Teaching with Technology to Engage Students and Enhance Learning

Daniel Lass[1] , Bernard Morzuch[2], and Richard Rogers[3]

Abstract:

Teaching technology effects on student learning in a large lecture introductory statistics course were tested. Findings show in-class personal response systems and on-line homework/quizzes significantly improve student exam scores. We infer proven small class techniques, participating in class and doing homework via technologies, can restore sound pedagogy in larger classes. The experiment was conducted using just one class, but factors usually unaccounted for in assessment research were controlled, especially the instructor and other materials. The technologies investigated here can provide learning benefits to students even in larger courses often criticized for their inability to provide students quality learning experiences.

_____

[1]Daniel Lass, Department of Resource Economics
University of Massachusetts, 211 Stockbridge Hall
80 Campus Center Way, Amherst, MA 01003
E: dan.lass@resecon.umass.edu   P: 413-545-1501   F: 413-545-5853

[2]Bernard Morzuch, Department of Resource Economics
University of Massachusetts, 213 Stockbridge Hall
80 Campus Center Way, Amherst, MA 01003
E: morzuch@resecon.umass.edu   P: 413-545-5718   F: 413-545-5853

[3]Richard Rogers, Office of the Provost
University of Massachusetts, 372 Whitmore Administration Building
181 President's Drive, Amherst, MA 01003
E: rogers@provost.umass.edu   P: 413-545-2554   F: 413-577-3980

Teaching with Technology to Engage Students and Enhance Learning

Daniel A. Lass
Professor of Resource Economics
University of Massachusetts,
Amherst, MA

Bernard J. Morzuch
Professor of Resource Economics
University of Massachusetts,
Amherst, MA

Richard T. Rogers
Professor of Resource Economics and
Faculty Advisor to the Provost for Undergraduate Education
University of Massachusetts
Amherst, MA

Teaching with Technology to Engage Students and Enhance Learning

Abstract

Teaching technology effects on student learning in a large lecture introductory statistics course were tested. Findings show in-class personal response systems and on-line homework/quizzes significantly improve student exam scores. We infer proven small class techniques, participating in class and doing homework via technologies, can restore sound pedagogy in larger classes. The experiment was conducted using just one class, but factors usually unaccounted for in assessment research were controlled, especially the instructor and other materials. The technologies investigated here can provide learning benefits to students even in larger courses often criticized for their inability to provide students quality learning experiences.

Teaching with Technology to Engage Students and Enhance Learning

Introduction

Teaching is central to any higher educational institution, but it is not the only activity expected from the faculty, especially at large, public research universities. As universities provide greater access to more students with more diverse backgrounds and training, faculty are asked to teach more students, yet increase research output, bring in more grant funding, along with providing more service to the university, the community, and the profession. And all of this is to be done while containing the cost of an education. This pressure creates a tension between research and teaching and, given that resources are fully employed, it stands to reason that to increase one activity is to reduce another.

Faculty respond to what is most highly valued at their institution. Siegfried and White analyzed determinants of salary for faculty of the University of Wisconsin Economics Department. They found that there were rewards for both teaching and research, but those for research exceeded the rewards for teaching. Faculty realize they must "publish or perish" and that merit pay increases are driven by research success. It is common for faculty to use grant funds to buy out their teaching time to allow more time for research, but seldom, if ever, is the reverse observed.

Becker, Highsmith, Kennedy and Walstad (1991) expressed concern about declines in research on teaching and factors that affect the delivery and effectiveness of instruction. At the same time, economics enrollments declined during the early 1990s (Becker, 1997). Economists responded with more attention to teaching research and more time teaching. The number of sessions devoted to teaching at annual meetings increased during the mid-1990s and into the 21st century (Becker 2000) and the median research university economist increased her or his percent of time teaching from 35 in 1995 to 40 percent. Becker and Watts (2001) hypothesized that increases in emphasis on teaching might lead to changes in the way economics was taught. But they concluded that little changed in the way economics is taught at universities. Becker and Watts (2001) write: "Our results show that the dominant picture of the U.S. undergraduate economics teacher continues to be a male, … who lectures to a class of students as he writes text, equations, or graphs on the chalkboard, and who assigns students readings from a standard textbook (p. 448)." Thus, despite reporting that more time is devoted to teaching, economists continue to teach as they traditionally have and are seemingly reluctant to consider alternatives to the chalk-and-talk standard.

Becker, Highsmith, Kennedy and Walstad (1991) suggested a number of teaching research questions that warranted attention, among them was the evaluation of new teaching technologies. Research on how classroom technologies affect learning is scarce. Becker and Watts (1995) discussed a number of options available to economists. But results of their national survey in 1996 (Becker and Watts, 1996) led them to conclude that economist typically did not adopt teaching innovations, perhaps for reasons of efficiency or due to the lack of incentives offered. Economists may also be reluctant to adopt given there is a lack of good empirical evidence that innovative teaching methods increase student learning. Recently, Ball, Eckel and Rojas (2006) tested whether a wireless interactive system (WITS) could improve student

performance in principles of microeconomics. They found that final exam scores for the section using WITS did improve. When controlling for other factors, they found that scores increased by 7.4 points for the WITS section over the control section, but the cost of the system used, wireless PDAs, would be an issue. From the instructor's view, they found that students rated the instructor higher in communicating subject matter, higher overall and found the course more stimulating. These results suggest some incentives for faculty considering course revision and adoption of technology. If personnel committees pay attention to teaching evaluations, then there may be some payoff.

Cost may be an issue in faculty's abilities to employ technologies in classroom instruction. Becker and Greene (2001) suggested faculty are often unwilling to abandon the "chalk and talk" style of teaching of quantitative methods due to institutional constraints, poor funding and a reward structure that does not encourage innovation in teaching. Still others may not be convinced that there will be positive educational benefits. Judson and Sawada, in their review of over 30 years of literature suggest that it is "…the pedagogical practices of the instructor, not the incorporation of the technology as being key to student comprehension (p. 167)." They concluded that, while students will favor the use of classroom response systems, it is really the interactive engagement that occurs that is important to gains in student comprehension. Use of a classroom response system for quizzing alone may not provide benefits. Using a classroom response system to stimulate thought and discussion is more likely to achieve the intended results. Use of a classroom response system can improve conceptual learning as students hear, see and apply the concepts. As Kennedy suggests, "…brilliant expositions seldom cause students to fully understand; such understanding comes through working out problems based on the concept…(p. 489). For example, Kennedy advocates use of computer simulations to illustrate key concepts such as a sampling distribution for statistics or econometrics. With a classroom response system and enough random samples, students in large classes can replace the need for simulations.

Most all faculty must teach, but they need not teach the same type or number of courses. The number of courses that one must teach is negotiated by new hires and department stars as they strive to insure time for research needed for the publications that will result in either tenure or an enhanced professional reputation or both. Especially desired are the small graduate courses where the material and the students are aligned with the professor's research. The dreaded courses often are the large undergraduate general education courses. Even once the type and number of the courses have been determined, the quality of the teaching is often left unmeasured and only becomes an issue if it is so bad that many students complain. Developing new technologies and new teaching methods that will improve students' performances are desirable if they can be done without increasing time costs for teaching. Developing new techniques would also be desirable without demonstrated learning benefits if they release time for research. New teaching methods/technologies that draw resources from research should be subjected to rigorous cost/benefit analyses. As Siegfried and White concluded, "…Whether or not they are attractive depends on the quantitative magnitude of the improved economic education benefits relative to the costs of foregone research findings (p. 315)."

What is at risk in this time allocation between teaching and research is the quality of the undergraduate student's learning. Faculty who were asked to teach ever larger numbers of

diverse students soon discovered that many undergraduates enrolling in introductory courses were willing to accept a course structure that asks little of them and, thus, in return they ask little of the faculty. Course projects, graded homework, and term papers were cut from the syllabus as it would be timely and costly to even grade the work, much less check for its originality. The number of exams was reduced, while overall grades at the institution went up. Faculty members found this approach to teaching lessened student demands on their time and reduced student complaints over grades, allowing more time for their research. Students were left to work on their own, hopefully driven by their desire to learn the material. The top students did, but many students were seduced into procrastination or even denial of any need to know *this* material. They became confused as to what they should be buying with their ever larger tuition payments: a diploma whose value depends on the reputation of the institution or knowledge gained from learning. Students even turned to web sites to find those courses/sections with instructors who have an easy grading history before selecting their courses. Taken to its extreme, this reduced emphasis on student learning led to devastating results as students failed to learn and often failed more advanced courses and eventually out of the university. Such scenarios have led to serious challenges that higher education is not fulfilling its full obligations, especially to its undergraduate students.

Higher education escaped much of the criticism leveled by parents and politicians at the primary- and secondary-education levels. However, during the last decade that scrutiny has broadened to include higher education. The Association of American Colleges and Universities (2002) has called "… for a dramatic reorganization of undergraduate education to ensure that all college aspirants receive not just access to college, but an education of lasting value (p. vii)." They argue that change is needed because "… even as college attendance is rising, the performance of too many students is faltering. Public policies have focused on getting students into college, but not on what they are expected to accomplish once there (p. vii)."

The issues are many, complex, and costly; they require corrective action from all involved. It is not that we are at a loss as to what should be done, rather we lack clear incentives to reinvest in proven methods to improve student learning. In 1987, Chickering and Gamson published "Seven Principles for Good Practice in Undergraduate Education" to assist those interested in improving teaching and learning. In their original work, they identified the following for good practice in undergraduate education:

1. encourages contact between students and faculty,
2. develops reciprocity and cooperation among students,
3. encourages active learning,
4. gives prompt feedback,
5. emphasizes time on task,
6. communicates high expectations, and
7. respects diverse talents and ways of learning.

Large lecture classes pose the most difficult challenges in applying the "Seven Principles" and are often pointed to as the reason students are failing to learn at our large, public universities where enrollments in some introductory courses surpass 1,000 students. In the typical large class the instructor lectures to a subset of the course's actual enrollment of usually anonymous students who take notes without the opportunity to ask questions or interact with the

instructor or fellow students. Hiring additional quality professors to teach smaller classes would address the problem but come at a huge cost that society is unwilling to pay. Teaching technologies are seen as a possible method to enhance student learning and avoid the common convention that higher quality requires higher costs (Twigg, 2003). Becker and Greene, however, express this concern: "Nevertheless, institutional constraints such as inappropriately supported and maintained computer labs, for example, may work against an instructor's eagerness to abandon chalk-and-talk methods. In addition, the reward structure may not recognize the efforts of instructors who bring technology and current issues into their teaching. As pointed out by Becker and Watts (1999), traditional end-of-semester student evaluations of instructors may actually deter innovation because they seldom include questions about the instructor's use of technology in real-world applications. Change can only be expected if students, instructors and their institutions are amenable to it (p.181)."

The three authors of this paper share a passion for teaching. They joined other colleagues at their university in an effort to redesign their large-enrollment courses to incorporate the "Seven Principles" by using technology in appropriate ways. They were encouraged by Chickering and Ehrmann (1996) who revisited these principles by addressing the "… cost-effective and appropriate ways to use computers, video, and telecommunications technologies to advance the Seven Principles ... (p. 3)", by de Vry and Brown (2000), and by course redesigns done by The Center for Academic Transformation at Rennselaer Polytechnic Institute.

Most faculty have to look beyond their institution to find the funds necessary to implement major teaching changes. We were fortunate to have had the Davis Educational Foundation show an interest in our efforts to redesign our large courses to improve learning by integrating technology into our teaching methods. But what technology actually works? Many questions remained unanswered, leading Ehrmann (2003) to ask: "… Is it true that research has never proved that technology improves learning … (p. 6)?" Much has been published, but much remains to be done as Singer (2006) suggests.

The authors of this study used their collective experiences to explore needed changes in the one class they all taught, Introductory Statistics. They strongly embraced the "Seven Principles" and challenged themselves to successfully apply these principles in their classes in general and to Introductory Statistics in particular. All three authors are winners of teaching awards. Two of the authors first adopted WebCT to improve communication and reinstate graded homework. Hand written homework had been abandoned because it was costly to grade by hand and their past analysis showed no positive association between the homework grades and exam performance in their courses. In their course redesigns, they used a carrot-and-stick approach. Specifically, before each lecture students were required to complete an on-line assignment. Students were permitted do the assignment an unlimited number of times before the start of a lecture (actually, 20 minutes before lecture began to discourage last minute, last ditch, frantic attempts). The highest score before the deadline was recorded as the grade, but the student could reference the material anytime for future study. This approach was used to encourage students to stay on pace and to avoid the all too common "coast and cram" studying style. One author was an early adopter (in 1998) of a Personal Response System (PRS) for use in class and another adopted it soon thereafter. Both professors adopted these tools based on the theory of sound teaching and their belief that they addressed the "Seven Principles." However, at the time they

adopted these two technologies they did not have strong empirical evidence that the technologies improved learning. They each analyzed their own course data suggesting positive correlations between these technology tools and student exam performance, but their assessments failed to meet many of the criticisms in Singer's (2006) top ten list.

The third author of this research was a cautious observer of these technology efforts; he was intrigued but not convinced. He was also a "gold standard" for teaching quality as his teaching awards affirm. He possesses unique gifts, including the ability, and willingness to expend the effort, to learn and recall the names of each and every one of his students, even as class size exceeded 200 students. That skill alone goes a long way to offset the negatives associated with large lecture courses as he could use this "personal" touch to motivate and to cajole, just as is customarily done in a smaller-sized course. However, one cannot call on each student in a class of 200 students during a 50-minute lecture as can be done in a class of 20 students.

He also was the ideal professor to test whether these technology tools lead to improved learning. In a controlled experiment, he adopted two technologies, on-line web-based learning (OWL) and a personal response system (PRS), one at a time, over the course of two semesters after first beginning with a controlled, no-technology semester. Effectively, an experimental design was established and data were recorded to assess the usefulness of these teaching tools. This study reports on the results of this experiment.

The Two Teaching Technologies Used

Online Web-based Learning

The online web-based learning system (OWL) used in this experiment is an automated, web-based homework system developed at the authors' home university in 1996. Much of OWL's development has been funded by external curriculum development awards from the National Science Foundation and Department of Education (FIPSE), which have allowed the creation of many powerful, customized features for specific domains such as chemistry, mathematics, and computer science. OWL has improved student performance, when used effectively by the instructor rather than merely tacked on to an existing course that was not redesigned to incorporate seamlessly the OWL activities.[1]

OWL was developed when the early Course Management Systems, e.g., WebCT, were gaining acceptance on higher education campuses. OWL was not designed to be a course management system but instead to do online homework exceedingly well. Its features include easy content delivery, interoperability allowing use with other tools, and many customized features all of which make it more powerful than what could be done in WebCT.

The two early adopters (in 1998 and 1999) first used WebCT in their introductory statistics courses to assign homework with the quiz feature of WebCT rather than have teaching assistants grade homework that was passed in once a week. These two professors shared the task

---

[1] More information on the online web-based learning system is available upon request.

of creating quality questions. The quiz feature in WebCT allowed for random draws from a pool of questions and did allow for basic formula-driven questions based on a range of values for the parameters.[2] However, questions were limited to basic arithmetic formulas, and the system stored the values as part of the question rather than use a more robust algorithm to pick values. In 2000, as part of the Davis Education Foundation Grant the WebCT quiz tool was dropped in favor of OWL because of its more advanced features for presenting material and problems to students. The Davis Grant funds were used to develop OWL content material, including "Stats Tools" (i.e., interactive widgets that aided students' understanding and calculations) and a data base of questions.

One of the most attractive features of OWL over WebCT is the feedback feature built into its questioning tool. In OWL, the feedback can be tied to the actual data values used by the student during that visit to the question. Thus, students see an explanation that follows their calculations step by step, and they can discover exactly where they went wrong. In WebCT, students are presented with "an example" using similar, but not exactly the same, numbers the students encountered, which proves much less valuable to students. Immediate quality feedback is indeed one of the best of the "Seven Principles" for improved learning as students' maximum curiosity is captured as to why they missed a question.

OWL was used to deliver pre-lecture activities to students that were low stress, enjoyable introductions to the material planned for the upcoming lecture. Often they included a content page explaining new jargon and concepts, an interactive activity, and some basic questions to check understanding. These pre-lectures were designed to encourage students to keep up with the course by rewarding timely completion of relevant learning activities. Once the actual in-class lecture began, the student could access the material for review only but could not improve that pre-lecture's score. For the statistics course used in the experiment, there were 33 pre-lecture assignments during the course of the semester. Pre-lecture assignments were not due on the days of scheduled exams. Student performance on pre-lecture material was incorporated into the course grade accounting for 5% of the student's course grade.

In addition, OWL was used for higher stakes pre-exam quizzes where students would answer exam-like questions. The data values and parameters that defined these questions would be chosen randomly by the OWL system offering an unlimited number of questions. During each of the semesters in this study, there were three exams administered outside of class and a final exam administered during the final exam week. Each of the four exams lasted two hours. All questions were open-ended and objective. True/false and multiple-choice questions were never used. Students seemed rather surprised that such a large course required open-ended responses to be hand written and hand graded.

Approximately four days prior to each exam, an OWL quiz was opened to all students enrolled in the course. Each quiz consisted of approximately 12 questions. The sentence structure of each question was identical for all students. Anything related to quantitative input to the question, however, differed for each student. For example, one student might be given an hypothesis test problem and be required to find the critical value of a t-test statistic at the five-

---

[2] For example, students could be asked to compute the mean and standard deviation of the numbers X_1, X_2, X_3, ..., X_k, where values were picked at random.

percent level with 29 degrees of freedom, while a colleague may have to find the critical value at the one-percent level with 17 degrees of freedom. For each quiz, the student had three tries at each question, but the parameters would change for each attempt. Since this was a quiz, feedback to a question was not immediate and each quiz closed the day before a given exam was administered. Once the quiz was closed, detailed feedback was provided and the students could review the questions with immediate feedback. Students' performances on OWL quizzes accounted for 10% of their course grade.

Personal Response System

The Personal Response System (PRS) is a classroom communication system that allows students to respond to questions posed by the instructor. Such systems have been around for decades, but only in the last few years have they been widely used in higher education (Judson and Sawada). This system can be used for polling, evaluating student comprehension, and quizzes. As such, it allows for active participation and learning even in large classes. Each student has a wireless transmitter, commonly called a "clicker," resembling a remote control. Students use their clickers to answer questions projected in front of the class. Aggregated results are then displayed in a bar chart for all to see. PRS can be used to ask questions, gather live data to use throughout a lecture, gauge student understanding, and introduce new topics by challenging students' intuition about a subject. All students can participate without other students knowing their personal answers. After each PRS activity, they discover the class's collective opinions or the percentage of their fellow students who knew the answer to a question. The key to effective use of PRS is the quality and the timing of the questions asked in a lecture. PRS should not be used merely to take attendance, but to increase active attendance by improving lectures and engaging students.

PRS was adopted during the final semester of this study. The tool was used during 33 of the 42 lecture periods. The discrepancy between actual number of lectures and lectures during which PRS was used was primarily due to the relative novelty of the technology at the time. This being a large course, and with much maneuvering at the beginning of a semester due to adding and dropping courses, the instructor wanted to eliminate all possible reasons for students not having their clickers. In addition, the professor experienced problems with the technology during two class periods resulting in no PRS during those particular lectures.

The challenge to using PRS was in weaving it into the lecture. Obviously, this addition takes up valuable class time. Serious effort had to be made in making the transition from lecture material to PRS questioning as seamless as possible. Prior to PRS adoption, the instructor always posed questions to individual members of his audience to get a sense of their understanding. Initially, PRS felt like a burden. Gradually, and because of practice, the instructor came to realize that PRS was filling the roll of "hearing" at once what all of his students had to say about a question rather than just one student.

The instructor liked to use PRS at the beginning of his lectures. He asked questions that addressed material that was most recently covered. From the instructor's perspective, this served as a barometer regarding whether he was successfully communicating the material. He found PRS to be an attention-getter. It seemed to promote focus for the day's material. Questions

varied from brief to somewhat involved prose statements, depending upon the nature of the topic that was being addressed. Answers were in the form of multiple choice and he normally included two or three PRS questions during each lecture period. For each question, a student would earn credit in two parts: simply attempting a question, irrespective of correctness, earned partial credit; and a correct response earned the remainder of the credit. PRS questions accounted for 5% of the student's course grade.

Teaching Philosophy of the Instructor

The professor whose course was used in this study has taught Introductory Statistics for 28 years to more than 8,000 students with course sizes now exceeding 200 students. If one were to select at random a college student who has taken a statistics course and ask for an opinion about the course itself, the student would probably describe it as boring. Anyone who has taught statistics can feel those vibes in the classroom and therefore understand that assessment. This professor's motivation in teaching is to put a dent in that stigma. He is driven to promote clarity for a subject that is renowned for getting a bad rap. Sometimes, this criticism has poor delivery by the instructor as its source. Often, adoption of a fresh twist or pedagogy is all that is needed to promote understanding and, therefore, appreciation.

Two principles guide this instructor's lecture delivery: rehearsal prior to presentation and student involvement. He always practices a lecture before presentation putting himself in the position of being the student who has never before been exposed to the material. He consciously asks himself if connections are being made with the material so that a first-time observer is able to grasp it. In addition, after each lecture, he immediately constructs a detailed log of what transpired during the lecture. Regarding the second principle, he has developed a style in teaching that directly involves his students in each lecture. A prerequisite for involvement is learning the names of students, irrespective of class size. This requires a time investment he sees as necessary because it is an integral part of the pedagogy that makes statistics more personal to a student. Anonymity in a classroom is a key ingredient for boredom taking hold. Knowing a student's name removes anonymity, promotes a sense of responsibility, and serves to combat apathy.

Over the years, he has been cautious about the pace at which he presents material. He feels the need to proceed at a pace that permits students to see what is going on while he presents a topic rather than feverishly copying notes as they observe material being flashed on a screen. His pedagogy has been to occupy a lot of chalkboard space, to write big, and to involve students (by calling on them personally) as he presents the material. This pedagogy is a natural barrier against moving too quickly and gives students time to see. Constantly moving from one end of a 25-foot-long chalkboard to the other likewise promotes eye movement by the students reducing the chance of falling asleep due to eyes being focused on a nonstationary target.

Through many years of delivery and practice, he has concluded that a statistics lecture can be looked at as a performance of a lively art. He finds it easy to remain excited about this subject and to carry the excitement into the classroom year after year. For the past several years he has made his course notes available to students in a Course Reader containing detailed notes. Feedback from

students is that the Course Reader is excellent regarding explanation of the material and presentation of examples.

For a good number of years he has felt comfortable with this style. He has been strictly a name-knowing/chalkboard-using instructor. He did not think that it could be improved. Nothing was ever broken; so he did not dare bother to change anything. As corroborating evidence, student evaluations have not hinted that the course was boring. On the contrary, many have commented that it was one of the best courses that they have taken and as the course evaluation data will show it continued to win such praise.

Davis Educational Foundation Grant Opportunity

The professor's two colleagues have taught the same introductory statistics course and have led the way in teaching with technology, especially for large classes. The grant from the Davis Educational Foundation provided him the opportunity to soften his "not-broke-don't-fix" philosophy and to conduct a well-designed assessment of these two teaching tools over a three semester period where one new tool would be introduced per semester. The grant allowed the construction of models that make comparisons in student achievement among his traditional style, his traditional style with an on-line homework system introduced, and his traditional style with both the on-line homework system and PRS included. The data gathered provide statistical evidence about lingering questions as to the value of these tools.

During the first semester of this experiment (Spring 2002), he continued his award-winning style with no changes, but data were gathered for this "control semester." In Spring 2003 (the next time he taught the same course), he replaced his "recommended practice problems" with student assignments using the online learning tool that colleagues had been building to support the basic introductory statistics courses. No other changes were made; he used the same textbook, and his Course Reader and his examination methods remained the same. The on-line homework system was used on two fronts: (1) for pre-lecture activities where students were encouraged by points and low-stress activities to learn about the topic of the day; and (2) for higher-stress quizzes that tested mastery of material that was already covered and that would be on the next exam. Finally, in Fall 2003 he adopted the in-class PRS technology for engaging students with questions during a lecture. Each student responded in class to individual questions with a "clicker." Again, he benefited from the efforts of his colleagues who had been using PRS for several semesters, but he adapted the use of both tools to fit his style of teaching.

Course Assessment Data

As part of the Davis research project, students were surveyed to learn how they perceived the changes made in teaching pedagogy. Students were asked about their perceptions of the course *workload*. It is interesting to find out whether students perceived the workloads to be higher in Spring 2003 and Fall 2003 than in Spring 2002, given that the course now required more graded activities. Students were also asked whether they were inclined to complete more of the *assigned readings* during Spring 2003 and Fall 2003 than in Spring 2002. Finally, students

were asked to *rate the course*. Our interest is whether, relative to Spring 2002, OWL pre-lecture assignments, scheduled OWL quizzes, and in-class PRS questions during Spring 2003 and Fall 2003 have negative, positive or no perceived impact on how students rate the course. These questions and the values associated with each response are shown in Table 1.

The critical part of assessing changes in teaching pedagogy is the impact on student performance. On-line pre-lecture exercises and quizzes provide instant feedback to students; thus, we expect that these tools should improve student performance. Use of PRS during class encourages attendance and participation during class. Students are more actively involved during class, which is expected to have a positive effect on grades. We asked students about their *grade expectations*. The grade categories for this question as well as grade categories for the actual total points earned through the semester are shown in Table 1. Comparisons of means and tests of differences in means are shown in Table 2.

We used a two-sample t-test to assess the possibility of a difference in the means of the key variables presented above between Spring 2002 and Spring 2003. Testing at a five-percent level of significance, evidence suggested that students felt there to be a significant increase in the course workload as the course moved from "recommended practice problems" to graded problems. The students also had significantly higher grade expectations and significantly higher actual grades for Spring 2003 relative to Spring 2002. While students found that they had to do more work, they did not react negatively when evaluating the course as the variable measuring overall course rating was not statistically lower.

When comparing Spring 2002 and Fall 2003, similar results held for course workload and expected grade. In Fall 2003, however, students did not have significantly higher grades than in Spring 2002 in this course and as a group their University grade-point averages were lower (Table 3). An additional interesting finding is that students rated the course higher in Fall 2003 than in Spring 2002, albeit at a six percent level of statistical significance.

Data were collected for three semesters: Spring 2002, Spring 2003, and Fall 2003. Course data include detailed information on scores for four exams, scores for four online (OWL) quizzes, scores for 33 OWL pre-lecture exercises, and scores from sets of questions on each of 33 lecture days when PRS was administered. Spring 2002 represents the "control semester" for this experiment. During this semester, online and PRS technologies were not used, and course data include only the detailed exam scores. During Spring 2003, OWL pre-lecture exercises and quizzes were introduced. During the final semester of the experiment, Fall 2003, the personal response system (PRS) was added. Data were also gathered for all three semesters from the university administration on a number of student characteristics. These include SAT scores, high school GPA, current GPA, and cumulative GPA. These are intended to measure innate ability and effort.

Summary statistics are presented in Table 3 for the combined course and student characteristic information for each semester. The numbers of observations illustrate two concerns. First, while students who withdrew from the course have been excluded, there are also students who do not withdraw officially from the courses but fail to attend exams later in the semester. For example, the number of students taking exams between the first and final exams decreased by eight in 2002, fourteen in 2003 and thirteen in 2004. Considering the model of student learning as a

recursive system results in our dropping these students from the final data set used in the regression analyses that follow.

Comparing performances and characteristics of Spring 2003 and Fall 2003 to those of Spring 2002, we find that Exam 1 scores were statistically greater in both Spring 2003 and Fall 2003. With the exception of the Fall 2003 average final exam, there were no other statistically significant differences among Spring 2002, Spring 2003, and Fall 2003. The Fall 2003 final exam average was statistically lower than the Spring 2002 final exam average.

There were also few differences in student characteristics among Spring 2002, Spring 2003, and Fall 2003. We did find that student SAT Math scores were statistically lower in Spring 2003 when compared to Spring 2002, as were the SAT Total scores. We also found that the average Fall 2003 cumulative GPA was statistically lower than the Spring 2002 average. Thus, we find that there are few statistical differences among the students taking statistics during these three semesters.

Empirical Models:

To analyze the effects of on-line and PRS teaching technologies on student performance, we proceed as follows. First, we use the control group, Spring 2002, to estimate a basic model of student performance by regressing student performance (exam score) on student characteristics. The set of student performance measures (all exam scores) is assumed to follow a recursive system; i.e., performance on a successive exam during the semester is specified to depend on the previous exam score. The entire performance model for the ith individual is:

$$Exam1_i = \beta_{10} + \beta_{11}CumGPA_i + \beta_{12}HSGPA_i + \beta_{13}Late_i + \beta_{14}Male_i + u_{1i};$$
$$Exam2_i = \beta_{20} + \beta_{21}CumGPA_i + \beta_{22}HSGPA_i + \beta_{23}Late_i + \beta_{24}Male_i + \alpha_2 Exam1_i + u_{2i};$$
$$Exam3_i = \beta_{30} + \beta_{31}CumGPA_i + \beta_{32}HSGPA_i + \beta_{33}Late_i + \beta_{34}Male_i + \alpha_3 Exam2_i + u_{3i};$$
$$Final_i = \beta_{40} + \beta_{41}CumGPA_i + \beta_{42}HSGPA_i + \beta_{43}Late_i + \beta_{f4}Male_i + \alpha_4 Exam3_i + u_{4i}.$$

The response variables Exam1, Exam2, Exam3, and Final are assumed to depend on characteristics of the student. The cumulative GPA (CumGPA) and high school GPA (HSGPA) are included to capture the innate abilities of the student as well as their work ethic. We expect that students who are innately gifted and those who work hard will have higher high school and cumulative GPAs and will perform better in statistics. Students who are anxious about taking a mathematics-related course, in this case statistics, may wait until well into their college careers to take statistics. Thus, we include a binary variable (Late) indicating whether students have waited until their final 30 credits to take statistics. We expect a negative sign for the variable Late in each equation of the model. We include a binary gender variable (Male) in the model as a control. We have weak prior beliefs about the contributing effect of this variable but include it to capture possible gender effects that remain apparent in SAT Math scores; i.e., that males tend to have higher SAT Math scores than females.

The course curriculum is cumulative, building on earlier concepts and material. Thus, students who make the effort to keep up with the course and learn the material as they go through the course are expected to perform better. We anticipate that prior exam performances will affect current

exam performance. The nature of the course and the structure of the model should lead to the most recent exam having the strongest effect on the current exam score. Because the model is recursive, the stochastic nature of the exam scores included in the set of predictor variables will not affect the properties of the estimators. Each equation can be estimated using ordinary least squares (OLS).

All exam questions were objective and open-ended. Typically, a situation was described, data were provided, and the student was required to execute the proper statistical procedure. Calculations were the obvious requirement for each problem. Partial credit was always awarded for proper set-up. The intent was to force the student to apply the methodology, complete with prose and recommendations, the way that an individual would be required to present a similar analysis in a professional setting. During the control phase of this experiment, we recognized that necessity of maintaining a common structure among all examinations over the three semesters. With the experience of applying statistics for well over 25 years, it was not difficult to write questions that accommodated this requirement. All exams except the final (which was administered during finals week) were administered outside of the regularly scheduled class period, usually on Wednesday evenings. The venue consisted of three large lecture halls. This offered students breathing space and minimized the possibility of wandering eyes. Requests for additional time were always honored. The lowest of the three exam grades counted 10% toward the final grade. The remaining two exams each counted 20%. The final exam counted 30%.

Teaching assistants graded all exams by hand. Partial credit was awarded according to a strict set of guidelines written by the instructor. The same set of guidelines applied for all three semesters. The instructor sampled the first subset of exams graded by each TA to determine if grading directions were being followed to the letter. If there were inconsistencies, the TA was required to regrade everything up to that point.

During Spring 2003 and Fall 2003, on-line web-based learning (OWL) exercises (Spring 2003) and PRS (Fall 2003) were added to the course pedagogy. Thus, the Fall 2003 model, including both on-line and PRS components, is:

$$
\begin{aligned}
Exam1_i &= \beta_{10} + \beta_{11} CumGPA_i + \beta_{12} HSGPA_i + \beta_{13} Late_i + \beta_{14} Male_i \\
&\quad + \gamma_{11} PRS1_i + \gamma_{12} Pre1_i + \gamma_{13} OWL1_i + u_{1i} ; \\
Exam2_i &= \beta_{20} + \beta_{21} CumGPA_i + \beta_{22} HSGPA_i + \beta_{23} Late_i + \beta_{24} Male_i + \alpha_2 Exam1_i \\
&\quad + \gamma_{21} PRS2_i + \gamma_{22} Pre2_i + \gamma_{23} OWL2_i + u_{2i} ; \\
Exam3_i &= \beta_{30} + \beta_{31} CumGPA_i + \beta_{32} HSGPA_i + \beta_{33} Late_i + \beta_{34} Male_i + \alpha_3 Exam2_i \\
&\quad + \gamma_{31} PRS3_i + \gamma_{32} Pre3_i + \gamma_{33} OWL3_i + u_{3i} ; \\
Final_i &= \beta_{40} + \beta_{41} CumGPA_i + \beta_{42} HSGPA_i + \beta_{43} Late_i + \beta_{44} Male_i + \alpha_4 Exam3_i \\
&\quad + \gamma_{41} PRSf_i + \gamma_{42} Pref_i + \gamma_{43} OWLf_i + u_{4i} .
\end{aligned}
$$

The teaching technology variables PRS1, Pre1 and OWL1 represent students' average scores for the period prior to Exam1. Thus, these variables represent students' efforts on the teaching technology components prior to Exam1. PRS2, Pre2 and OWL2 represent average scores for the period between Exam1 and Exam2. The same is true for the variables PRS3, Pre3 and OWL3, as well as PRSf, Pref

and OWLf. All are included to capture incremental effects of teaching technologies; cumulative effects are included in the effects of prior exams scores on current exam scores.

The Fall 2003 model offers several possibilities for hypothesis testing. We expect that technology will help students learn statistics; we expect positive effects for PRS, OWL pre-lecture and OWL quiz scores. Thus, we test the following hypotheses:

$$H_0: \quad \gamma_{jk} \quad \leq \quad 0; \qquad H_a: \quad \gamma_{jk} \quad > \quad 0; \quad \text{for } j = 1, 2, 3, 4 \text{ and } i = 1, 2, 3.$$

In addition to these right-tail tests, we test the joint hypothesis that the teaching technologies used have no effect on student learning as measured by exam scores:

$$H_0: \quad \gamma_{j1} = \gamma_{j2} = \gamma_{j3} = 0; \qquad H_a: \quad \text{at least one } \gamma_{jl} \neq 0.$$

The joint test can be conducted as an F-test for each of the four models ($j = 1, 2, 3, 4$).

Our goal is to test the effects of the teaching technologies introduced on student performance as measured by exam grades. To do so, however, it is important to consider the level of prior preparation that students might be expected to have with the material before they enter the course. Also, it is important to consider the motivation and incentive that a student brings to preparing for each of the four exams. In this way, we are able to suggest which of the equations provides the best test of the teaching technologies used. We had prior beliefs about how each equation would fare in terms of its predictive ability. Exam 1, for example, concentrates on descriptive statistics. This material is very basic and many students have seen it in high school. Consequently, we felt that probable prior exposure to this material may taint the potential explanatory power of our predictor variables for Exam 1. After Exam 1, the material achieves a new plateau in terms of level of difficulty and an even higher plateau after Exam 2.

We felt that Exams 2 and 3 would be the most appropriate equations for testing the technology because the playing field among students now has evened out in terms of prior exposure to this more difficult material. Students would not have mastered this material previously. Thus, we would expect students to prepare earnestly for these exams because the material is more difficult. Also, students recognize that doing well on these two exams lessens the pressure associated with the Final Exam. Finally, the Final Exam may be hindered by end-of-semester problems. Students may have resigned to certain grade expectations and decided that their scores on the Final Exam will not alter their final grades. Even if they are wrong, these preconceived notions affect their ability and willingness to prepare for the Final Exam. Other students begin to panic and realize that they must do extremely well on the Final Exam to get their desired grade, often just a passing grade to prevent the dreaded need to retake the course. In conclusion, we believe that the results for Exams 2 and 3 should provide the best test of the teaching technologies used.

Model Results

A basic model was estimated using the "control" sample from Spring 2002 semester. No technology treatments were included during Spring 2002; the course was taught as a standard lecture with a weekly discussion session for those who chose to attend. For the basic model, students' high school GPA and cumulative university GPA were used to measure innate abilities and historic effort. The cumulative GPA included the current semester, but the grade points from the statistics course were removed from the cumulative GPA. Alternative measures of innate abilities would be SAT scores. However, SAT scores were not available for all students resulting in a loss of degrees of freedom. Also SAT scores do not measure students' efforts. Preliminary tests showed that high school GPA and cumulative GPA performed as well or better in explaining Spring 2002 exam scores. In regression models where SAT scores, high school GPAs and cumulative university GPAs were included, tests supported the hypothesis that SAT verbal and math scores did not explain a statistically significant portion of the variation in exam scores. Thus, we settled on the use of cumulative GPA and high school GPA as measures of students' abilities and effort. We also dropped students from the data set who did not take all four of the semester exams. Thus, a consistent data set for 141 students was used for all models – Exam 1 through the Final Exam.

Estimated regression results are shown in Table 4. The estimated regression models fit the data well; the model explained between 38.1 percent of the variation (Exam 1 scores) to 54.4 percent of the variation (Exam 3 scores). The proportions of explained variation in exam scores were statistically important as shown by the calculated F-statistics. In all cases, the models explained a statistically significant portion of the variation in exam scores at the one percent level of significance or better.

The results show that students' cumulative GPAs are important explanatory variables. For Exam 1, both the high school GPA and cumulative university GPA have strong positive and statistically important effects on the exam score. Holding other factors constant, a student with a high school GPA that is one point higher (i.e., a 3.0 versus 2.0) would be estimated to earn an additional 6.8 points on Exam 1. We estimate that a B-student at the university (a cumulative GPA of 3.0) would earn an additional 13.5 points compared to a C-student (cumulative GPA of 2.0). The estimated effects of high school GPA were statistically significant only for Exam 1, and estimated effects of cumulative university GPA appear to diminish throughout the course. However, Exam 2, Exam 3 and the Final Exam all depend upon prior exam scores and these prior exam scores depend upon students' innate abilities through the recursive nature of the model. Every point a student scores on Exam 1 is estimated to improve the Exam 2 grade by 0.36 points. There is a stronger effect of Exam 2 scores on Exam 3 grades; each point on Exam 2 is estimated to improve the student's Exam 3 grade by 0.73 points. This increase in prior exam effect is logical; Exam 2 tests students on their knowledge of probability, while Exam 3 moves on to probability applications with the normal distribution. Students' final exam scores are higher by 0.55 points for every point earned on Exam 3; this is support for the correspondence between understanding probability distributions and hypothesis testing. As discussed above, we expected that predicting final exam scores would be challenging, but in Fall 2003 the results were the strongest found.

We had weak expectations for gender effects on learning statistics. Shibley Hyde, Fennema and Lamon (1990) reviewed a hundred studies and found gender differences were small, but when found favored males. Sosin, Blecha and Agarwal (2004) found that women did not perform as well as men on microeconomic questions, while there was no statistical difference on macroeconomic questions. We do not find statistical support for that hypothesis in our data. Indeed, the only statistically important effect for the male binary variable was a strong negative effect on the final exam. We anticipated, based on our collective experiences about students' attitudes toward the course, that students taking the course late in their careers may have been putting it off as they believe they are relatively weak in statistics. While the estimated coefficients for all four models (Exam 1 through Final Exam) are negative for the "late" variable, these effects are not statistically different from zero at our chosen significance level.

Our basic model suggests that better students who have historically done well in their academics also do better in statistics. Our measures of innate abilities and historic effort (high school and university cumulative GPA) do explain a statistically important portion of the variation in exam scores. We also find that prior exam scores are important explanatory variables for subsequent tests in statistics. It is clearly important to success in this statistics course that students arrive prepared, get a good start, and continue to keep up with the course throughout the semester.

The basic models estimated for Spring 2002, our control semester, were then used to predict Spring 2003 and Fall 2003 exam scores. During Spring 2003, the instructor incorporated on-line web-based learning (OWL) technologies in the curriculum. We included variables to measure performance on pre-lecture OWL exercises and OWL quizzes. For Fall 2003, PRS was introduced. During that semester, we incorporated variables measuring how well students performed on in-class quizzes (PRS), OWL pre-lecture exercises, and OWL quizzes. We tested hypotheses that the additional teaching technology variables do not affect the exam scores. If the teaching technologies have no effect on learning (as measured by exam scores) then these variables would not explain a statistically important portion of the variation in exam scores. Individual parameter tests are then used to illuminate individual variable effects on exam scores.

Spring 2003 results support our conclusions for the basic Spring 2003 model. In the interest of brevity, we do not include our Spring 2003 results; they are available upon request. We found strong cumulative university GPA and prior exam effects. Joint hypothesis tests supported the inclusion of pre-lecture OWL exercise and OWL quiz variables for Exam 1, Exam 2 and Exam 3. We find strong positive effects for OWL quizzes for Exam 1, Exam 2 and Exam 3. OWL pre-lecture exercises were statistically significant and positive for Exam 3 We did not find statistically important effects for the teaching technology variables for the Final Exam. Final exams are relatively difficult to predict as we discussed above.

Estimated regression models for Fall 2003 are shown in Table 5. The models include the variables in the basic model used for Spring 2002 as well as student measures for participation and effort on in-class PRS quizzes, pre-lecture OWL exercises and OWL quizzes. Included for each exam are the PRS, pre-lecture and quiz scores for that section of the course. Thus, for Exam 1, we include the PRS quiz averages, the average pre-lecture scores, and the OWL quiz score for exercises prior to Exam 1. For Exam 2, we include those exercises and quizzes that covered

material between Exam 1 and Exam 2. Again, the recursive model includes prior exam scores, which depend upon the technology scores relevant to that section. Therefore, Exam 2 scores indirectly depend upon PRS quizzes before Exam 1. We find that the models fit these data well with R-square values ranging from 0.47 (Exam 1) to 0.66 (Final Exam). We find that the models explain statistically important proportions of the total variations in exams scores; all calculated F-statistics for the models are highly significant.

Regression results for Exam 1 of Fall 2003 are consistent with results we found for the basic model of Spring 2002. We find statistically significant positive effects of both high school and cumulative university GPAs. The magnitudes of these effects are consistent with those for Spring 2002. While signs are different (positive versus negative) for gender and taking the course late, these variables are again statistically insignificant. We also find that there are strong positive effects of prior exams on current exam scores, consistent with our Spring 2002 results.

We find that the teaching technology variables play an important role in student learning as measured by exam scores. The F-test statistics shown in the final row of Table 4 test the restrictions that all three teaching technology parameters are jointly zero. The null hypothesis of no teaching technology effects is rejected in all cases at the one percent level of significance. Thus, we conclude that the teaching technologies employed have important effects on student learning.

Beginning with Exam 1, we find that OWL pre-lecture exercises have strong positive effects on Exam 1 scores. Each additional one-point increment in the pre-lecture average translates to about 0.30 points on Exam 1. Positive, but statistically unimportant effects are estimated for both PRS and OWL quizzes.

Strong pre-lecture effects are again observed for Exam 2. While PRS effects were estimated to be negative, they were not statistically important. The OWL quiz effect increased in magnitude, but is again statistically unimportant at the five percent level of significance.[3] For Exam 3, we observe strong positive effects for PRS and OWL quizzes. These estimates were both statistically important. The pre-lecture estimate decreased in magnitude; an additional point on the pre-lectures increased Exam 3 scores by 0.07 point.[4] The estimated technology effects for the Final Exam were all positive and highly significant. Each additional point earned on PRS quizzes added 0.09 points to the Final Exam score, while additional points on pre-lecture exercises and OWL quizzes added 0.11 and 0.12 points to the Final Exam score, respectively. The combined effect of an additional point on each technology component was 0.32 points for the final exam. These results suggest that nearly one-third of the final exam score can be explained by students' performances on the teaching technology components of the course. These strong results on the final exam were impressive given our expectation that the final exam scores would prove difficult to predict.

---

[3] The p-value for the OWL quiz estimate for a one-tail (right-tail) test was 0.0732. Adjusting our level of significance to 10 percent, consistent with strong priors for a positive effect would lead us to conclude that the effect of OWL quizzes was statistically important.
[4] Note again that the p-value for a right-tail test is 0.0557, very close to our chosen level of significance.

The trends in PRS effects are those that we might expect. While the professor incorporated the on-line exercises and quizzes of OWL during Spring 2003, the use of PRS was new for him during Fall 2003. Also, employing a classroom communication system is quite unlike using OWL where much of the material had been developed by his colleagues, modified by him to fit his course and style and then presented online. PRS requires a new thinking about how to ask good questions by the professor during his lectures. The use of PRS was somewhat different from his approach of asking individual students to participate during class. He had agreed to use PRS for this project and he put the kind of effort into learning to use the technology that we would expect from an award-winning teacher. PRS represented unknown territory and he was learning appropriate use. The weak early semester results likely reflect his learning curve. Once he became engaged with the teaching tool and gained experience from application during class, he obtained the results for his students that support our hypothesis; the PRS teaching technology can improve student performances. We have seen the importance of using PRS well in other departments as some teachers have great results and others find that students hate this interactive teaching tool. The key is that the teacher understands what makes for good questions that engage rather than frustrate or infuriate students.

Conclusions

Our findings should not surprise many teachers as we found statistically significant support for the hypothesis that increased and immediate quality feedback and increased time on task improve student performances. One could summarize our findings as: "Students should seek quality instructors, go to class and do their homework if they wish to learn the material." Thus, despite the cutting edge technology employed, the advice is very old fashioned, well-tested, and proven for success in school. Of course, it is easier to encourage students to follow that advice when class size is 20, not 200 or 500 or 1,000 or more as some class sizes have reached. We infer from our findings that proven small class techniques employed through technology can help to restore such sound pedagogy in larger classes.

We cannot conclude more than we have shown. The test we conducted was narrow as we measured the effects of two teaching technologies on learning in one class during one semester. Nevertheless, it was also a precise test as we controlled for many factors usually unaccounted for in other assessment research, especially the instructor and materials other than the added changes. This study is just a small step towards answering the many questions regarding the usefulness of teaching technologies. These teaching tools were applied by a renowned teacher who puts his students first. We would be the first to say that adopting technology without a strategic plan focused on pedagogical problems will prove unsuccessful. A bad lecture with PRS is still a bad lecture, but now with bad questions forcing students to attend a class of questionable value to earn silly attendance points that are not associated with additional learning. But, these technology tools in the hands of a dedicated teacher can provide learning benefits to students even in the larger courses that are often criticized for their inability to provide students with a quality learning experience.

Table 1. Assessment questions asked of students and their categorical response values.								
	Categorical Values							
Categorical Variable	1	2	3	4	5	6	7	8
Workload	one of the lightest	lighter than average	about average	heavier than average	one of the heaviest			
Readings Completed	all or almost all	about three-quarters	about half	about one-quarter	almost none			
Overall Rating	one of the best	better than average	about average	worse than average	one of the worst			
Grade Expected	A	AB	B	BC	C	CD	D	F
Grades Earned	90-100	85-89	80-84	75-79	70-74	65-69	57-64	0-56

Table 2. Comparison of means for student assessments of introductory statistics: Spring 2002 versus Spring 2003 and Fall 2003.

Variable	Spring 2002			Spring 2003			Fall 2003		
	n	Mean	Standard Deviation	n	Mean	Standard Deviation	n	Mean	Standard Deviation
Workload	139	2.928	0.738	166	3.162*	0.78	148	3.115	0.796
Readings Completed	139	2.057	1.075	166	1.945	1.08	147	2.211	1.068
Course Rating	139	2.064	0.986	166	2.096	0.929	148	1.899	0.823
Grade Expected	139	3.467	1.602	166	2.975*	1.366	148	3.068	1.697
Total Points Earned	154	75.882	15.554	182	79.6*	12.071	157	76.4	14.76

* Means are statistically different at the 5 percent level of significance or better.

Table 3. Descriptive statistics by semester for introductory statistics courses used in experiments.

Variable	Spring 2002			Spring 2003			Fall 2003		
	n	Mean	Std Dev	n	Mean	Std Dev	n	Mean	Std Dev
Exam 1	149	68.6	17.4	191	77.7	12.8	158	75.4	16.6
Exam 2	146	73.6	14.7	187	75.6	13.7	156	70.5	18.2
Exam 3	145	72.2	19.1	180	69.7	17.8	152	73.3	18.4
Final Exam	141	76.0	18.4	177	75.6	18.6	145	69.3	21.7
SAT Verbal	126	547	81	168	536	82	127	537	86
SAT Math	126	567	75	168	543	77	127	565	80
SAT Total	126	1114	132	168	1079	141	127	1102	140
High School GPA	149	3.22	0.54	191	3.31	0.47	158	3.20	0.52
Cumulative GPA	149	2.89	0.59	191	2.88	0.58	158	2.70	0.66
Current Semester GPA	149	2.71	0.83	191	2.76	0.91	158	2.54	0.98
Total Credits Taken	149	57.8	26.5	191	60.1	26.2	158	50.7	31.7
Gender (Male = 1)	149	0.537	0.500	191	0.429	0.496	158	0.557	0.498
Taking Course Late in Career	149	0.121	0.327	191	0.147	0.355	158	0.120	0.326
Pre-Lectures for Exam 1	NA	NA	NA	191	92.0	16.0	158	90.0	16.2
Pre-Lectures for Exam 2	NA	NA	NA	191	83.9	23.3	158	83.1	21.4
Pre-Lectures for Exam 3	NA	NA	NA	191	80.5	30.6	158	79.4	26.9
Pre-Lectures for Final	NA	NA	NA	191	77.0	32.0	158	69.4	33.2
OWL Quiz for Exam 1	NA	NA	NA	191	87.7	17.0	158	85.5	30.1
OWL Quiz for Exam 2	NA	NA	NA	191	82.3	27.1	158	70.4	34.8
OWL Quiz for Exam 3	NA	NA	NA	191	71.9	34.4	158	74.4	33.2
OWL Quiz for Final	NA	NA	NA	191	78.3	33.9	158	69.2	39.8
PRS for Exam 1	NA	NA	NA	NA	NA	NA	158	59.6	26.9
PRS for Exam 2	NA	NA	NA	NA	NA	NA	158	58.1	29.4
PRS for Exam 3	NA	NA	NA	NA	NA	NA	158	49.2	28.7
PRS for Final	NA	NA	NA	NA	NA	NA	158	43.4	33.3

Table 4. Estimated regression models of student learning in introductory statistics, Spring 2002.

Variable	Exam 1 Coefficient	Exam 1 St. Error	Exam 2 Coefficient	Exam 2 St. Error	Exam 3 Coefficient	Exam 3 St. Error	Final Exam Coefficient	Final Exam St. Error
Intercept	9.50	8.24	23.85	6.24	-6.64	7.80	12.99	8.32
Exam 1			0.363*	0.065				
Exam 2					0.732*	0.092		
Exam 3							0.549*	0.079
Final Exam								
Cumulative GPA	13.46*	2.36	10.06*	1.98	5.38*	2.52	5.39*	2.69
High School GPA	6.83*	2.57	-1.65	1.98	2.47	2.30	3.36	2.58
Taking Course Late in Career	-1.91	3.48	-0.267	2.63	-0.993	3.12	-2.43	3.48
Gender (Male = 1)	-2.31	2.33	2.29	1.76	3.51	2.09	-7.25*	2.36
R-square	0.381		0.495		0.544		0.503	
F-statistic	20.96*		26.51*		32.28*		27.27*	
n	141		141		141		141	

* Estimated coefficient is statistically different from zero at the 5 percent level of significance, or better.

Table 5. Estimated regression models of student learning in introductory statistics, Fall 2003.								
	Exam 1		Exam 2		Exam 3		Final Exam	
Variable	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept	-5.26	9.67	-10.69	8.79	2.10	7.00	-2.22	8.35
Exam 1			0.349*	0.087				
Exam 2					0.329*	0.063		
Exam 3							0.560*	0.088
Cumulative GPA	11.06*	2.19	3.89	2.66	5.23*	2.09	7.42*	2.70
High School GPA	5.13*	2.40	6.61*	2.58	3.52	2.17	-3.72	2.65
Taking Course Late in Career	4.65	3.28	2.74	3.52	-0.399	2.94	0.350	3.55
Gender (Male = 1)	1.58	2.19	2.80	2.30	3.53	1.89	-0.714	2.35
PRS	0.035	0.045	-0.041	0.045	0.142*	0.037	0.094*	0.040
Pre-Lectures	0.296*	0.085	0.233*	0.070	0.069	0.043	0.108*	0.050
OWL Quiz	0.045	0.044	0.060	0.039	0.104*	0.035	0.117*	0.038
R-Square	0.471		0.485		0.605		0.662	
Regression F-Statistic	17.44*		16.00*		25.99*		33.25*	
n	145		145		145		145	
Joint Test for Technology (F)	6.04*		5.83*		12.15*		12.42*	

* Statistically different from zero at the 5 percent level of significance or better.

References

Association of American Colleges and Universities. "Greater Expectations: A New Vision for Learning as a Nation Goes to College." National Panel Report, Washington, DC. 2002.

Ball, Sheryl, Catherine Eckel and Christian Rojas. "Technology Improves Learning in Large Principles of Economics Classes: Using our WITS." *The American Economic Review*, May 2006, 96(2), pp. 442-446.

Becker, William E. and William H. Greene. "Teaching Statistics and Econometrics to Undergraduates." *The Journal of Economic Perspectives,* Autumn 2001, 15(4), pp. 169-182.

Becker, William, Robert Highsmith, Peter Kennedy and Willam Walstad. "An Agenda for Research on Economic Education in Colleges and Universities." *American Economic Review,* May 1991, 81(2), pp. 26-31.

Becker, William E. and Michael Watts. "A Review of Teaching Methods in Undergraduate Economics." *Economic Inquiry*, October 1995, 33(4), pp. 692-700.

Becker, William E. and Michael Watts. "Chalk and Talk: A National Survey on Teaching Undergraduate Economics." *American Economic Review*. May 1996, 86(2), pp.448-453.

Becker, William E. and Michael Watts. "How Departments of Economics Evaluate Teaching." *American Economic Review*. May 1999, 89(2), pp. 344-350.

Becker, William E. and Michael Watts. "Teaching Economics at the Start of the 21st Century: Still Chalk-and-Talk." *American Economic Review*. May 2001, 91(2), pp. 446-451.

Chickering, Arthur W. and Stephen C. Ehrmann. "Implementing the Seven Principles: Technology as Lever." *AAHE Bulletin*, 1996, 48, pp. 3-6.

Chickering, Arthur W. and Zelda F. Gamson. "Seven Principles for Good Practice in Undergraduate Education." *AAHE Bulletin*, 1987, 39(7), pp. 3-7.

de Vry, Janet R. and David G. Brown. "A Framework for Redesigning a Course." In *Teaching with Technology*, ed. D. G. Brown, Anker Publishing Company, Bolton, MA, 2000.

Ehrmann, Stephen C. "New Ideas, and Additional Reading" http://www.tltgroup.org/programs/seven.html , October 2003.

Ehrmann, Stephen C. 1995. "Asking the Right Questions: What Does Research Tell Us About Technology and Higher Learning?" Change. The Magazine of Higher Learning, March/April 1995, XXVII(2), pp. 20-27.

Judson, Eugene and Daiyo Sawada. 2002. "Learning from Past and Present: Electronic Response Systems in College Lecture Halls." *Journal of Computers in Mathematics and Science Teaching,* 21(2):167-181.

Kennedy, Peter. "Teaching Undergraduate Econometrics: A Suggestion for Fundamental Change." *The American Economic Review*, May 1998, 88(2), pp. 487-492.

Siegfried, John and Kenneth White. "Financial Rewards to Research and Teaching: A Case Studey of Academic Economists." *The American Economic Review*, May 1973, 63(2), pp. 309-315.

Singer, David. "Ten Ways You Might Be Fooling Yourself about Assessment." *Campus Technology,* April 2006. [http://www.campus-technology.com/article.asp?id=18178].

Shibley Hyde, Janet, Elizabeth Fennema, and Susan Lamon. "Gender Differences in Mathematics Performance: A Meta-Analysis." *Psychological Bulletin*, 1990, 107(2), pp. 139-155.

Sosin, Kim, Betty J. Blecha and Rajshree Agarwal. "Efficiency in the Use of Technology in Economic Education: Some Preliminary Results. *American Economic Review*, May 2004, 94(2), pp. 253-258.

Twigg, Carol A. "Improving Learning and Reducing Costs: Lessons Learned from Round I of the Pew Grant Program in Course Redesign." Center for Academic Transformation, Rensselaer Polytechnic Institute, Troy, NY. 2003.