

DEVELOPMENT AND DEPLOYMENT OF A WEB-BASED COURSE EVALUATION SYSTEM

Trying to Satisfy the Faculty, the Students, the Administration, and the Union

Jesse M. Heines and David M. Martin Jr.

Dept. of Computer Science, University of Massachusetts Lowell, Lowell, MA 01854, USA

Keywords: User Interfaces, Security, Privacy, Anonymity, User Perceptions, Web Application Deployment

Abstract: An attempt to move from a paper-based university course evaluation system to a Web-based one ran into numerous obstacles from various angles. While development of the system was relatively smooth, deployment was anything but. Faculty had trouble with some of the system's basic concepts, and students seemed insufficiently motivated to use the system. Both faculty and students demonstrated mistrust of the system's security and anonymity. In addition, the union threatened grievances predicated on their perception that the system was in conflict with the union contract. This paper describes the system's main technical and, perhaps more important, political aspects, explains implementation decisions, relates how the system evolved over several semesters, and discusses steps that might be taken to improve the entire process.

1 OUT WITH THE OLD, IN WITH THE NEW

Course evaluations have been around for eons. Although many professors view them as nothing more than a popularity contest ("If I give a student a good grade I get a good evaluation, if I don't, I don't"), when students respond honestly and conscientiously they can provide invaluable feedback for improving the quality of instruction.

1.1 Paper-Based Evaluations

Paper-based forms are typically filled out in one of a course's final class meetings by those who happen to be present that day. Absentees rarely get another chance to complete the forms.

The collected data are not always subjected to rigorous analysis. Administrators may "look over" the forms and get a "feel" for students' responses, but if student input is not provided in machine-readable format, the general picture is hard to quantify. Many professors ignore the results due to a lack of interest and/or perceived importance or simply the inconvenience of having to go request to see the forms from a department administrator and then wade through raw data rather than machine-generated summaries.

The real tragedy is that while students' free-form responses have the greatest potential to provide real insight into the classroom experience, when provided on paper-based forms these responses are sometimes illegible or so poorly written that trying to make sense of them is exceedingly difficult. All too often these valuable comments therefore have little impact on professors' teaching.

In our institution, as in many others (see Hernández *et al.* (2004) as an example), paper-based evaluations suffered from all of these problems.

1.2 Web-Based Evaluations

Web-based forms have the potential to address many of the shortcomings of their paper-based counterparts. They can be filled out by students anytime and anywhere. Responses can be easily analyzed, summarized, and displayed in tables or graphs. Student responses to free-form questions will at least be legible, even if they remain unintelligible. Professors and administrators alike can do far more than simply "look them over." Most importantly, online data collection and reporting ensures that students opinions can at least be heard, which is the first step in having professors listen to those opinions and use them to positively impact the quality of instruction.

2 SYSTEM EVOLUTION

2.1 Single Professor Version

Our first online course evaluation system was deployed at the end of the Fall 2001 semester in two computer science courses taught by a single professor. The questions in this system were hard-coded, response data was stored in XML, and reports were generated using XSL. The system had no login procedure. Students were simply trusted to complete only a single evaluation form for each class. In addition, students were allowed to see summaries of the response results. Given the trustworthiness of the students and the fact that the professor did not mind students seeing the response results, more formal constraints were simply not necessary.

2.2 Department Version

In the Fall 2003 semester, the Dept. of Computer Science began planning for its reaccreditation and was in need of concrete course evaluation data. The department voted to adopt the system that had now been used for four semesters, provided that a number of improvements were made.

Most of the hard-coded data had to be moved to data files so that the system could accommodate multiple professors and far more courses. Alternate question types were needed to satisfy the desire for richer course-specific questions. A faculty login system was added to protect individual professors' response results, but student logins were still not required. This led to our first observance of angry students "stuffing the ballot box," a shortcoming that clearly needed to be addressed. In addition, the time needed to support the system became a major factor. This was to become a much larger issue as the system user base expanded in subsequent semesters.

Despite these issues and the fact that not as many students used the system as we had hoped (345 evaluation forms were submitted for 28 courses), for the first time individual professors were able to see their results virtually instantaneously (see Figure 1).

2.3 College Version

In the Spring 2004 semester, our dean asked us to expand the system for use by the entire College of Arts & Sciences. The larger scope of this task made it impractical to continue using XML as the data store, so we decided to rewrite the entire system to use MySQL. This was, of course, a major undertaking. We tackled the problem of "stuffing the

ballot box" by introducing a "one ticket, one evaluation" scheme (described below).

With the system now being used outside our own department, special care had to be taken to ensure that its functionality conformed to all provisions in the faculty union contract dealing with course evaluations. This was problematic, to say the least, due to the rigid interpretation of the contract's wording by the union's executive board.

2.3.1 One Ticket, One Evaluation

We designed a system of "survey keys" that could each be used to submit a course evaluation. Our requirements were:

- (1) Each survey key should be uniquely associated with one section of one course in one semester.
- (2) Each survey key should provide the ability to submit one and only one evaluation for its associated course.
- (3) Each student should receive one survey key for each course in which he or she was enrolled that semester.
- (4) Submitted evaluations should not be traceable to student identities in the associated section, course, and semester.
- (5) It should be infeasible to forge a valid survey key.

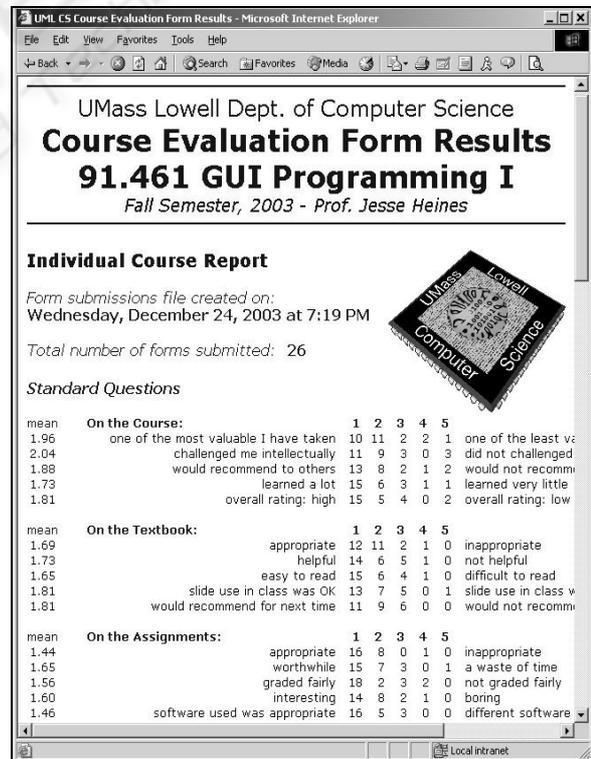


Figure 1: First Individual Course Report

Requirements (1)-(3) ensure that students could not evaluate courses they were not enrolled in, or evaluate a single course multiple times. Requirement (4) ensures that each student's evaluation is anonymous, and (5) ensures that only survey keys issued through the official process can be used to submit an evaluation.

A survey key in our system is an encrypted "ballot," which is a 64-bit entity consisting of:

- a 4-bit revision field (currently 0)
- a 24-bit course ID, allowing for 16,777,216 unique course/section/semester descriptors
- a 16-bit slot ID, allowing for classes of up to 65,535 students (with increasing forgery risks as the class size grows; see below)
- a 20-bit padding field of zeros

The course and slot IDs are both allocated sequentially. A separate table translates the course ID into the actual course, section, and semester as shown in the course catalog.

Using this encoding, it was straightforward to meet requirements (1)-(3). Each course/section/semester is encoded as a single course ID. The system generates one slot ID per enrolled student. To prevent multiple uses of a ballot, the system database keeps track of those 64-bit ballots that have been used to submit an evaluation and prevents their reuse. Toward the end of the term, we distributed the encrypted ballots (survey keys) to professors, who in turn could give them to students to submit evaluations. A card containing a survey key is illustrated in Figure 2.

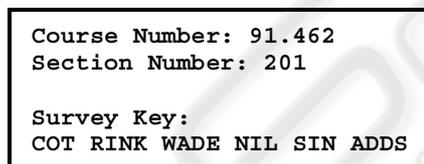


Figure 2: Student Survey Key Card

Requirement (4), achieving anonymity, was also straightforward. Since each survey key enabled a student to submit only one evaluation, they could be distributed to the students arbitrarily. We envisioned simply throwing the key cards in a hat and passing it around. To emphasize that the survey keys were not tied to individual students, we encouraged students to trade their card with any other student in their same course and section.

As mentioned, survey keys are encrypted ballots. Specifically, given a ballot B , its corresponding survey key is $S = \text{Words}(3DES_K(B))$. The $\text{Words}()$ function is a one-to-one mapping between 64 bits and a series of six English words; this mapping is specified in RFC 2289 as part of the S/Key one time password system. The 3DES key K is randomly

chosen and specific to the course ID associated with that survey key. K is stored in the system database, and no human ever needs to use it directly.

To evaluate a course, the student enters both the course and section number, along with the six-word survey key. The Web site backend uses the course and section number to locate the appropriate 3DES key and then decrypts the survey key to determine the 64-bit ballot. If the survey key does not decrypt to a legitimate ballot, the system rejects it.

Using this scheme, the probability of guessing a valid survey key is quite small, as was desired for Requirement (5). A forger must try to present a course number, a section number, and a survey key to the Web site. If the course has $128 = 2^7$ students, then there are only 2^7 valid survey keys for that course. Since the forger does not know the 3DES key, the probability that the forger is successful is $2^{-(64-7)} = 2^{-57}$. We felt that this provided adequate security for our needs, especially because each attempt would have to be an HTTP transaction to our single evaluation server, and this limits the achievable rate of brute force key guessing attacks. However, this same calculation shows that larger class sizes make brute force forgery attacks easier; for this reason we do not recommend using this system for classes consisting of more than several hundred students.

We did not intentionally record any IP addresses associated with evaluation submissions, but neither did we use end-to-end encryption or any network anonymization techniques to further hide evaluators' identities or prevent the hijacking of survey keys.

2.3.2 Deploying the Survey Key System

Putting this scheme into practice was not without its problems. 22,316 survey key cards for 1,237 courses were generated and printed up to 10 per sheet organized by course. The sheets were perforated so that the cards could quickly and easily be broken apart and distributed to students during class.

A meeting was called of all department heads in the college to explain the system and distribute these 3,084 sheets. Only about half of them attended. Department heads were then supposed to distribute the sheets of survey key cards to individual professors during a faculty meeting in which they also passed on information about their use. Only two or three held such meetings.

About half of the department heads didn't distribute the sheets at all for one reason or another, some out of protest against the system. Most of those who did distribute them did so without any accompanying instructions. Thus, some professors had no idea what to do with them.

Those professors who did attempt to use the system had numerous problems. Many didn't notice that their username and password appeared at the top of the first sheet of student survey keys for each class and contacted us to ask how to get onto the system. Others were confused because their username and password were different from those on the university e-mail system. Still others thought the online system was in some way tied to the paper-based evaluations and asked when the results from the paper-based forms would be visible online. Clearly, reliable information about the system did not get out to professors as intended. As a result, system support became a very time-consuming task, with the system developers exchanging about 137 e-mail messages (total received and sent) with 20 faculty members (about 10% of the college).

We have no way of knowing exactly how many survey key cards were distributed to students, but we do know that only 1,458 evaluations were submitted for 246 courses in 15 (out of 37) departments. This was our first inkling that the system's deployment would face far more formidable political obstacles than technical ones. It was not to be our last.

2.3.3 Question Database and Union Objections

The system's question database was designed to accommodate a hierarchy of five potential question sources: university, campus, college, department, and course. The deans of our two college divisions drafted questions for their respective divisions to be asked on all evaluation forms and sent them to department heads for approval. With a few minor changes, these questions were used to populate the college-level question tables. No questions were placed in the university- or campus-level tables. Department heads could request us to add questions to the department-level table, and a few did. Individual professors were free to add their own questions to the course-level table using a custom editor built specifically for that purpose (see Figure 3). Some did, but others sent us questions to add for them.

Despite this process, the faculty union threatened a grievance against one of the deans. The union contract states: "Individual faculty members in conjunction with the Chairs/Heads and/or the personnel committees of academic departments will develop evaluation instruments which satisfy standards of reliability and validity." Members of the executive board felt that the fact that the dean had originated some of the questions constituted a breach of contract, even though he had sent them to department heads for approval and revised them based on their input. Once again, political issues seemed to undermine our efforts to improve the effectiveness of our

course evaluation procedures.

2.4 University Version

Despite these problems, in the Fall 2004 semester the Provost asked us to enhance the system once again for use by the entire university. The MySQL database design created for the College of Arts & Sciences proved robust enough to handle the entire university, which comprises:

- 82 departments (including "areas")
- 965 faculty (including adjuncts)
- 2,357 courses
- 9,480 students
- 35,564 course registrations

However, given the disappointing experience we had with student survey key card distribution in the previous semester, we decided to try to devise a system that took department heads and faculty out of the loop and delivered survey keys directly to students.

2.4.1 Revised Survey Key Distribution Scheme

The Registrar provided us with student ID numbers and the courses for which each of those students was registered. Our first idea was to set up stations where students could swipe their ID cards and a program would look up their schedules, generate 6-word survey keys for each of their courses, and either print these or e-mail them to the student. This idea was not implemented for several reasons:

- discussions with students revealed that they did not believe that such a system would truly be anonymous
- the university has a high percentage of commuting students who do not tend to congregate in any particular central location, so many stations would be needed
- staffing the stations would be problematic
- student input also revealed that most simply "wouldn't bother" to go to such stations to retrieve their survey keys

The last point above may have been prophetic. The system that we did implement (described below) resulted in an even lower response than the one that used survey key cards, even though the number of students involved was doubled. Potential reasons for this low response rate are discussed in the concluding section of this paper, and we are currently interviewing students to try to understand why it was so, but the fact remains that only 1,230 evaluations were submitted for 471 courses in 55 departments, 16% less than the 1,458 submitted for 246 courses the previous semester.

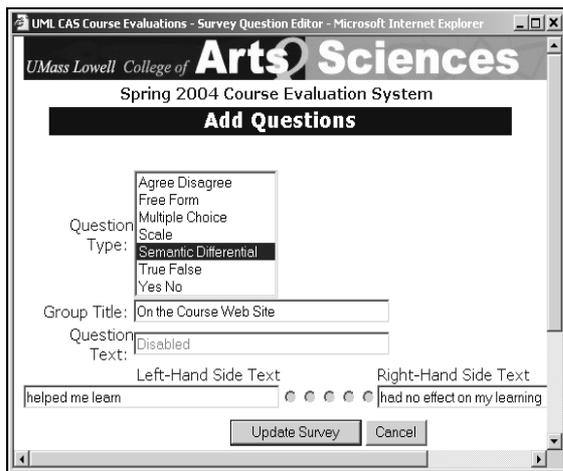


Figure 3: Custom Question Editor

The scheme we implemented involved a two-tiered system. The first page presented students with two links, one to *retrieve* their course survey keys and another to *access and submit* an anonymous course survey (see Figure 4). To retrieve their keys, students entered their Student ID and PIN (Figure 5) and were then shown the survey keys for each course in which they were enrolled (Figure 6). They then returned to the application's initial page (Figure 4) and selected the second student option to get to the course login form.

As mentioned above, only 1,230 evaluations were submitted using this scheme. This number is particularly disappointing considering that 3,388 survey keys were retrieved. Thus, only 36% of the retrieved keys were used. It is clear that the reason for this low response rate needs to be investigated.

2.4.2 Changes to the Question Database

To address the union's concerns about deans being involved in the question selection process, the common questions that we had at the college level in the previous semester's database were cloned into each of the department-level tables. The survey question editor shown previously in Figure 3 was revised so that it could be used to edit department-level questions as well.

The database change and department-level question editor allowed department heads to modify the standard questions to make them more appropriate to their individual departments, or even to delete questions they felt were not applicable to their departments. Only a few department heads did this, but at least these changes satisfied the union's objections to the deans' involvement.

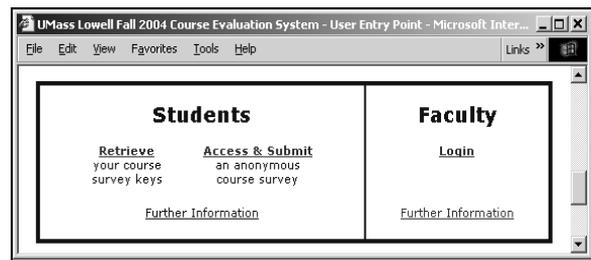


Figure 4: First Page Choice Point

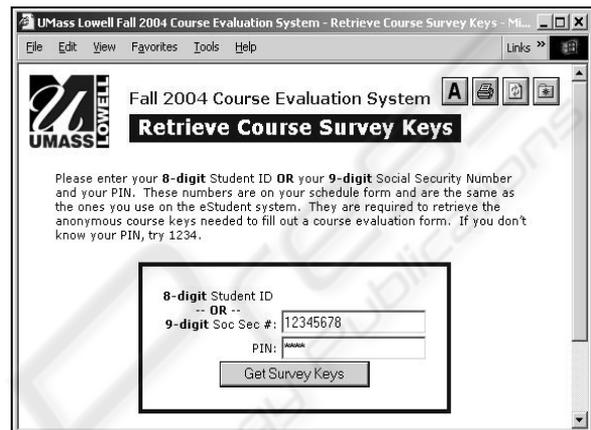


Figure 5: Student ID and PIN Entry Page

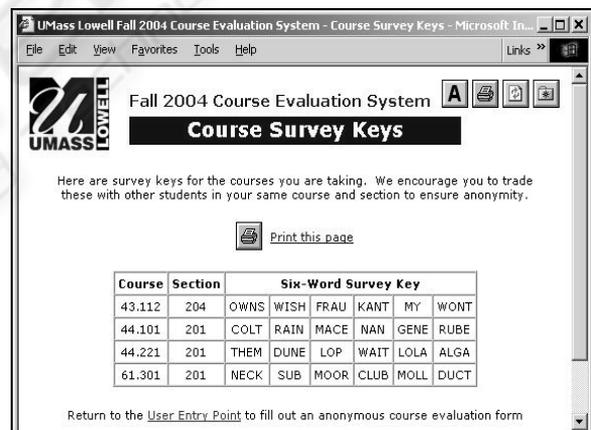


Figure 6: Retrieved Course Survey Keys

2.4.3 More Union Objections

The faculty union contract requires tenured faculty members to be evaluated "in a single section of one course per semester." The faculty menu (Figure 7) therefore allowed faculty to "turn off" student access to those courses that they didn't wish to have evaluated. All one had to do was click the  button next to a course to toggle it to . The problem, according to the union, was that "all the buttons [were initially] in the ON position." This, they felt, again

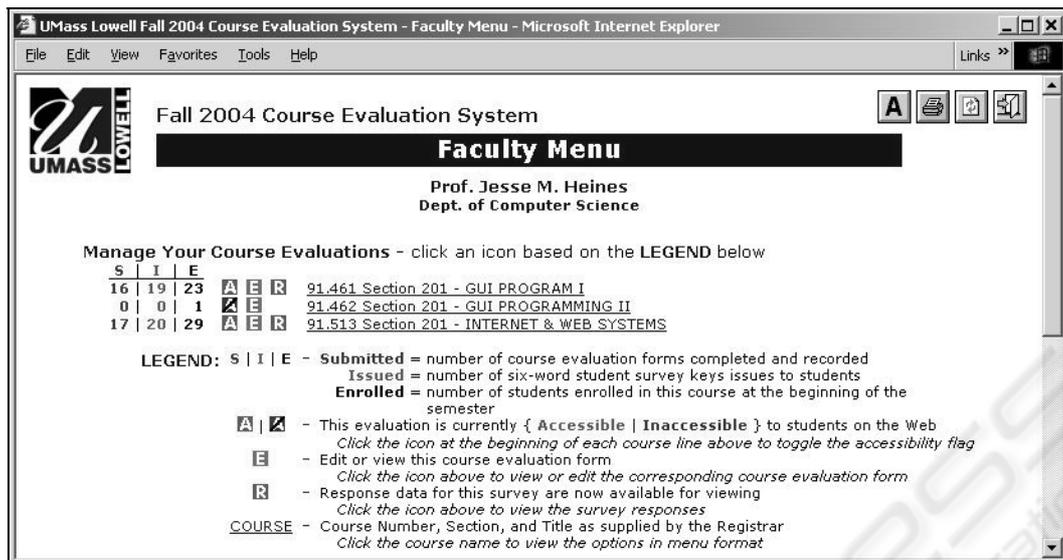


Figure 7: Faculty Menu with Web Accessibility Buttons

constituted a breach of contract, but they recognized that administratively switching all of the buttons to the OFF position half-way through the evaluation period “would create mass confusion among both faculty and students.”

To resolve this problem, the Provost and union agreed on two points. First, the Provost sent a second announcement to all faculty explicitly reminding them that the use of the online system this semester was completely voluntary. Second, the Provost and union agreed that all Web accessibility “buttons” would be set to the OFF position for future semesters, so that faculty would need to opt-in to use the system going forward.

The union was also concerned with security of the database. They asked, “Is there a zero possibility of a breach in computer security? Is the potential for a breach ... greater or less than [that] of evaluations manually compiled mainly by graduate students?” We assured the union that we had taken steps to protect the system and its database from being compromised, but of course we could not guarantee that the system could not be broken into. We agreed to look into this matter further and to add SSL/TLS protection for the next semester.

Due to these concerns, the Provost and union agreed that all student response data should be deleted from the database three weeks after the beginning of the next semester, which would be February 15, 2005. To accommodate those faculty who wish their data to be kept (for example, for including in their applications for promotion and tenure), we have put in place a mechanism by which those who could so inform us, thus preventing their data from being deleted.

As one can imagine, the enlarged user base and the additional union objections demanded that even more time be spent on system support. This semester the system’s developers exchanged about 368 e-mail messages (total received and sent) with 84 different faculty members (about 18% of the full time faculty). In addition, we exchanged another 29 e-mail messages with 10 different students.

3 DISCUSSION AND FUTURE CONSIDERATIONS

3.1 Response Rates

The low response rate we experienced is not unique. Stephen Thorpe (2002) reported only slightly higher response rates at Drexel University with paper- as well as Web-based systems, although he noted significantly higher response rates for courses in students’ majors. Furthermore, Thorpe compared paper- to Web-based evaluations and concluded that “no pattern was evident to suggest that the Web-based course evaluation process would generate substantially different course evaluation results from the in-class, paper-based method.” Robinson *et al.* (2004) reported response rates of 38%, 44%, and 31% over three semesters even though students were sent multiple reminders via e-mail.

Mount Royal College experimented with an “online teaching assessment system” that was available to students throughout the term. Ravelli (2000) reported that “fewer than 35% of students completed the online assessment at least once during the term”

and addressed this issue in follow-up student focus groups. He discovered that “students expressed the belief that if they were content with their teachers’ performance, there was no reason to complete the survey. Therefore, the lack of student participation may be an indication that the teacher was doing a good job, not the reverse.”

Given the existence of such public Web sites as RateMyProfessors.com, we had thought that the simple existence of our system would trigger a *Field of Dreams* reaction among students: if we built it, they would come. Obviously, we were wrong.

The reasons why we were wrong are not clear at this time. A presentation on the system was made to the Student Government Association, an article was printed in the campus newspaper, and detailed directions were posted on the first page of the application’s Web site. (Notice the Further Information links in the First Page Choice Point shown previously in Figure 4.) At least some professors strongly encouraged their students to use the system and were deeply disappointed that they didn’t, as demonstrated by the e-mail shown in Figure 8.

I checked my online evaluations and discovered that no student filled one out. Can you check this out and confirm [it]? If true, I am surprised and dismayed, since I devoted 15 minutes of one class encouraging students to fill them out -- not just for me -- but for all of their classes.

Two possibilities exist:

- 1) You toggled [Web accessibility to OFF on December 22, the last day of the final exam period] per the Provost's memo before my students had a chance to use the system. (I toggled it back [ON] just yesterday [January 2].)
- 2) Students are students.

[I think the reason is] more likely to be the second possibility. Based upon this data (0 for 36 students), I am inclined to use a paper version distributed in class [in] the future.

Figure 8: Faculty E-Mail re Response Rates

There has been discussion about offering some sort of lottery prize that students would automatically become eligible for if they filled out at least one course evaluations form, but we haven’t been able to figure out how to make that possible without compromising the system’s anonymity.

We suspect some combination of several reasons why the system was not used more by students.

- Students may not trust the system’s anonymity and fear that professors may “get back at them” if they submit less than flattering evaluations. We certainly stressed the system’s anonymity and the fact that professors can’t see the results until after the deadline to turn in final grades. Still, when

we replaced the “pick a card from the hat” system with the “copy the key from one Web screen to another” system, we removed the physical manifestation of the system’s anonymity and ultimately asked the students to trust our assertion that the system kept their identities secret.

- Despite our efforts to “get the word out,” many students still may not have known that the system existed.
- The process of retrieving course survey keys and then logging in to fill out an evaluation for each course may be more confusing or difficult to students than we imagine.
- As implied by our colleague in his e-mail shown in Figure 8, students may simply be too lazy or apathetic to bother with the system.
- The end of the semester and final examinations are traumatic to many students, and evaluating a course honestly and usefully can take real effort. When students are asked to add course evaluations to their list of duties at this time, it may be inevitable that doing so ends up with a lower priority than studying, relaxing, worrying, and recuperating.

We intend to sponsor focus groups or at least talk to groups of students early next semester to try to understand why the system wasn’t used more extensively. We consider this a critical task, as it is clear that we need to increase student response rates significantly to make our online course evaluation system successful.

3.2 Faculty Information Distribution

It was rather disheartening to realize that numerous faculty don’t read official university announcements sent to their university e-mail accounts.

A memo from the Provost and system developers was sent to all faculty e-mail accounts on December 5, 2004. This message contained general information about the system’s use and specific information providing the faculty member’s username and password. Further information was posted on the site via a link that was accessible without logging in (see the First Page Choice Point shown previously in Figure 4). Many faculty claimed to have never received this e-mail for a variety of reasons such as those illustrated in the two e-mail excerpts in Figure 9.

These problems were exacerbated by the fact that the system was not officially announced until the final two weeks of the semester. The reason for the delay was that the Provost wanted the Dean’s Council to approve the system before announcing it to the general faculty, and that Council only met at the beginning of each month. Therefore, we weren’t able to get approval until early December.

Thank you for your prompt reply and information. I couldn't find your [December 5th] e-mail until I searched my Netscape "junk" folder -- I forget that Netscape makes its own decisions on what is and isn't junk.

I guess I probably cause a lot of problems by not using my [university] account. I use only the account I have with my ISP. ... I used to have a number of accounts, but it got tedious keeping track of them so I give out the above e-mail address on my syllabus and use it with my online classes. ... I like to avoid all the "memos" that come around on the [university] account, although clearly I have missed something important in your communication.

Figure 9: Excerpts from Two Faculty E-Mails re Account Use

These problems were exacerbated by the fact that the system was not officially announced until the final two weeks of the semester. The reason for the delay was that the Provost wanted the Dean's Council to approve the system before announcing it to the general faculty, and that Council only met at the beginning of each month. Therefore, we weren't able to get approval until early December.

It is clear that a greater effort must be made to get information to faculty in a more timely manner. We discussed the possibility of making presentations at department or college faculty meetings, and perhaps we will be invited to do so next semester. Whatever the approach, it is clear that faculty have to better understand how the system works before it will be used successfully.

3.3 Union Support

Despite all the concerns raised by the union, we were heartened that their communications always expressed respect for our efforts and never accused us of trying to thwart collective bargaining agreements. On the contrary, one communication specifically recognized that we, too, are union members and stated that we could "be trusted in terms of ... [programming] skill, intentions, and good will."

That communication also acknowledged that "if the system works well, it will save resources in terms of paper, printing/copying, man-hours summarizing the data, and classroom time. These are all real values for the campus and the membership." It went on to say, "The system ... provides for questions related to the particular department and to the particular course. It also provides for student comments. These are real positive differences when compared to the present in-class system." It con-

cluded that "there are enough valuable aspects of the new system that could be of great value to the campus, if it operates properly and is properly secured, to consider it seriously."

Thus, despite all the rhetoric about what the system does and does not do, we are strongly encouraged by the administration's desire to see an online course evaluation system in place and the union's recognition that such a system "could be of great value." Hopefully we can address all the concerns of both the union and the administration so that the system can achieve its full potential.

ACKNOWLEDGMENTS

A large amount of the backend database programming on this project was also done by UMass Lowell graduate students Shilpa Hegde and Saurabh Singh. In addition, Shilpa wrote a fair amount of the code that generated summary reports and Saurabh wrote the custom question editor.

REFERENCES

- Haller, N., Metz, C., Nesser, P. and Straw, M., 1998. *A One-Time Password System*. RFC 2289, available from <http://www.ietf.org/rfc/rfc2289.txt>.
- Hernández, L.O., Wetherby, K. and Pegah, M., 2004. *Dancing with the Devil: Faculty Assessment Process Transformed with Web Technology*. Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services, Baltimore, MD, pp. 60-64.
- Ravelli, B., 2000. *Anonymous Online Teaching Assessments: Preliminary Findings*. Annual National Conference of the American Association for Higher Education, Charlotte, NC. ERIC Document No. ED445069.
- Robinson, P., White, J., and Denman, D.W., 2004. *Course Evaluations Online: Putting a Structure into Place*. Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services, Baltimore, MD. pp. 52-55.
- Thorpe, S.W., 2002. *Online Student Evaluation of Instruction: An Investigation of Non-Response Bias*. 42nd Annual Forum of the Association for Institutional Research, Toronto, Ontario, Canada. ERIC Document No. ED472469.