration			
7. students complete blocks 2 and 3 on sheets			
8. students discuss their interpretations			
9. TF explains demonstration			
10. answer questions			
11. end			

Please note down anything that may be relevant to the study (*e.g.*, problems with demo, deviations from the indicated mode). Use the back if necessary.

* If performing the demo in observe mode, please also unobtrusively complete the attached attendance list, making any necessary additions. (Section attendance is still up to the student, but it is essential that we know who saw which demo in which mode for the purposes of the study.) Attendance for predict and discuss modes is automatically recorded with the PRS data.

Viewgraph for loaded beam demonstration, Physics 1a
(for predict *and* discuss *modes)*

A plank of negligible mass is supported at its two ends by platform scales. When a block of metal is placed at the center of the plank, halfway between the scales, the scales have the same reading *X*. If the metal block is now placed over the right-hand scale, the two scale readings are:

1. right scale = *X*, left scale = *X*
2. right scale = *X*, left scale = 0
3. right scale = 0, left scale = *X*
4. right scale = 2*X*, left scale = 0
5. right scale = 0, left scale = 2*X*
6. right scale = 1.5*X*, left scale = 0.5*X*
7. right scale = 0.5*X*, left scale = 1.5*X*
8. none of the above

Worksheet for loaded beam demonstration, Physics 1a
(for discuss *mode)*

A plank of negligible mass is supported at its two ends by platform scales. When a block of metal is placed at the center of the plank, halfway between the scales, the scales have the same reading *X*. The metal block is now placed over the right-hand scale.

A. What are the two scale readings now? Why?

B. Record your observation of the demonstration.

C. Compare your prediction (A) to your observation (B). Do they agree?

__ Completely (1) __ Mostly (2) __ Somewhat (3) __ Not at all (4)

When prompted by your TF, please use your PRS transmitter to register this comparison using the 1-4 scale listed above.

D. After discussing your prediction and the demonstration with your neighbors, record why your prediction and the reasoning behind it were correct or incorrect (use the back of this sheet if you need more room).

Appendix C:

End-of-Semester Demonstration Test for Physics 1a

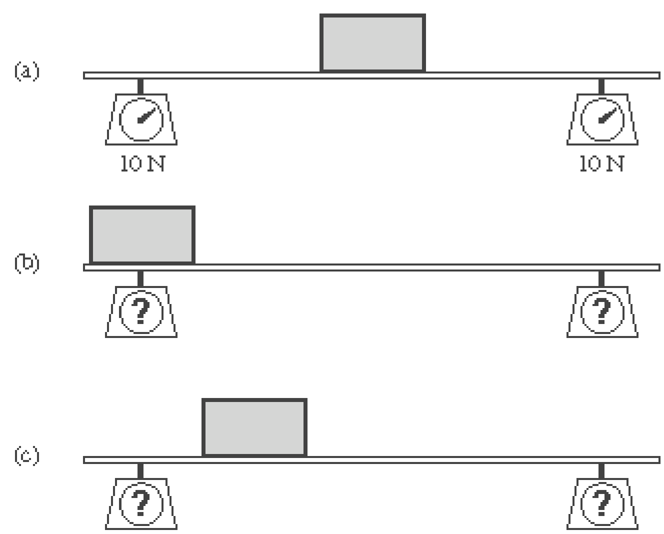
Physics 1
Computer Test 3

On this electronic assignment:		
Collaboration with others is	**not** permitted	
Use of chapter summaries	permitted	
Use of other printed reference materials is	**not** permitted	
Use of other online or electronic reference is	**not** permitted	
Use of calculator is	permitted	

This test has ten questions. Please answer each question and provide a brief explanation of your answer. Each question is graded separately; you receive full credit if your answer demonstrates genuine effort regardless of its correctness. You have up to two hours to complete the test, but it is not expected that you will need the entire time. You may submit earlier, but please keep in mind that you will not be able to change your responses after they are submitted.

---

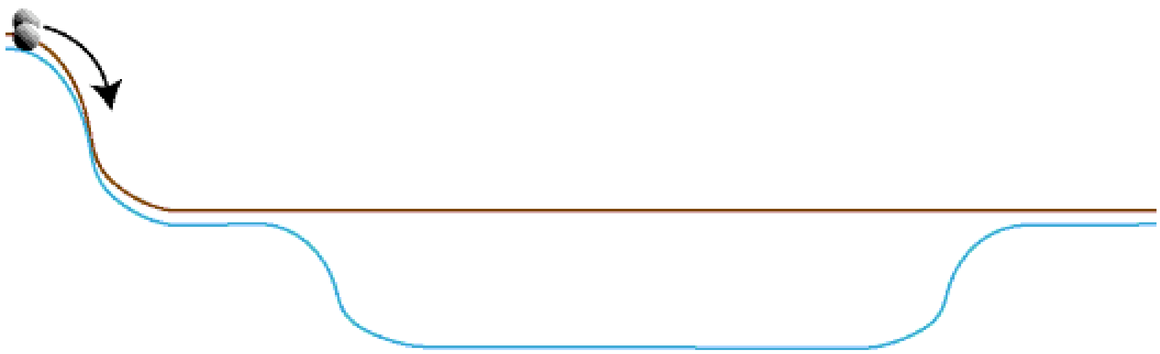
1. A plank of negligible mass is supported at its two ends by platform scales. When a block of metal is located at the center of the plank, halfway between the scales, the scales have the same reading of 10 N as shown in (a).



If the metal block is now placed over the left-hand scale, as in (b), what are the readings on the scales? Explain your answer briefly.

What are the readings when the block is placed halfway between the left-hand end and the center of the plank, as in part (c) of the diagram? Explain your answer briefly.

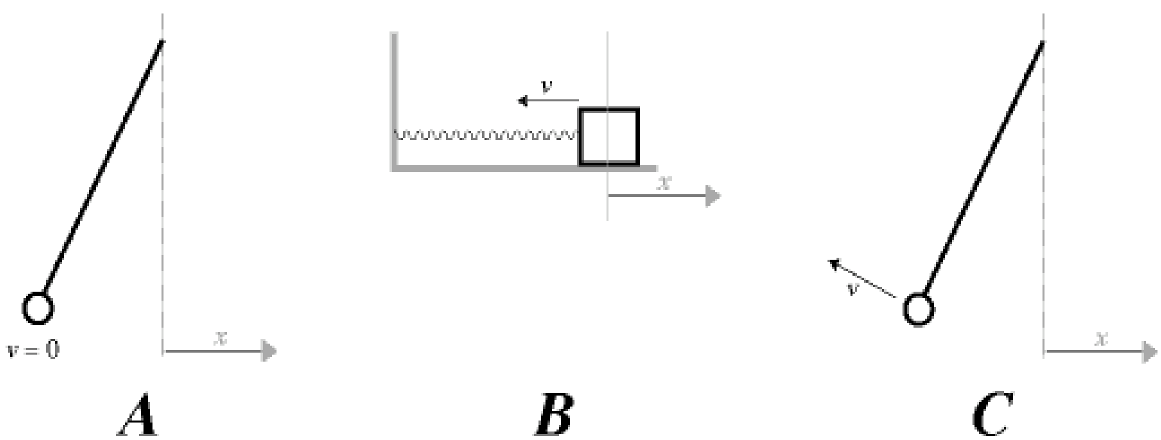
2. Consider the motion of two identical balls on the tracks shown in the diagram below. The tracks begin and end at the same vertical height, and the horizontal width of both is the same. One track dips significantly lower in the middle than the other.



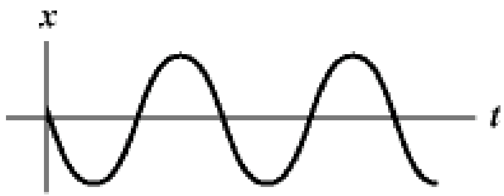
The balls are released simultaneously from the left ends of the two tracks. Which ball reaches the right end of the tracks first? Briefly explain your reasoning.

3. Consider the following three situations:

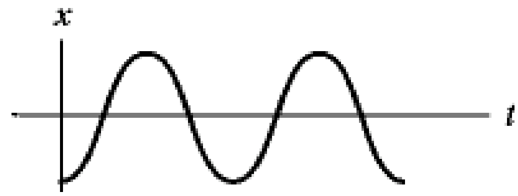
A. a pendulum held at rest away from equilibrium as shown and then released at $t = 0$
B. an object on a spring oscillating and passing through its equilibrium position as shown at $t = 0$
C. a pendulum swinging as shown at $t = 0$



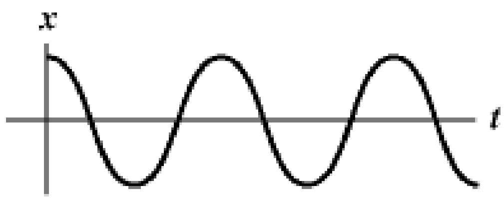
Each of these situations may be described by one (or more) of the following *x(t)* graphs:



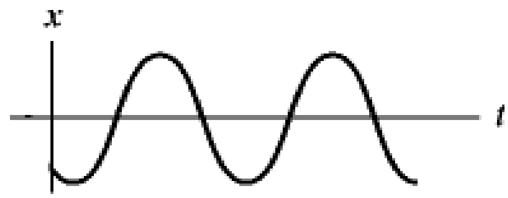
1



2



3



4

For each of the three lettered situations, identify which of the numbered *x(t)* graphs above best corresponds. In the text box next to each, explain briefly which features of the graph(s) you used to establish the correspondence.

Situation A 1 2 3 4

Situation B 1 2 3 4

Situation C 1 2 3 4

---

4. A remote-controlled model car is placed on a lightweight foam "roadbed," which in turn is set on a dozen empty soda cans which are free to roll under the roadbed. The inertia of the car and that of the roadbed are equal. What happens when the car starts to drive forward? Briefly explain your reasoning.

If the inertia of the car were much greater than that of the roadbed, what would happen when the car starts to drive forward? Briefly explain your reasoning.

---

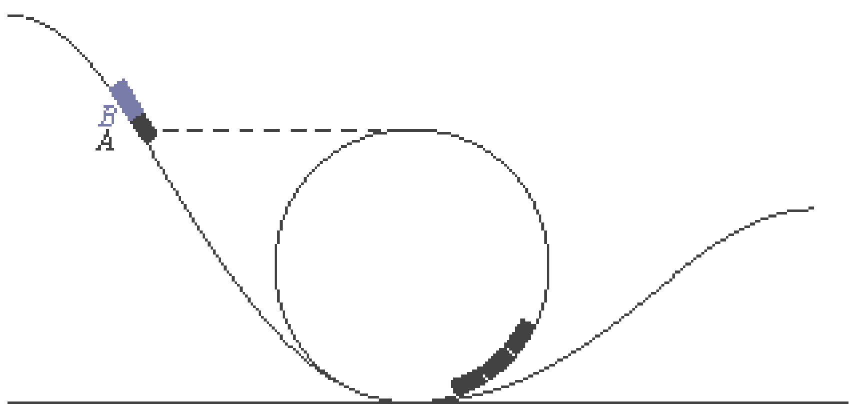
5. You are walking out of Cabot Science Library carrying a tall stack of books. The magnitude of the contact force exerted by the top book on the book immediately below it is *F*. Now suppose you stop walking and drop the stack of books. While the books are falling, is the magnitude of that contact force greater than, equal to, or less than *F*? If it is possible with the information given to determine the value of the contact force, give its value; if not, indicate what additional information you would need to determine its value. Briefly explain your reasoning.

---

6. A person attempts to knock down a large wooden bowling pin by throwing a ball at it. The person has two balls of equal size and inertia, one made of rubber and the other of putty. The inertia of the pin is much greater than that of the balls. On striking the pin, the rubber ball bounces back, while the ball of putty sticks to the pin. Which ball is more likely to knock the pin over? Briefly explain your reasoning.

---

7. A roller-coaster track includes a complete loop as shown in the diagram. Cars move along the track with negligible friction and are *not* held in place on the track by rails.



If a car is held on the initial part of the track at exactly the same height as the top of the loop (position *A* in the diagram), does it travel all the way around the loop successfully? If the car is held at position *B*, slightly higher than the top of the loop, does it travel all the way around the loop successfully? Explain your answer briefly.

---

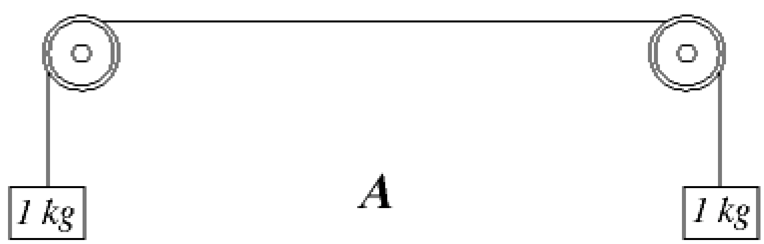
8. Consider the space shuttle starting at rest on its launch pad and then taking off. Is the magnitude of the change in momentum of the shuttle, from when the shuttle was initially at rest to when it is airborne, greater than, less than, or equal than the magnitude of the change in momentum of the Earth during the same time period? If you need more information to answer this question, indicate what information you need. Explain your answer briefly.

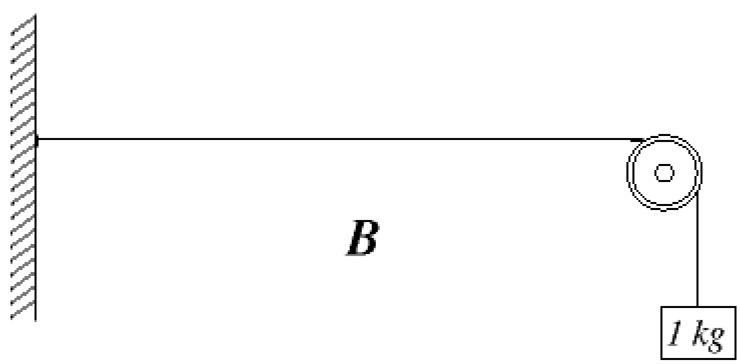
Is the change in the shuttle's kinetic energy greater than, less than, or equal to the change in the Earth's kinetic energy? As the shuttle gains kinetic energy, does the Earth gain or lose kinetic energy? If you need more information to answer this question, indicate what information you need. Explain your answer briefly.

---

9. Consider the following two situations (shown in the diagram below).

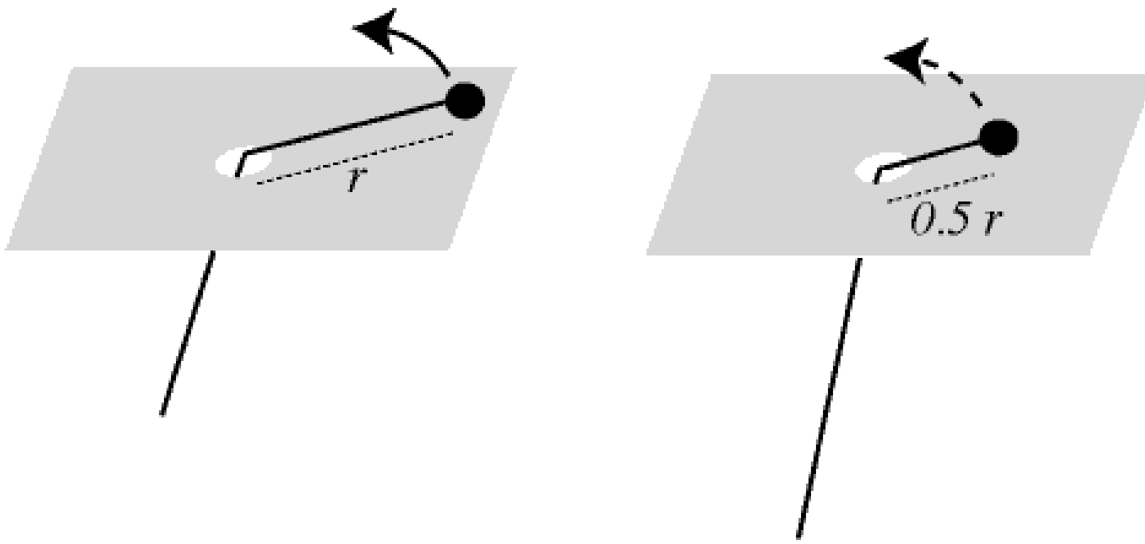
A. A string is attached to two 1-kg blocks over two pulleys. The tension in the string is T_A.
B. A string is attached to a wall at one end and to a 1-kg block at the other end over a pulley. The tension in the string is T_B.





Is $|T_A|$ greater than, equal to, or less than $|T_B|$? Explain your answer briefly.

10. A puck sits on a flat tabletop. The puck is fastened to one end of a string; the other end of the string goes through a hole in the tabletop and is held underneath the table. The puck is given a push so that it begins to travel in a circular path centered on the hole (the string is taut). The string is then pulled partway through the hole, so that the length of string on the tabletop is half what it originally was. Comparing the speed of the puck from before the string is shortened to after, by what factor does the tangential speed of the puck change? Explain your answer briefly.



You have completed the test! Verify that you have answered all questions, place a checkmark in front of the statement below, then click the "Submit" button.

☐ I, <<First>> <<Last>> (ID ########), certify that the work I am submitting is entirely my own.
☐ I received no help and I did not refer to books or notes (other than the provided chapter summaries) while completing this assignment.

Appendix D:

End-of-Semester Demonstration Test for Physics 1b

Physics 1b – Spring 2002

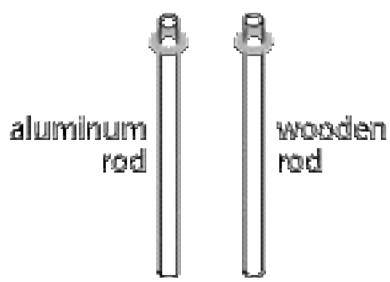
Online Test 2

On this electronic assignment:	
Collaboration with others is	**not** permitted
Use of other printed reference materials is	**not** permitted
Use of other online or electronic reference is	**not** permitted
Use of calculator is	permitted

This test has nine questions, some of which have more than one part. Please answer each question and provide a brief explanation of your answer. Each question is graded separately; you receive full credit if your answer demonstrates genuine effort regardless of its correctness. You have up to two hours to complete the test, but it is not expected that you will need the entire time. You may submit earlier, but please keep in mind that you will not be able to change your responses after they are submitted.

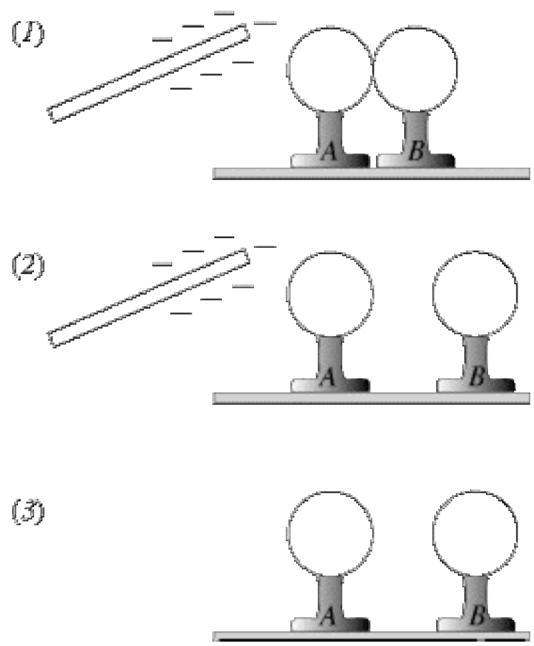
---

1. Two rods of the same diameter, one aluminum (conducting, non-magnetic) and one wooden, are held vertically. If identical magnetic rings are placed around the top of each rod and released at the same time, which ring reaches the bottom first? Explain briefly. (Neglect any friction between the ring and the rod.)



---

2. Two uncharged metal spheres, *A* and *B*, are placed on insulating glass stands and placed next to each other, so that the spheres touch. A negatively charged rod is brought near sphere *A* without touching it. Then, with both the rod and sphere *A* kept stationary, sphere *B* is moved so that it is no longer in contact with sphere *A*. Finally, the rod is removed.



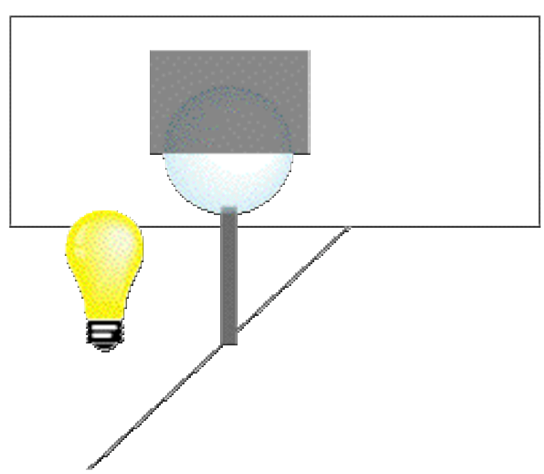
(a) After the rod is removed, is sphere *A* positively charged, negatively charged, or neutral? Explain briefly.

(b) Is sphere *B* positively charged, negatively charged, or neutral? Explain briefly.

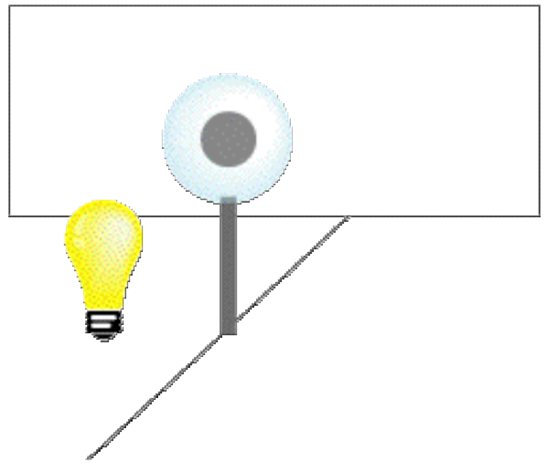
---

3. A light bulb is placed to the left of a converging lens at a distance greater than the focal length of the lens. The image of the bulb is formed on a screen to the right of the lens.

(a) What happens to the image if you cover the top half of the lens with a card? Explain briefly.

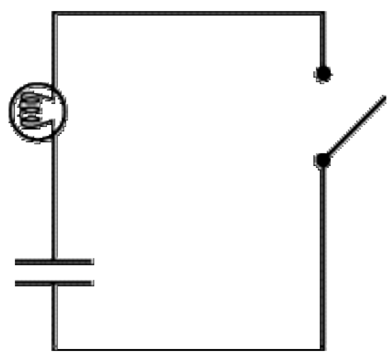


(b) What happens to the image if you cover the center of the lens with a disc? Explain briefly.



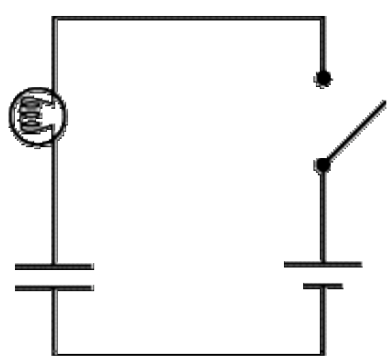
---

4. A light bulb, a charged parallel plate capacitor, and a switch are connected to form a circuit as shown below. The switch is initially open so that the capacitor remains charged. Describe what will happen to the bulb after the switch is closed. Explain briefly.



---

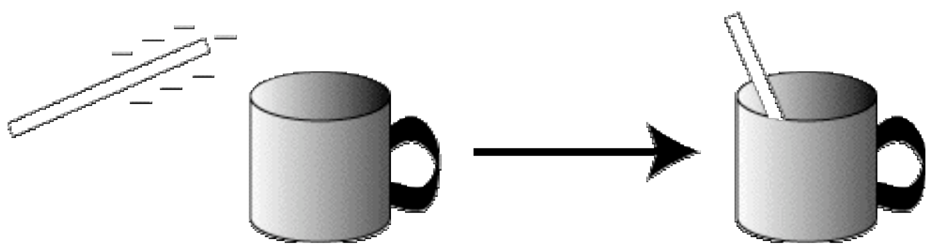
5. A light bulb, an uncharged parallel plate capacitor, a battery, and a switch are connected to form a circuit as shown below. The switch is initially open. Describe what will happen to the bulb after the switch is closed. Explain briefly.



6. A wire passes through a hole in a glass plate so that the wire is perpendicular to the plane of the glass plate. A thin layer of iron filings is sprinkled on the glass plate all around the wire. Then the wire is connected to a battery so that current flows through the wire. When current is flowing through the wire, do the iron filings rearrange at all? If so, how? Explain briefly.

7. Two identical capacitors are placed side by side and connected as shown below. How does the capacitance of the two connected capacitors, C_{comb}, compare to that of just one of the capacitors, C_0? Explain briefly.

8. A large metal cup is charged by touching the bottom of the inner surface of the cup with a negatively charged rod. The charged rod is then removed.

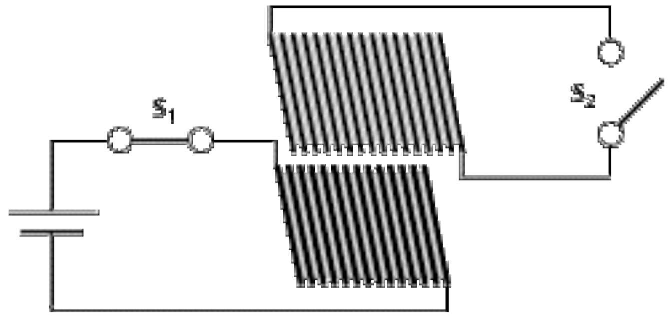


(a) Is the charge density on the inner (bottom) surface of the cup positive, negative, or zero? Explain briefly.

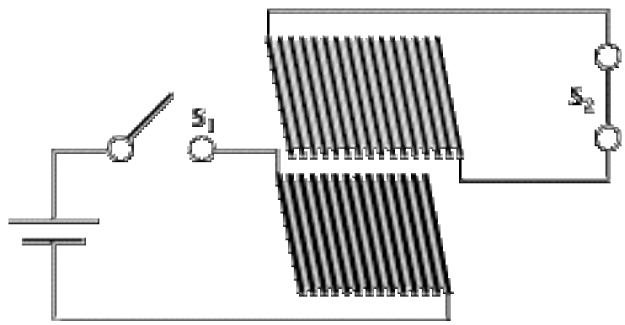
(b) Is the charge density on the outer (side) surface of the cup positive, negative, or zero? Explain briefly.

9. Two conducting coils are placed next to each other. Each coil has a switch connected to it; when the switch is open, current cannot flow through the coil. A battery is connected to the first coil so that current flows through it when its switch, S_1, is closed.

(a) Initially, switch S_1 is closed so that the battery drives a steady current through coil 1, and switch S_2 is open, as shown in the figure below. If S_2 is then closed, is a current induced in coil 2? Explain briefly.



(b) Now let switch S_1 initially be open and switch S_2 closed, as shown in the figure below. No current is flowing through either coil. If S_1 is now closed, is a current induced in coil 2? Explain briefly.



Appendix E:

Genetics vocabulary survey (pre- and post-test)

Biological Sciences 50 – Spring 2001
Beginning of Semester Survey

Adam Fagen (afagen@fas), one of the Teaching Fellows for this course, is doing research on biology and education, and is very interested in finding out about the background of Bio Sci 50 students. This survey is designed to collect information on your previous experience in biology. You can help the teaching staff gain a better sense of your preparation and experience by completing this survey.

Your responses on this survey will have no bearing on your grade. However, you will positively affect your own experience in the course by responding thoughtfully and honestly.

Thanks for your feedback!

Terminology

The following table contains vocabulary taken from the glossary in the back of the text and represent some of the important terms in genetics. For each word, indicate your familiarity with the term in the context of biology and genetics (several have non-scientific meaning as well) on a 1-5 scale as shown. Remember that we do not expect or assume you to know all – or even any – of these terms prior to the course.

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
allele	1	2	3	4	5
amino acid	1	2	3	4	5
anaphase	1	2	3	4	5
annealing	1	2	3	4	5
anticodon	1	2	3	4	5
autosomes	1	2	3	4	5
backcross	1	2	3	4	5
bacteriophage	1	2	3	4	5
carrier	1	2	3	4	5
cell fate	1	2	3	4	5
centromere	1	2	3	4	5
chiasma	1	2	3	4	5
chromosome	1	2	3	4	5
cis-acting	1	2	3	4	5
clastron	1	2	3	4	5
clone	1	2	3	4	5
consensus sequence	1	2	3	4	5
cyclic AMP	1	2	3	4	5
denaturation	1	2	3	4	5
diploid	1	2	3	4	5
dominance	1	2	3	4	5
electrophoresis	1	2	3	4	5
endonuclease	1	2	3	4	5
eukaryote	1	2	3	4	5
evolution	1	2	3	4	5
founder effect	1	2	3	4	5

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
frameshift mutation	1	2	3	4	5
gamete	1	2	3	4	5
gene	1	2	3	4	5
genetic code	1	2	3	4	5
genome	1	2	3	4	5
genotype	1	2	3	4	5
hemizygous gene	1	2	3	4	5
heterozygous	1	2	3	4	5
housekeeping gene	1	2	3	4	5
hybrid	1	2	3	4	5
imprinting	1	2	3	4	5
inbreeding	1	2	3	4	5
intron	1	2	3	4	5
knockout mutation	1	2	3	4	5
lagging strand	1	2	3	4	5
linkage	1	2	3	4	5
major groove	1	2	3	4	5
meiosis	1	2	3	4	5
messenger RNA	1	2	3	4	5
metaphase	1	2	3	4	5
natural selection	1	2	3	4	5
nondisjunction	1	2	3	4	5
nonsense mutation	1	2	3	4	5
nucleus	1	2	3	4	5
Okazaki fragment	1	2	3	4	5
open reading frame	1	2	3	4	5

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
operon	1	2	3	4	5
pedigree	1	2	3	4	5
phenotype	1	2	3	4	5
plasmid	1	2	3	4	5
polymerase	1	2	3	4	5
polymerase chain reaction	1	2	3	4	5
population	1	2	3	4	5
primer	1	2	3	4	5
probe	1	2	3	4	5
promoter	1	2	3	4	5
protein	1	2	3	4	5
recessive	1	2	3	4	5
recombination	1	2	3	4	5
reporter gene	1	2	3	4	5
repressor	1	2	3	4	5
restriction enzyme	1	2	3	4	5
reverse transcriptase	1	2	3	4	5
ribosome	1	2	3	4	5
ribozyme	1	2	3	4	5

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
semiconservative replication	1	2	3	4	5
sex-linked	1	2	3	4	5
silent mutation	1	2	3	4	5
sister chromatids	1	2	3	4	5
Southern blot	1	2	3	4	5
spooling	1	2	3	4	5
sticky end	1	2	3	4	5
telomere	1	2	3	4	5
template	1	2	3	4	5
testcross	1	2	3	4	5
transcription	1	2	3	4	5
transfer RNA	1	2	3	4	5
transgenic	1	2	3	4	5
translation	1	2	3	4	5
transposable element	1	2	3	4	5
wildtype	1	2	3	4	5
X chromosome	1	2	3	4	5
zygote	1	2	3	4	5

Now please briefly *define the following words or describe their biological significance.*

anticodon	genetic code
autosomes	genome
cell fate	intron
consensus sequence	linkage
eukaryote	nonsense mutation
gamete	operon

primer	sticky end
semiconservative replication	telomere
spooling	transposable element

Concepts

Please answer the following multiple choice and short answer questions. Please provide an answer for each question, even if it is a wild guess. Remember, you are not being graded on this survey.

How does the genetic material in one of your skin cells compare to the genetic material in a liver cell?

1. The genetic material is identical in both cells.
2. The two cells share many common genes, but largely contain genes specific to its cell type.
3. Each cell contains only the genetic material for its cell type.
4. It depends upon which specific cells are compared.

What does it mean, at the *molecular* level, to say that a trait is dominant?

You are asked to serve on a jury considering a paternity case. A single mother of a baby is trying to get the father to pay his share of child support. The man she has identified as the baby's father denies the charge. They agree to perform a genetic test; they obtain DNA samples from the child, mother, and accused father and look at the length of DNA fragments resulting from a restriction enzyme digest in a certain region of the human genome. The child shares some of the DNA fragments with the mother but also has several fragments not shown in the mother's sample.

(a) If the man's sample contains fragments which match up exactly with those of the mother, what, if anything, could you conclude about the accused man's paternity?

(b) If the man's sample contains fragments which match up with those of the child *not* present in the mother, what, if anything, would you conclude about the accused man's paternity?

(c) If the man's sample contains fragments dissimilar to both those of the mother and child, what, if anything, could you include about the accused man's paternity?

Background

Please indicate how many years of each of the following subjects you took in high school and whether or not you took the Advanced Placement exam for that subject (and score, if you remember it):

Biology	_____ Years	AP	Y / N	AP Score (if taken)	1 2 3 4 5
Chemistry	_____ Years	AP	Y / N	AP Score (if taken)	1 2 3 4 5
Physics	_____ Years	AP	Y / N	AP Score (if taken)	1 2 3 4 5

Please circle any of the following Harvard courses which you have already completed:

Bio Sci	1	2	10	11	14	25	51	52	53	54	80	Other ___
Biology	10	17	19	20	21	22	24	Other ________			
Chemistry	5	7	10	17	20	27	30	40	60	Other ________	
Physics	1a	1b	11a	11b	15a	15b	15c	16	Other ________		
Science Core	B-16	B-23	B-27	B-29	B-40	B-44	B-46	B-48	B-54		

If you taken coursework in the biological sciences other than in high school or at Harvard (e.g., other colleges, summer programs), please briefly describe:

Please circle the classroom instructional method that you believe is best for you in learning biology:

Lecture   Discussion/Seminar   Laboratory/Hands-on   Demonstration

What are your career goals, if any, right now?

Why did you decide to take this course? What do you hope to gain from it?

What concerns do you have about taking this course?

Name (*only for research, not grades**): ____

Class: '01 '02 '03 '04 GSAS Other _____________   Sex: M / F

Concentration (actual or predicted): ____

Are you pre-med? Y / N   Country in which you attended high school: _________

**Names are only asked so we can compare individuals' responses at the end of the semester.*
The survey has no effect on your grade!

Biological Sciences 50 – Spring 2001
End of Semester Survey

Adam Fagen (afagen@fas), one of the Teaching Fellows for this course, is doing research on biology and education, and is very interested in finding out what BS 50 students have learned during the semester. This survey (and a similar one you may have taken at the beginning of the semester) will help to measure the effectiveness of the course in helping you to learn introductory genetics. You can help the teaching staff to best meet your needs by responding to this survey thoughtfully and honestly.

Your responses on this survey will have no bearing on your grade. This will help to measure the course, not your performance in it. Thanks for your feedback!

Terminology

The following table contains vocabulary taken from the glossary in the back of the text and represent some of the important terms in genetics. For each word, indicate your familiarity with the term in the context of biology and genetics (several have non-scientific meaning as well) on a 1-5 scale as shown. We did not necessarily discuss all of these words, nor do we expect that you are familiar with all of them.

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
allele	1	2	3	4	5
amino acid	1	2	3	4	5
anaphase	1	2	3	4	5
annealing	1	2	3	4	5
anticodon	1	2	3	4	5
autosomes	1	2	3	4	5
backcross	1	2	3	4	5
bacteriophage	1	2	3	4	5
carrier	1	2	3	4	5
cell fate	1	2	3	4	5
centromere	1	2	3	4	5
chiasma	1	2	3	4	5
chromosome	1	2	3	4	5
cis-acting	1	2	3	4	5
clastron	1	2	3	4	5
clone	1	2	3	4	5
consensus sequence	1	2	3	4	5
cyclic AMP	1	2	3	4	5
denaturation	1	2	3	4	5
diploid	1	2	3	4	5
dominance	1	2	3	4	5
electrophoresis	1	2	3	4	5
endonuclease	1	2	3	4	5
eukaryote	1	2	3	4	5
evolution	1	2	3	4	5

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
founder effect	1	2	3	4	5
frameshift mutation	1	2	3	4	5
gamete	1	2	3	4	5
gene	1	2	3	4	5
genetic code	1	2	3	4	5
genome	1	2	3	4	5
genotype	1	2	3	4	5
hemizygous gene	1	2	3	4	5
heterozygous	1	2	3	4	5
housekeeping gene	1	2	3	4	5
hybrid	1	2	3	4	5
imprinting	1	2	3	4	5
inbreeding	1	2	3	4	5
intron	1	2	3	4	5
knockout mutation	1	2	3	4	5
lagging strand	1	2	3	4	5
linkage	1	2	3	4	5
major groove	1	2	3	4	5
meiosis	1	2	3	4	5
messenger RNA	1	2	3	4	5
metaphase	1	2	3	4	5
natural selection	1	2	3	4	5
nondisjunction	1	2	3	4	5
nonsense mutation	1	2	3	4	5
nucleus	1	2	3	4	5

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
Okazaki fragment	1	2	3	4	5
open reading frame	1	2	3	4	5
operon	1	2	3	4	5
pedigree	1	2	3	4	5
phenotype	1	2	3	4	5
plasmid	1	2	3	4	5
polymerase	1	2	3	4	5
polymerase chain reaction	1	2	3	4	5
population	1	2	3	4	5
primer	1	2	3	4	5
probe	1	2	3	4	5
promoter	1	2	3	4	5
protein	1	2	3	4	5
recessive	1	2	3	4	5
recombination	1	2	3	4	5
reporter gene	1	2	3	4	5
repressor	1	2	3	4	5
restriction enzyme	1	2	3	4	5
reverse transcriptase	1	2	3	4	5
ribosome	1	2	3	4	5

How well do you know these terms in the context of biology/genetics ?	Totally unfamiliar	Word is vaguely familiar	Hazy understanding	Okay understanding	Confident understanding
ribozyme	1	2	3	4	5
semiconservative replication	1	2	3	4	5
sex-linked	1	2	3	4	5
silent mutation	1	2	3	4	5
sister chromatids	1	2	3	4	5
Southern blot	1	2	3	4	5
spooling	1	2	3	4	5
sticky end	1	2	3	4	5
telomere	1	2	3	4	5
template	1	2	3	4	5
testcross	1	2	3	4	5
transcription	1	2	3	4	5
transfer RNA	1	2	3	4	5
transgenic	1	2	3	4	5
translation	1	2	3	4	5
transposable element	1	2	3	4	5
wildtype	1	2	3	4	5
X chromosome	1	2	3	4	5
zygote	1	2	3	4	5

Now please briefly *define the following words or describe their biological significance.*

anticodon	genetic code
autosomes	genome
cell fate	intron
consensus sequence	linkage
eukaryote	nonsense mutation
gamete	operon

primer	sticky end
semiconservative replication	telomere
spooling	transposable element

Concepts

Please answer the following multiple choice and short answer questions. Please provide an answer for each question, even if it is a wild guess. Remember, you are not being graded on this survey.

How does the genetic material in one of your skin cells compare to the genetic material in a liver cell?

1. The genetic material is identical in both cells.
2. The two cells share many common genes, but largely contain genes specific to its cell type.
3. Each cell contains only the genetic material for its cell type.
4. It depends upon which specific cells are compared.

What does it mean, at the *molecular* level, to say that a trait is dominant?

You are asked to serve on a jury considering a paternity case. A single mother of a baby is trying to get the father to pay his share of child support. The man she has identified as the baby's father denies the charge. They agree to perform a genetic test; they obtain DNA samples from the child, mother, and accused father and look at the length of DNA fragments resulting from a restriction enzyme digest in a certain region of the human genome. The child shares some of the DNA fragments with the mother but also has several fragments not shown in the mother's sample.

(a) If the man's sample contains fragments which match up exactly with those of the mother, what, if anything, could you conclude about the accused man's paternity?

(b) If the man's sample contains fragments which match up with those of the child *not* present in the mother, what, if anything, would you conclude about the accused man's paternity?

(c) If the man's sample contains fragments dissimilar to both those of the mother and child, what, if anything, could you include about the accused man's paternity?

Background

Which topic(s) in the course do you still not feel confident about?

Which topic(s) in the course that you now understand were most difficult for you to master?

Which topic(s) did you think you understood before the course but now realize that you did not?

Which topic(s) were not discussed in enough detail?

Which topic(s) were discussed in too much detail?

Which study strategies were most effective in helping you to learn the material?

Which study strategies were most effective in helping you to perform well on exams?

Did you find that the exams accurately assessed your understanding of the material? Why or why not?

Which parts of the course were most surprising or unexpected for you?

Has the course led you to alter your career goals or concentration? If yes, please explain briefly.

Name (*only for research, not grades**): ____

Class: '01 '02 '03 '04 GSAS Other ____________ Sex: M / F

Concentration (actual or predicted): ____

Are you pre-med? Y / N Country in which you attended high school: ________

**Names are only asked so we can compare individuals' responses over the semester. The survey has no effect on your grade!*