Innovation from Within the Box:
Evaluation of Online Problem Sets
in a Series of Large Lecture Undergraduate Science Courses

Evonne Schaeffer*
Tina Bhargava
John Nash
Charles Kerns
Scott Stocker

Stanford Learning Lab
Stanford University

Paper to be presented at the annual meeting of the American Educational Research Association (session 27.62). April 10-14, 2001, Seattle, Washington.

*Email correspondences regarding this paper should be addressed to Evonne Schaeffer <evonne@stanford.edu>.

Background

Student learning at research universities is often compromised by institutional constraints and traditions which undervalue innovative teaching practices. Here we address the context of the large lecture science class within the research university and our introduction of a technology-mediated solution to enhance the learning experience for students. Although typically it is difficult to sustain innovation after its introduction, the intervention described here has been adopted in several departments across campus, and is poised to be distributed and adopted more broadly in the near future.

The Challenge

Traditionally, the large lecture format is an efficient way to simultaneously educate many students. Sometimes large lectures are the most appropriate teaching format (Shulman, 2000); other times, constraints associated with large lecture courses work against the students' learning needs. The anonymity that typically defines the large lecture setting is often accompanied by infrequent and incomplete feedback both to students on their performance, and to faculty on their students' understanding. Communication between the students and the teaching team and among the students themselves is often sporadic; lecture attendance is often weak. Students in large lecture courses are often treated as passive consumers of information and are often not given responsibility for their learning. Because of these less than ideal conditions, student learning is often compromised (Boyer Commission, 1998; Friedlander & Kerns, 1998; Twigg,1999).

It is well known that in the classroom setting that regular, ongoing feedback is helpful for learning (AAHE, 1996; Bransford, Brown, & Cocking, 2000; Chickering & Gamson, 1987; Cross, 1998; Kulik & Kulik, 1988). In large lectures, partly because of the sheer number of students, feedback is typically infrequent and serves a summative function (Angelo, 1990). Many times, the first systematic feedback that students receive is their midterm exam score, which may come too late to allocate adequate resources in time for acceptable improvement on the final exam.

Faculty in the Program in Human Biology at Stanford University were concerned about addressing the shortcomings found in their series of large lecture core courses. Human Biology is the second most populous undergraduate major at Stanford University. The goal of this interdisciplinary program is to provide students with a broad biological and social science understanding of how humans interact with each other and their environments. All students majoring in Human Biology (approximately 200 – 250 per year) are required to take the year-long A- and B- sides of the "core" sequences. Figure 1 shows the two side-by-side five-unit courses students take each quarter, typically in their sophomore year. Each of these courses is lecture based, with four 1-hour lectures and one 1-hour discussion section meeting ("section") per week. Course assistants lead the sections in groups of about 20 students. Sections are typically used to review the voluminous content from the previous week's lectures and reading assignments. Even in the smaller section setting, it was difficult for the course assistants to monitor all of their students' performance and to provide timely, individual feedback on a regular basis.

	Fall	**Winter**	**Spring**
A-side	Genetics, Evolution, and Ecology	Developmental and Cell Biology	Human Physiology
B-side	Culture, Evolution, and Society	Biology and Culture in Human Development	Environmental and Health Policy

Figure 1. Human Biology A- and B- side required core sequences.

In working with the Stanford Learning Lab, the head faculty of the Human Biology program expressed concern that while the students were proficient at memorizing the course content, they were not adept at synthesizing and applying it. In addition, the faculty were concerned about timely feedback, because it was often the case in their program that assigned homeworks were not collected, and, when they were, they were not graded in a way that allowed for timely, personal, formative feedback. They also felt that didactic review of material was not the best use of section time and that sections could be more interactive and tailor instruction to match the students' level of understanding.

The Context

In general, introducing and sustaining technological and pedagogical change within the research university is challenging. Institutional constraints and traditions make teaching innovation in the research university setting rather limited. There are many layers of organization that are fundamentally unwilling or unable to experiment with pedagogical change, including the faculty, departments, and the institution as a whole (Boyer Commission, 1998; Massey & Wilger, 1999; National Research Council, 1999).

Lazerson et al. (2000) state, "Research remains the primary avenue to individual and institutional prestige" (p.13). Although it may seem redundant to stress that research is the primary focus at research universities, it sets the context for their argument that "the efforts to change teaching and improve learning are essentially battles over institutional values, rewards, and behaviors" (p.19).

This tension is most clearly seen with junior faculty. Senior faculty, given their experience with the university's reward system, often counsel their younger colleagues in a direction that values research over teaching and learning. "The junior faculty member who seems to give disproportionate time and attention to freshman/sophomore courses may well be counseled toward more 'productive' redirection; if interest is shown in experimental or interdisciplinary courses at the baccalaureate level, movement toward tenure or promotion may be stalled" (Boyer Commission, 1998, p. 32). Donald Kennedy (1997), President emeritus of Stanford University, also recognizes the tension for tenure-track faculty members: "Much of the innovation in teaching is being accomplished by young faculty members. It is a welcome and exciting development, but we cannot help worrying about the futures of those who make such commitments while still untenured. We cannot yet assure young academics that their departments will be as interested in their teaching as in their research" (p.96).

When given the choice, faculty often opt for teaching more material, rather than teaching less material well: quantity over quality. Content coverage is especially a concern in science courses, where one course may be the prerequisite building block for many courses to come (National Research Council, 1996). The traditional model focuses on what should be taught and not how it should be taught or how learning should be assessed (Massy & Wilger, 2000, p. 2).

When the focus does shift to how to improve teaching, even promising new teaching practices are rarely sustained. Innovations rarely survive beyond the initiator. The dissemination of the innovation is the "final critical issue" (Brown, 1992). Although directed at the K-12 audience, Brown's conclusion holds true for higher education as well: "it is not sufficient to argue that a reasonable end point is an existence proof, although this is indeed an important first step" (p. 171).

Towards an Innovation

Recognizing the many layers of constraint when working with the Human Biology faculty, members of the Stanford Learning Lab advocated an incremental approach to addressing the challenges discussed above. Fundamentally we were still operating "within the box": classes were still didactic large-lectures, material presented was still voluminous, problem sets were still predominately multiple-choice, and so on. By-and-large, the faculty did not have to change their routines. The intervention we proposed, designed, and evaluated built on the foundation of weekly problem sets. The web-based environment, called Coursework, facilitated the online weekly problem sets: students responded to multiple-choice questions and supplied rationales for their responses; students received feedback soon after they turned in their problem sets; the Coursework system provided student performance data to the teaching team; and it facilitated personalized email exchanges from the teaching team to the students.

Similar Programs

Projects similar to ours, which use online multiple-choice questions in the large lecture setting, have demonstrated motivational and learning benefits. Buchanan (1998) discussed the use of a web-based multiple-choice tool in a large introductory psychology class. Rather than being given the solutions, students were given the resources to find the correct answers for problems that they missed. The web platform enhanced accessibility, and students reported that they liked using it, and that it helped them to learn. Paull, Jacob, and Herrick (1999) discussed an automated homework system for use in electrical engineering. They compared two classes, one using traditional problem sets, the other using the online version. Their findings show that students who had their problems online performed better on the exams than students who had traditional homeworks. Thoennessen and Harrison (1996) reported on the use of a computer-assisted homework system in a large physics class. Students reported that they spent significantly more time on the assignments compared to traditional classes. They found that homework completion and performance significantly correlated with the final exam score.

Intervention and Implementation

The intervention discussed in this paper addresses several of the Boyer Commission's (1998) and the National Research Council's (1999) recommendations for reforming undergraduate education at the research university, especially in terms of using technology effectively, enhancing communication, and evaluating the effects of those interventions. The online problem set system was implemented and evaluated in the Spring 1998, and then again in the Fall 1998 – Spring 1999 academic year. The online problem sets were meant to systematically engage students in the content, because "students who are actively engaged in learning for deeper understanding are likely to learn more than students not so engaged" (Cross, 1998, p. 4). The Chair of the program in Human Biology required the eight faculty coordinators in the sequence, about half of whom are junior faculty, to use the online problem sets in their courses.

The Process

The weekly problem sets contained several multiple-choice problems, each with space for an open-ended rationale. Whereas the multiple-choice format allowed for automated grading, the rationales were meant to promote a reflective approach to learning (Anastasi & Urbina, 1997) and to model the practice of giving scientific justifications (Edgington, 1997; Eflin, 1995).

Each quarter, the online weekly problem sets were worth between 5% and 15% of the students' final grades. Usually between 10 and 15 problems were assigned each week. The timing differed somewhat from quarter to quarter, but typically lectures occurred Monday through Thursday; the problem sets were released on Thursday and were due on Sunday. Students received their scores and the correct rationales Sunday evening. Course assistants would email their students Monday through Wednesday, and then the cycle would repeat. Figure 2 illustrates the type of problem found in the weekly problem set. Figure 3 gives an example of the class-wide feedback of solution and correct rationale released after the due time. Item statistics for the course and for each individual section were then available to the faculty and course assistants. Figure 4 gives an excerpt of the "frequently missed questions." The faculty and course assistants used these student performance data when discussing their section plans for the week. The course assistants and faculty used the email feature of the system to provide students with tailored feedback regarding their responses and rationales. Figure 5 shows a personalized email sent to a student. Coursework automatically embeds the student's answer choice and rationale within the body of the email message.

Though the problem sets were automatically graded, there was no time savings for the course assistants. How they spent their time differed, presumably geared toward less rote tasks. The course assistants were integral to the entire process of the online problem sets: from constructing the items and providing individualized feedback, to incorporating the students' performance back into the curriculum.

2. *(1 pts)* **An investigator plans to use Seligman's (1975) procedures to study the effects of learned helplessness on immune function in rats. What control condition or control group is most important for an unambiguous interpretation of the results?**

- a. a group of rats given a pleasant event instead of an aversive event in the presence of the CS
- b. a group of rats given the same amount of an aversive event, but the event is under the rat's control
- c. a group of rats exposed to the CS but given neither pleasant nor aversive stimuli
- d. a group of rats given some kind of virus that impairs the immune system by a nonbehavioral means

Please enter your rationale here

3. *(1 pts)* **Carefully controlled laboratory studies by learning theorists had an enormous impact on ideas about childrearing. These ideas are no longer so popular, in part because we now understand that the experiments of Skinner and others in the behaviorist tradition:**

- a. lacked reliability.
- b. were biased by nativist views.
- c. lacked internal validity.
- d. lacked external validity.
- e. would not be replicable using more modern measurement techniques.

Please enter your rationale here

Figure 2. Excerpt from a problem set that students see before the due date/time.

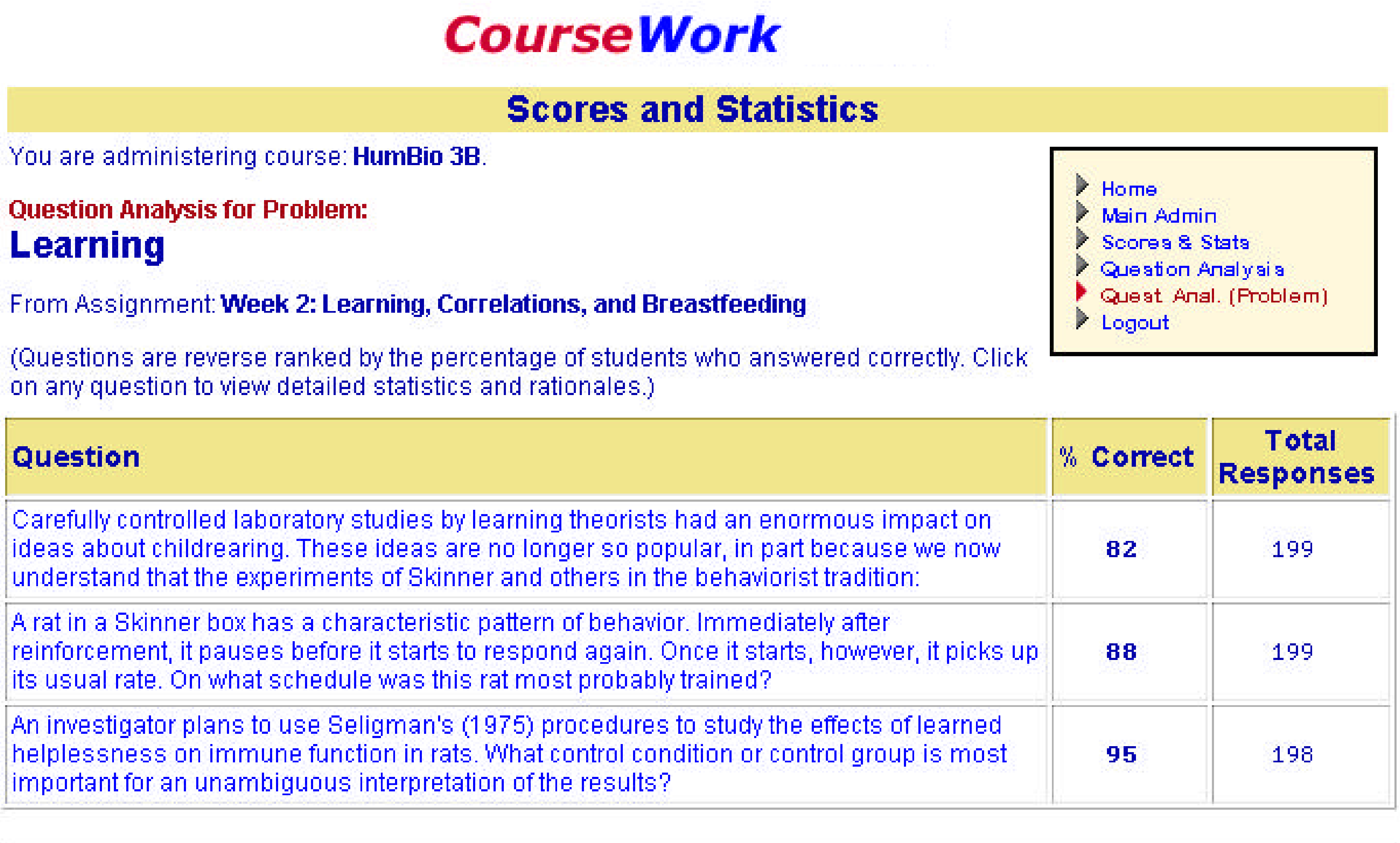
2. *(1 pts)* **An investigator plans to use Seligman's (1975) procedures to study the effects of learned helplessness on immune function in rats. What control condition or control group is most important for an unambiguous interpretation of the results?**

- ✗ a. a group of rats given a pleasant event instead of an aversive event in the presence of the CS
- ✓ b. a group of rats given the same amount of an aversive event, but the event is under the rat's control
- ✗ c. a group of rats exposed to the CS but given neither pleasant nor aversive stimuli
- ✗ d. a group of rats given some kind of virus that impairs the immune system by a nonbehavioral means

The best control subjects match the experimental subjects in every detail except for the variable under investigation. In this instance, answer (B) would be the best control group because it matches the experience of the control and experimental subjects except for the fact that the control subjects have control over the shock -- which is the variable under investigation. See Gleitman page 146 for reference.

Student Answer:	c
Student Rationale:	This would be a control group while the group of "b" would be the actual test group.
Score:	*0 pts*

Figure 3. Example feedback class receives after the due date/time.



CourseWork

Scores and Statistics

You are administering course: **HumBio 3B**.

Question Analysis for Problem:

Learning

From Assignment: **Week 2: Learning, Correlations, and Breastfeeding**

(Questions are reverse ranked by the percentage of students who answered correctly. Click on any question to view detailed statistics and rationales.)

- Home
- Main Admin
- Scores & Stats
- Question Analysis
- Quest. Anal. (Problem)
- Logout

Question	% Correct	Total Responses
Carefully controlled laboratory studies by learning theorists had an enormous impact on ideas about childrearing. These ideas are no longer so popular, in part because we now understand that the experiments of Skinner and others in the behaviorist tradition:	82	199
A rat in a Skinner box has a characteristic pattern of behavior. Immediately after reinforcement, it pauses before it starts to respond again. Once it starts, however, it picks up its usual rate. On what schedule was this rat most probably trained?	88	199
An investigator plans to use Seligman's (1975) procedures to study the effects of learned helplessness on immune function in rats. What control condition or control group is most important for an unambiguous interpretation of the results?	95	198

Figure 4. Excerpt from the administrator's view of frequently missed questions.

From: Bob D. Seeay <bobdseeay@stanford.edu>
Date: Jan 20, 1999 2:35 PM
Subject: Re: Week 2: Learning, Correlations, and Breastfeeding
Assignment: Week 2: Learning, Correlations, and Breastfeeding
Question: An investigator plans to use Seligman's (1975) procedures to study the effects of learned helplessness on immune function in rats. What control condition or control group is most important for an unambiguous interpretation of the results?

Hi Peyto ,

> On the question, "An investigator plans to use Seligman's (1975) procedures to study the effects of learned helplessness on immune function in rats. What control condition or control group is most important for an unambiguous interpretation of the results?",
>
> you answered:
> c. a group of rats exposed to the CS but given neither pleasant nor aversive stimuli
>
> BECAUSE:
> This would be a control group while the group of "b" would be the actual test group.

I wanted to be sure you understand why answer B (a group of rats given the same amount of an aversive event, but the event is under the rat's control) is the best choice here -- this will be a point you'll want to know for the midterm. Check the rationale (if you haven't already), and let me know if you have any questions, okay?

Hope you have a good weekend,
Bob

Figure 5. Example of a personalized email to student who missed the problem.

The Rationales

In the large lecture setting, where there are numerous students and limited resources, multiple-choice problems are appealing for their ease in scoring and their apparent objectiveness. However, simple responses to multiple-choice questions may not display a student's understanding. Students may give the right answer for the wrong reason, and vice versa. Our intervention addresses this concern by having students respond, not only with what they think is the right answer, but by providing a justification or rationale explaining why they think that response is right. Shulman (1999) argues that "learning is least useful when it is private and hidden; it is most powerful when it becomes public and communal" (p. 39). The rationales were meant to make the students' thinking more visible, both to the students and to the teaching staff.

The rationales served a second purpose as well: they were meant to develop the ability to justify and argue one's position, an important part of the process of scientific inquiry. Edgington (1997) argues that explanation is the very purpose of science and tasks that require explanation are commonly used to assess students' understanding. Eflin (1995) argues the importance of teaching how to frame scientific rationales.

Evaluation

Methodologically, we were constrained in terms of our research design and the types of information that we could collect (Schaeffer, Mabogunje, et al., 1999; Schaeffer, Michalchik, et al., 1999). We did not have the opportunity to create comparison groups or to collect data from the previous year that might have served as a baseline for several reasons: this intervention occurred in the context of an ongoing sequence of courses; faculty were reluctant to parse students into comparison groups; and the intervention was designed and developed on rather short notice. We had lengthy discussions about locating other potentially comparable groups that did not have access to the online problem set environment, but those discussions quickly turned to highlighting the difficulty in establishing comparability otherwise (e.g., inherent differences between students who choose the Human Biology major as compared with the Biology major). We were also constrained in terms of measures of student learning; we did not have permission from our institutional review board to collect student grades or to use the performances on the problem sets in our research. Future iterations may seek to enact tighter research controls and to obtain more direct measures of student learning.

Hypotheses Guiding the Evaluation

The graded online problem sets served multiple goals from the faculty, course assistant, and student points of view. These multiple goals are summarized in the following three overarching hypotheses. These hypotheses guided our evaluation activities over four quarters.

Hypothesis 1. Graded online problem sets (including providing rationales) will require students to be actively involved with the material on a weekly basis. Scores, together with solutions and correct rationales, will provide students with rapid and frequent feedback. Together they will enhance students' motivation, learning, and metalearning (student's assessment of their own learning).

Hypothesis 2. Graded online problem sets will help sections to be more useful. Having done the problem sets, students will be better prepared. Course assistants will have the student performance data and rationales beforehand, allowing them to better prepare remedies for misconceptions and foster a less didactic atmosphere.

Hypothesis 3. The problem sets and rationales will serve as a common basis for discussion and may serve to enhance student/course assistant, student/faculty, and course assistant/faculty communication.

Method

Our evaluation efforts drew on multiple methods and multiple perspectives. Besides interviews with the students, faculty, and course assistants, we administered surveys at the beginning and end of each of the quarters, held focus groups, observed lectures, sections, and student study groups, and conducted think-aloud protocols observing students completing their problem sets.

Result Highlights

Overall, approximately 90% of the students reported liking that the problem sets were presented online. Students have access to the web around campus and found that they liked submitting their homeworks online, saving them the trouble of walking across campus to turn something in on time. One student commented: “Very helpful to have problem sets and solutions on the web--could basically check and do homework anywhere.”

Throughout the academic year, between 70% and 79% of the students reported working with a partner or in small groups while completing their problem sets. Typically they would review the problems individually, come together with their peers to discuss their initial solutions and revisit their thinking, and then submit their responses and rationales individually. Overall, the problem sets guided the students’ reading and exam preparation.

Evaluating Hypothesis 1: Enhancing student involvement, learning, and motivation. In addressing Hypothesis 1, the data support that the students were actively involved with the material. Whereas only 20% of the students at the beginning of Fall quarter expected that the online problem sets would be particularly helpful for them to keep up with the course, at the end of each of the subsequent quarters between 55% and 65% (depending on the quarter and A- or B- side) reported that the problem sets were “very helpful” for keeping up with the course. Overall, between 79% and 91% of the students rated the online problem sets as “very” helpful for their learning. As one student wrote on the survey, “I think that [the problem sets] are an excellent way to keep on top of the material.”

Between 91% and 97% of the students said that the multiple-choice part of the problem sets were "somewhat" or "very" helpful for their learning and between 84% and 89% said that the rationales part of the problem sets were "somewhat" or "very" helpful for their learning. Between 85% and 94% said that the solution keys were "somewhat" or "very" helpful for their learning.

As a result of having to provide rationales, approximately 75% of the students reported spending more time thinking about the problems and approximately 68%

reported that they thought about the problems more deeply and critically. As one student wrote, “Without having to give rationales, it would have been easy just to guess. The rationales made me think about the material and learn it much better.” The rationales reportedly served a self-correcting function for another student: “I often realized I was wrong on a problem after being required to think it through in order to write a rationale.”

Although there is ample evidence to suggest that the students were more involved and presumably learned more because of the problem sets, students frequently commented on the quality of the questions themselves, and how that undermined their motivation and learning. Each problem set seemed to have at least one question that was difficult to understand. As one student said, [the problem sets] “were helpful but I thought that many questions were either too ambiguous or poorly presented." Another student commented, "...I felt that the questions tended to be very ambiguous and at times I found myself spending up to a half hour trying to figure out the answer to a question simply because it was worded strangely." Students were also aware that the problems on the problem sets were not always representative of the content emphasized in other parts of the course. One student wrote "I found [the problem sets] to be not very comprehensive and often very detail-oriented and nit picky which really only stressed me out and forced me to read into the handouts and readings while not getting the larger picture.” As one student said with caution, "They are helpful as long as they reflect what will be on the final."

The course assistants felt strongly that students were more engaged with the material than they would have been without the online problem sets. They reported that the problem sets “compel students to open their books, read their notes, and talk with other students enough so that they can't ever get more than a week behind." The course assistants stressed that students "actively talking with each other" while working on problem sets "enhances their learning more than anything."

A comment by one faculty member underscored the formative nature of the feedback: “this program forces [students] to see [when] they are not doing well. And therefore they will come in earlier for help and more for office hours. And go back to the material that they didn’t get and make sure that they got it.” The faculty felt that the problem sets especially helped students who don’t tend to ask questions. “And so what the problem sets did for all those who don’t ask questions [is to give them] an opportunity to find out that they don’t know the answers. But not for those who already ask questions; those who ask questions ask them anyway…” and get feedback on their understanding whether or not they are using the problem sets.

Another faculty member said, “I think Stanford students are so over committed in so many areas that just the rigors of needing to prepare something on time and knowing that you’ll get that instant feedback on how they did [is useful]. And then knowing that it still can ideally play roles the next week [in lectures] is encouraging.”

<u>Evaluating Hypothesis 2: Improving usefulness of sections.</u> Together with the faculty, the course assistants routinely reviewed the student performance data on the problem sets when planning that week’s sections. They would focus particularly on the questions frequently missed by the entire class. Sometimes they would discover ambiguities in the wording of the question that confused students; other times they recognized material that needed to be revisited, and would do so in sections. They often

focused on the rationales students provided for clues as to the basis for their misunderstandings.

Individually, course assistants often reviewed the problem set performance of the students in their sections before their section meeting. The course assistants found this useful for preparing to cover the material students found difficult.

In addressing Hypothesis 2, between 95% and 99% of the students rated sections “somewhat” or “very” helpful during the evaluation period. One student wrote, “sections were very helpful in pulling the material together and making me see the big picture as well as better understand the details.” Some sections were able to sustain more interactive formats; course assistants would use the problem set performance as a basis for calling on students in class. According to one student, “I was able to participate more because [my course assistant] would call on people in class. Whether or not they knew the answer, they'd have to talk at least a little bit. I felt that was good because no one person dominated the discussion.”

Evaluating Hypothesis 3: Improving communication among students, faculty, and course assistants. There is mixed support that communication between students and course assistants improved. From the course assistant’s point of view, communication between the students and the course assistants improved because the problem set performance often served as the basis for discussion in the sections. The course assistants told us that the students' rationales helped them get a better feel of where students were coming from, and allowed them to provide individual email feedback to address misconceptions and praise good reasoning. The message from the students was not so clear. Class email was rated as “somewhat” or “very” helpful by approximately 80% of the students at the end of Fall quarter as compared to 50% at the beginning of Fall quarter. Yet receiving personal feedback from the course assistants was rated as “somewhat” or “very” helpful only by between 51% and 58% of the students. As one student wrote: [I’d like to see] “more feedback on rationales from course assistants. I know they are busy already and can only do so much, but this would help.” Only between 51% and 65% of the students reported that interaction with the course assistants was “somewhat” or “very” helpful for their learning. In a course where most of the student-teacher “face-time” occurs with the course assistants, one might expect this number to be higher. The course assistants told us that perhaps students' expectations were raised too high by the promise of greater interaction with the teaching team.

Faculty differed in the amount of interactions with the students. Two faculty members in particular regularly sent emails to the students. The students told us in the middle of that quarter that even if they hadn't yet been directly contacted by the faculty, they knew of peers who had, and thus felt that it was a possibility that the faculty were caring about their work. Through the rationales, faculty felt they had a better idea of what students were thinking, and were able to use those rationales to evaluate how well they were teaching. One faculty member told us that, through the email system, he was able to get to know many more students in during the quarter than is typical. Approximately a third of the students reported that interaction with this faculty member outside of class was helpful for their learning; this seems remarkable given the nature of large lecture classes.

The caveat in interpreting these results is that the faculty interaction with students in the online environment greatly differed from faculty member to faculty member. Some faculty, though able to access student performance data readily, never did. Although all of the faculty knew about the email feature embedded in the software system, most did not use it. As one faculty said, " the advantage of this system is the emailing links [but] I've never used them." There was a clear sense from all the faculty involved that the enhanced communication takes more of their time. Although all the faculty recognized the benefits of increased personalized communications with the students, only two faculty members were successful at maintaining ongoing contact with the students. As one faculty member struggling with the pressures of tenure-track responsibilities said about emailing the students: "I put it in the 'wouldn't it be nice if I could?' category."

In contrast, course assistants did regularly use the email system to communicate with students. The system helped them to keep track of how often and when they sent each of their students emails. Many of the course assistants set goals such as emailing each student at least four times during the quarter.

Discussion

We have not attempted to change the institutional constraints that surround large lecture classes in the research university setting. Rather, we have worked within traditional models to improve the learning situation incrementally. In this way we have seen rather large-scale adoption of a technology-enhanced solution.

The online problem sets afforded the students and teaching team with weekly, personalized feedback that is often overlooked in the large lecture setting. The problem sets served as an anchor for the students. It guided their reading and exam preparation. Most students worked in small groups or in partners on the problem sets, and found those interactions helpful for their learning. The course assistants used the information they obtained from the student performance on the problem sets to help determine content to be covered in the weekly sections. As one faculty member explained, "The online homework submission expresses two important rules-of-thumb in learning activity: nothing is more important for the student than rapid feedback on their work; and for the professor, nothing is more important than the rationale for why students do what they do."

In addressing Hypothesis 1—Improved engagement and learning, our data support that the online problem sets have helped to increase student engagement in the material, and presumably through this enhanced engagement and routine feedback their learning has improved. In addressing Hypothesis 2—Improved sections, it is clear that section coverage is highly responsive to the students' performance on that week's problem sets. Sections are better tailored to match students' understanding, and students find sections incredibly helpful. In addressing Hypothesis 3—Enhanced communication, we have mixed support that the problem sets have helped to enhance communication between the students and the teaching team.

Involvement with the problem set system varied greatly from faculty member to faculty member. Although it was the program chair who mandated the use of the problem sets within this program, individual faculty determined their own level of interactions. Only one faculty member fully interacted with the system and maintained regular student email feedback; at least two other faculty regularly visited the website to view the student performance data. The remaining five faculty, though recognizing the benefits of such, felt that their time was better spent focusing on other things, including course preparation and their research programs.

Course assistants were instrumental in initiating review of student performance after each problem set was due. Even if the faculty did not visit the website to view the student performance data, they were still able to benefit from the feedback on the students' performance because the course assistants would "brief" them at their weekly meetings. What is unclear, however, is how the faculty changed their lectures based on the information they received on their student's understanding.

The course assistants carried much of the responsibility for writing the questions and uploading them into the system, and providing personal email feedback to the students. An interesting unintentional outcome: the course assistants learned a lot from the problem set system too. Rich learning occurred during the course assistant meetings when they developed the problem sets. There is clear benefit to the course assistants themselves to work together to construct well-worded multiple-choice problems. Faculty oversight on constructing the problem set questions varied. Regardless if faculty were involved in the question writing process, the course assistants quickly learned not only how difficult it is to write good questions, but how frustrating it is for students when presented with ambiguous or unrepresentative questions.

In many ways, the success of the online problem set system rests on the quality of the questions asked. Because the students' level of interaction with the problem sets was more visible with this online system, the importance of good multiple-choice questions was elevated, not only in terms of being well constructed and unambiguous, but also in terms of aligning with the material covered in lectures and exams. Students should feel that spending time thinking about the problems is time well spent, and will benefit them for the exams. Questions that are not representative of the important content can create a false sense of security: students feel that if they score well on the problem sets, it is indicative that they understand the material. There is more to writing good quality, representative multiple-choice questions than most people would assume (Haladyna, 1999; Osterlind, 1998; Stiggins, 1997).

One faculty member argued that the "cost" of a bad question on the problem set is much higher than for that on an exam; students would spend hours trying to sort through a poorly worded question on a problem set. Another faculty member commented on how students' rationales were sometimes just referencing page numbers in the text. It is unclear whether this faculty member saw how asking good, probing questions will pull for nontrivial rationales that require higher level thinking, whereas questions of fact and definition lend themselves to trivial rationales. In general, faculty and course assistants tend to underestimate how time consuming and difficult it is to write good, probing multiple-choice questions that require synthesis and deeper understanding; it is much quicker and easier to write questions of fact and definition. As Bransford et al. (2000) argue, "assessments must reflect the learning goals that define various environments. If the goal is to enhance understanding, it is not sufficient to provide assessments that focus primarily on memory for facts and formulas" (p. 154). The best use of this online problem set system will come with better training on writing multiple-choice questions that are representative of important content and probe for deeper understanding.

Many of the obstacles typically associated with introducing and sustaining innovation were apparently overcome in our study. First, our approach amounts to facilitating incremental change; we were still operating “within the box.” Second, the head of the Human Biology program, himself a senior faculty member, advocated this system and mandated the faculty coordinators in that program to use it. Third, although the faculty involvement with the system varied greatly, the course assistants served as a “safety net”: they were instrumental in maintaining the feedback loop between faculty and students.

In addition to its continued use in the Human Biology program, Coursework online problem set system has been adopted by several departments across campus (e.g., French, Spanish, Mechanical Engineering, etc.). We are planning a roll-out for a full-functioning learning management environment (including course outline, electronic course material reserves, a grade book, an asynchronous discussion tool, etc.) that will include the online problem sets as a module, which will soon be available at several different campuses across the country[1].

Note that further adoption of the software tool itself is not our goal; rather, improved teaching and learning is. With thoughtful use, this tool can facilitate timely, personal feedback, which is widely recognized to enhance learning, yet is often overlooked in the large lecture setting.

Future studies should look more systematically at student benefits of such a system. More direct measures of student learning would allow more systematic inquiry as to whether student benefit is related to the extent of email feedback received; to the quality of questions asked; to the extent of faculty responsiveness to student performance; as well as whether student characteristics (e.g., how well they monitor their own understanding) mediate the potential value of such a system.

---

[1] For more information, contact Scott Stocker <stocker@stanford.edu> or Charles Kerns <kernsc@stanford.edu>.

References

American Association for Higher Education. (1996, July 25). Nine principles of good practice for assessing student learning. <http://www.aahe.org/assessment/principl.htm> (2001, January 10).

Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Upper Saddle River, N.J.: Prentice-Hall.

Angelo, T. A. (1990). Classroom assessment: Improving learning quality where it matters most. New Directions for Teaching and Learning, 42. San Francisco: Jossey-Bass.

Boyer Commission on Educating Undergraduates in the Research University (1998). Reinventing undergraduate education: A blueprint for America's research universities. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). How people learn. Brain, mind, experience, and school (expanded edition). Washington, D.C.: National Academy Press.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences, 2(2), 141-178.

Buchanan, T. (1998). Using the World Wide Web for formative assessment. Journal of Educational Technology Systems, 27(1), 71-79.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. AAHE Bulletin 39(7), 3-7.

Cross, K. P. (1998). What do we know about students' learning and how do we know it? Keynote address. AAHE National Conference on Higher Education. <http://www.aahe.org/cross_lecture.htm> (2001, February 10).

Edgington, J. R.(1997, March). What constitutes a scientific explanation? Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Oak Brook, IL.

Eflin, J. (1995). Educating for scientific literacy. Journal on Excellence in College Teaching, 6(1), 31-43.

Friedlander, L., & Kerns, C. (1998). Transforming the large lecture course. Syllabus Magazine, 11(10), 53-56.

Haladyna, T. M. (1999). Developing and validating multiple-choice test items (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum.

Kennedy, D. (1997). Academic duty. Cambridge, MA: Harvard University Press.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. Review of Educational Research, 58(1), 79-97.

Lazerson, M., Ursula, W., & Nichole, S. (2000). Teaching and learning in higher education, 1980-2000. Change, May/June 2000, 13-19.

Massy, W.F., & Wilger, A. K. (1999). Faculty productivity. Stanford University: Stanford Institute for Higher Education.

Massy, W.F., & Wilger, A. K. (2000). Note on Education Quality Work. Stanford University: National Center for Postsecondary Improvement.

National Research Council (1999). Transforming undergraduate education in science, mathematics, engineering, and technology. Washington, D.C.: National Academy Press.

National Research Council (1996). From analysis to action: Undergraduate education in science, mathematics, engineering, and technology. Report of a convocation. Washington, D.C.: National Academy Press.

Osterlind, S. J. (1998). Constructing test items : multiple-choice, constructed-response, performance, and other formats (2nd ed.). Boston: Kluwer Academic Publishers.

Paull, T., Jacob, M., & Herrick, R. (1999). Automated homework in electrical engineering technology. Purdue University. <www.clearlearning.com/support/cases/purduePaper.pdf> (2001, February 12)

Schaeffer, E., Mabogunje, A., Johnson, M., Carrizosa, K., Steinbeck, R., Kerns, C., Billings, M., Marincovich, M., Sheppard, S. (1999). Human Biology 4A and the Stanford Learning Lab (Spring 1998). Stanford University. <http://sll.stanford.edu/projects/> (2000, September 10).

Schaeffer, E., Michalchik, V., Martin, N., Birks, H., & Nash, J. (1999). Evaluation of the collaboration between the Program in Human Biology and the Stanford Learning Lab: Web-based problem sets (Fall 1998). Stanford University. Available at: <http://sll.stanford.edu/projects/> (2000, September 10).

Shulman, L. (October, 2000). "From problem to learning: The promise and perils of problem-intensive pedagogies." Keynote speech presented at "PBL 2000: Promises, breakthroughs & lessons." Birmingham, Alabama: Stamford University.

Shulman, L. (1999). Taking learning seriously. In D. DeZure (Ed.), Learning from change. Sterling, VA: Stylus Publishing.

Stiggins, R. (1997). Student-centered classroom assessment (2nd ed). Upper Saddle River, N.J.: Prentice-Hall.

Thoennessen, M., & Harrison, M. J. (1996). Computer-assisted assignments in a large physics class. Computers and Education, 27(2), 141-147.

Twigg, C. (1999). Improving learning and reducing costs: Redesigning large-enrollment courses. Troy, N.Y.: The Pew Learning and technology Program, Center for Academic Transformation. <http://www.center.rpi.edu/PewSum/mono1.html> (2001, January 10).