AN ONLINE SYSTEM FOR EVALUATION OF CLASSROOM INSTRUCTION: DESIGN AND IMPLEMENTATION

by
Kevin J. Morgan

A Master's paper submitted to the faculty of the
School of Information and Library Science of the
University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of
Master of Science in Information Science

Chapel Hill, North Carolina

November 2002

Approved by:

_____
Advisor

Kevin J. Morgan, An Online System for Evaluation of Classroom Instruction: Design and Implementation. A Master's project for the M.S. in I.S. degree. November, 2002. 39 pages. Advisor: Paul Solomon.

While course content delivered via the World Wide Web is increasingly routine at major universities in the United States, barriers still exist for the online course and teaching evaluations. This paper describes the development and pilot of an online evaluation system designed to replace a pencil-and-paper evaluation system. Similar efforts described in the literature are reviewed. The creation and implementation of a database-backed Web evaluation application is specified. Post-pilot comparisons between old and new systems are discussed making use of several measures, including a post-participation survey of students involved in the pilot.

Headings:

Student Surveys – Test Format – Student Evaluation of Teacher Performance

Student Attitudes

Design/Information Systems

Table of Contents

Introduction

In the spring of 2002 the School of Nursing at UNC Chapel Hill convened a committee to consider moving the School's course and teaching paper-and-pencil evaluation instrument online. As Cummings and Ballantyne (1999) point out, the evaluation of instruction goes back to the origins of organized instruction when teachers in the Middle Ages, directly dependent on their students for their wages, often would simply not be paid by their students if the student deemed their efforts inadequate. Although derided by their harshest critics as no less draconian, written student evaluations have become a nearly universally employed tool in higher education over the course of the last half of the last century. A 2000 study of the 200 most wired campuses found that 94.3% had some from of course and teaching evaluation in place (Hieleski, 2000). As at many schools, evaluations at the School of Nursing serve a dual purpose: improving classroom instruction and providing school administrators with a measurable criteria in their decisions regarding promotion and tenure of faculty.

Description of Current System

The current system has been in place since the mid-90s when the two separate evaluation tools -- one to evaluate individual teaching performance, and the other, to evaluate overall course quality -- were developed by an earlier faculty group. Since the initial work of the committee considering the feasibility of an online evaluation system centered on duplication of this online system, it bears taking a look at that system at the outset before

moving on to considering the literature concerning online evaluations system or a description of the development and implementation of an online evaluation system prototype.

The current evaluation system has three types of evaluations: course, teaching and clinical group evaluations. Course evaluations use the tool referred to above geared to evaluating overall course quality. It consists of eight fixed response Likert scaled questions and 2 open ended questions. The items are printed on a single sheet of paper. Students are expected to "bubble in" answers to the Likert questions on a computer scanable (i.e. Scantron) form that is passed out in conjunction with the sheet of questions. Students handwrite responses to the final two open ended questions on the question sheet itself. Each course offered is required to give a course evaluation.

Teaching and clinical evaluations use the second tool referred to above, which consists of 13 fixed response items followed -- as in the course evaluations -- by two free response questions. More than half of courses in the curriculum use a team teaching approach. Generally this works as follows. Each course has a faculty member who serves as course coordinator. The coordinator manages the course and handles the majority of the lecturing duties. But the coordinator also enlists other faculty for lectures, covered by the course, on which they have expert knowledge. Teaching evaluations are then administered -- based on the discretion of the coordinator, usually in consultation with the various members of the course's faculty team -- for each faculty member that has made a contribution to the course.

Additionally, many of the core courses in the undergraduate program degree tracks have a clinical component where students, divided into sub-groups, are assigned to an actual functioning health care unit (e.g. a hospital wing, an area health clinic). Here they are able to apply knowledge gained in lectures and course assignments with actual patients. These clinical groups, as they are referred to, are monitored and mentored by a faculty member, usually referred to as the clinical faculty member. In ordered to monitor how these faculty members are handling their duties -- usually carried out outside of the school and thus outside the coordinator's purview -- students are asked to evaluate their clinical faculty member using the teaching evaluation tool.

The evaluation load is, thus, potentially very heavy on students. Pupils in a core class with a clinical component could be asked to complete anywhere from 5 to 10 evaluations for that course; for instance: the course evaluation, a teaching evaluation of the course coordinator, multiple teaching evaluations for the guest lectures (which may number as many as 5 additional faculty) and an evaluation of their clinical faculty (of which there maybe up to two depending on the number of "clinical rotations" a course requires). At the very least students are asked to fill out two evaluations: one course evaluation and one teaching evaluation for their course coordinator/faculty.

As stated the current system is a pencil and paper system. Students receive a printed page of evaluation questions and an associated Scantron "bubble" sheet. Students “bubble in” answers to the Likert-scaled questions and write out their answers to the final two free answer questions. A school support unit (CITES) is charged with administering the evaluations. Through a combination of course coordinator requests and listings of that semester's course offerings and enrollment data, the office prepares evaluation packets

containing a sufficient number of question sheets and the same number of the requisite bubble sheets. Each sheet has a printed code -- generated by an internal office template -- that identifies each type of evaluation for each type of class. The code is also 'bubbled' by hand on to each included answer sheet by an office staffer. (Earlier, students had been asked to "bubble" the codes but the policy was changed because of large number of errors that rendered response sheets useless.)

Evaluations are generally distributed at semester's end and administered during class time at the final class meeting -- although in some instances evaluations have been distributed during the semester for faculty who've completed their teaching contribution well before the end of the semester or for clinical rotations completing early in the semester. Faculty are asked to give students instruction on how the evaluations are to be filled out and why filling out evaluations is important for the school's future advancement. Once they've completed this task faculty leave the room and a previously designated student collects the evaluations once they've been completed and then returns them to the support office.

Results of the evaluations are distributed once the scantron data -- scanning of the sheets having been out sourced to a university wide office that tabulates such data -- has returned. The report sheets containing the aggregated data for all the Likert-items is collated with copies of all the question sheets that were returned with written comments for the final two questions.

Distribution of results is tiered. A table in the appendix details the process, but in general faculty members see the evaluations dealing with their own teaching work. Course coordinators see those teaching evaluations and the general overall course evaluations for the

courses for which they served as coordinator. Chairs of the school's two academic divisions receive all the evaluations, course or teaching, for faculty and courses under their governance. Finally the Associate Dean for Academic Affairs is delivered copies of all the evaluations.

In summary it is a rather complex and paper heavy system. Points of failure are many. Coordinators can fail to inform the support office of courses or instructors that require evaluations. Evaluations answer sheets can be miscoded by the staff. Evaluations end up unreturned or undistributed. And the coping and collating required to produce result documents is time intensive and tedious work. Space, is also a growing concern, because all of the results must be archived by the support office for the five previous years. For all of these reasons the School was interested in considering the possibility of replacing the current system with an online one.

Literature Review

Perceived benefits of online surveys

The benefits of using the World Wide Web for survey data collection are not only widely touted but readily evident to anyone with a basic functional knowledge of the medium. Distribution does not require the use of paper nor do geographical distances pose the same, if any, sort of barrier. Processing of results moves from the human realm to the machine realm making them available much more quickly, even instantaneously. A major component of HTML is the form object which contains representations of all standard response mechanisms -- fill-in-the-blank, free text, item checklists and multiple choice -- used on printed forms. With the integration of JavaScript into a page, it's possible to perform fairly complex data validation. Even without scripted checks a web-based evaluation

system ensures a "cleanness" of data far exceeding the paper alternative. Idiosyncrasies of handwriting are eliminated. The “radio button” form element can be implemented to prevent accidental multiple selection for items where only a single selection is desired (e.g. multiple choice questions). HTML “selection boxes” offer a similar functionality. Finally, the growing accommodation of multi-media file formats, either through browser plug-ins or "hard coded" functionality, coupled with ever expanding bandwidth, make possible the integration of audio, video and interactive elements into the online survey process.

Studies looked at for this paper involved trading a paper survey system for an online system in an academic environment. But perceived benefits of online surveys cited were always similar, if not exactly the same, as the perceived benefits generally associated with general survey administration. Reduced costs for materials, distribution and administration were frequently cited as major inducements for looking to an online system. Online evaluations free up time for instruction that may have previously been allocated to the evaluation process. (Cummings and Ballantyne, 1999) Allowing students flexibility to complete evaluations on their own schedule, when they want to, also contributes to potentially richer, more considered responses. Students have been shown to write longer responses, as well. (Cummings, 1999) In instances where “bubble” sheets are used to record responses the potential for students to loose their place and “double bubble” items, rendering the responses unusable, are completely eradicated. As is the problem of indecipherable open ended responses because of poor handwriting. Also, there is no chance that a respondent can be identified because of their handwriting.

Finally, while not directly a potential dividend of an online evaluation system, the trend of higher education toward use of the Internet as a standard course tool would seem to

be driving evaluations online regardless of the benefits that might be perceived. This trend is also likely to be advanced as courses or parts of courses are delivered in web-based formats. At the School of Nursing more than 80% of classes make use of Blackboard courseware to maintain course syllabi, transmit assignments, and even administer tests. Completing evaluations online seems a logical step. Indeed, this is a necessary step when it comes to courses and programs offered over the Internet in a distance fashion, with students rarely -- if ever -- attending classes in person.

Barriers to effective online surveys

However, a survey of the literature also reveals an abundant number of concerns in administering evaluations online; some, in fact, potentially significant enough to undermine any perceived benefits. Essentially these break down into three basic questions. Are the people that respond to a web survey representatively the same as those who might respond via another method? Secondly, among those who respond, are the answers given the same answers that they might give where they to respond via a different mode? Finally, and this is the concern that seems to be the most prevalent in the literature, are people less likely to respond to a web administered survey (or evaluation) than they would be were the survey administered via some other medium? Let's consider these -- in this order -- in the context of student evaluations.

For the opinion survey as a whole, the largest threat to accurate inferences from web surveys is that of coverage and sampling error (Couper, 1999). In the language of survey science, a mismatch exists between the target population and the frame (or those people who the surveyors have some sort of conduit of access to...for instance an Internet e-mail address). The classic example of such an error is the poll taken prior to the 1936 Presidential

election that predicted a win for Alf Landon over the incumbent Franklin Roosevelt. Roosevelt went on to win by a landslide margin and hindsight showed that the poll had been skewed because the frame used were people with telephones, a group that did not accurately represent the target of the entire national population. A, perhaps, more readily available example is that of surveying Internet users (via the Web) and attempting to make an inference about the entire population. Sampling error is related but more explicitly a concern in the cases of probability sampling. In short, in the words of Couper (1999: 473), "it arises from the fact that not all member of a frame are measured and the potential that a subsequent sample will produce a survey group with significantly different characteristics from the previous sample."

For a student population this concern is less paramount than it may in fact be when conducting surveys of the general population. Internet access is much more prevalent on the typical college campus in the U.S. than it is in society at large. It is unlikely that the School of Nursing would have even considered moving evaluations on line had the technological environment not evolved to the point where it was fairly conducive. Students in the Nursing program have access to several computer labs and the campus at large features many such labs, free e-mail and networked file storage space and a LAN with high bandwidth access to the Internet. Additionally, the University, in 2000, instituted a requirement for all incoming freshmen to purchase laptops, accelerating the reliance of both students and teachers on computers and Internet technologies.

With computer access a given, there is evidence that students electing to submit a survey via the Internet will not differ demographically from the student population as a whole. Handwerk, Carson and Blackwell (1999) studied the respondents to a survey of

college undergraduates regarding their satisfaction, in retrospect, with their decision to attend the particular institution in which they were enrolled. A sample of 3000 of the institution's students was divided with half asked to complete a paper version of the survey; the other half asked to complete the survey via a web form. Responses were not anonymous and a post-survey comparison of respondents found that there where no substantial differences between respondents to the paper survey and respondents to the online version with regard to sex, race, class (year) or housing situation. The only demographic difference discernible was a higher response rate among traditional age college students (defined here as 18-24) for the online survey.

Tomsic, Hendel, and Matross (2000) draw somewhat contradictory conclusions. They looked at demographic data from respondents to a survey of student satisfaction, administered in three successive years. Distributed survey packets included a web address and unique access code allowing students to complete the survey online if they so chose. Few students actually opted to respond online, overwhelmingly choosing to respond to the paper form. Evaluation of respondent characteristics found students from technological disciplines were more likely to respond to the online version, leading to the conclusion that online respondents might be different as a group. However, the study results point out that those choosing the online survey nearly doubled between the first and second years of the survey. The researchers conclude that while selection of the online format may indeed be an indication of a subject's comfort level with technology, the comfort level of the student population as a whole is broadening over time as computers and the Internet become more frequently used tools in higher education and American society at large.

Within the controlled, relatively homogeneous population of university students, the potential for coverage affects seem to be minimized. Similarly, there seems to be ample evidence that measurement error -- that is, a subject's potential to respond to an evaluation item differently given the different "delivery" format of the item – is not a problem. Thorpe (2002) administered three evaluations to two courses over subsequent semesters. In the Fall section students were equally divided and half were asked to complete a paper evaluation and half an online evaluation. Similarly the students enrolled in the Spring semester version of the class, which had two sections, were administered the evaluation in the same fashion. Of the 23 items on the evaluations -- all of which required Likert scaled responses -- only 3 items showed a measurable statistical difference in response between the two evaluation modes. The second course showed a statistical difference in only 2 items. The researcher did not discern a pattern in these differences between the classes and concluded that there was not a significant difference in responses based on mode. Cummings and Ballantyne (1999) found a similar result when they compared ratings gathered via an online evaluation form with the results gathered for the same class in the previous year via a paper form. Despite varying paper versus online response rates between the two student satisfaction surveys administered 9 months apart in 1996, Antons (1997) did not note a change in validity scores between the paper and online surveys, which remained consistent.

The final factor -- non-response bias, then, was the primary concern raised in the literature. Indeed, a survey conducted in 2000 of the 200 most wired campuses (as defined by Yahoo) found such. Administered by the Interactive and Distance Education Assessment Laboratory at Rensselaer Polytechnic Institute, the survey sought to gauge concerns and benefits associated with a change in mode of course evaluation administration, from paper

to electronic. Of the school's responding, 53% cited reduced return rate (from the return rates of paper evaluations) was the overriding problem in converting. Concerns such as user authentication, quality of response, and cost were all considered secondary concerns that appeared, by comparison, solvable.

Several of the studies looked at in this literature review found substantially lower response rates for online evaluation methods as opposed to paper administration. Cumming and Ballantyne (1999), for example, examined the administration of an online evaluation to 280 engineering students at Murdoch University in Australia. In prior years response rates to the paper evaluation form had averaged 65% returned, but the online system achieved only a 31% return rate. In a slightly different set-up, Tomsic (2000) offered students the choice of completing an evaluation online or via paper. For the three administration periods looked at respondents choosing the web form never rose about 14.9%, while those filling out the paper survey never fell below 85.1%.

In his review of the literature about online survey methods, Couper (1999) cites three major reasons that online surveys suffer from a non-response bias. First, analogs to tried and true motivational tools used in standard mail (reader paper, in this context) surveys have not been discovered. The second reason deals with technical difficulties. These arise from either a user's lack of familiarity and/or comfort with the Internet and computers in general. They may also arise from problems that users experience in the network transmission of survey materials, though these can also include hardware and software failures. Finally, Couper notes that some users see a potential for compromised confidentiality.

These reasons and more were found in studies of non-response in an academic environment for. Students that were more comfortable with technology and use of the Internet invariably responded in greater number to web forms when given the choice between the web and paper formats (Tomsic, 2000; Handwerk, 1999). In general, access to working technology was found to be a fundamental prerequisite for online responses. In Cummings and Ballantyne's (1999) work faculty reported, in post survey interviews, being unprepared to explain and prepare students for the online survey. That was postulated as a major role in the response drop off seen in that study. Motivations via reminder e-mails had only minimal affect at boosting results. Students also remarked that with no potential penalty for not responding they simply choose not to allocate time for completion of the evaluation.

However, despite substantial belief that online survey's suffer a non-response bias, it may, in fact, be more chimera than fact. All three of the studies mentioned in the preceding paragraph would seem to support the conclusion implicit in Couper's assertion that a fundamental reason for non-response bias is the newness of the web survey medium and, in particular, that the tried and true motivational tools of the paper survey have yet to be discovered. Non-response seems to be qualified in all three studies by, for instance, availability of web access and experience with web surveys. In fact, Thorpe's study found that response rates for the online survey actually exceeded the response to the paper form for two of the three class sections looked at. Only one class showed response rates that might be expected when comparing in-class versus web-based evaluation methods.

Notably the more recent studies support Tomsic's (2000) postulation that as the Internet becomes a more standard teaching tool and as students exposed to the Internet at a younger age begin to matriculate, a student's preference for a survey format may be the web

rather than paper. However, Thorpe (2002) is quick to point out that while the results seem to belie the standard concern, results at this point may not be generalizable across the entire breadth of higher education.

Development of the Online System

The fundamental goal of the committee considering conversion to an online evaluation was to replicate the current paper-and-pencil question online. Several enhancements, however, were discussed in initial meetings. The foremost of those was creating a system in which faculty could, if they desired, add questions to their evaluations. The school would require that all evaluations cover the "core" items (mentioned above in discussion of the paper system), but faculty would be able to add other questions to this core -- presumably from some sort of question bank -- to gather information about particular items that might be relevant to their particular course. Additionally, committee members envisioned an online system with a more streamlined process for scheduling an evaluation so that rather than the standard end of semester evaluation period, evaluations could be carried out at any point during the semester. With the prevalence of team teaching across the curriculum, such a refinement would be a major step beyond the paper system. Also, ready access to older, archived data was identified as a key possible enhancement.

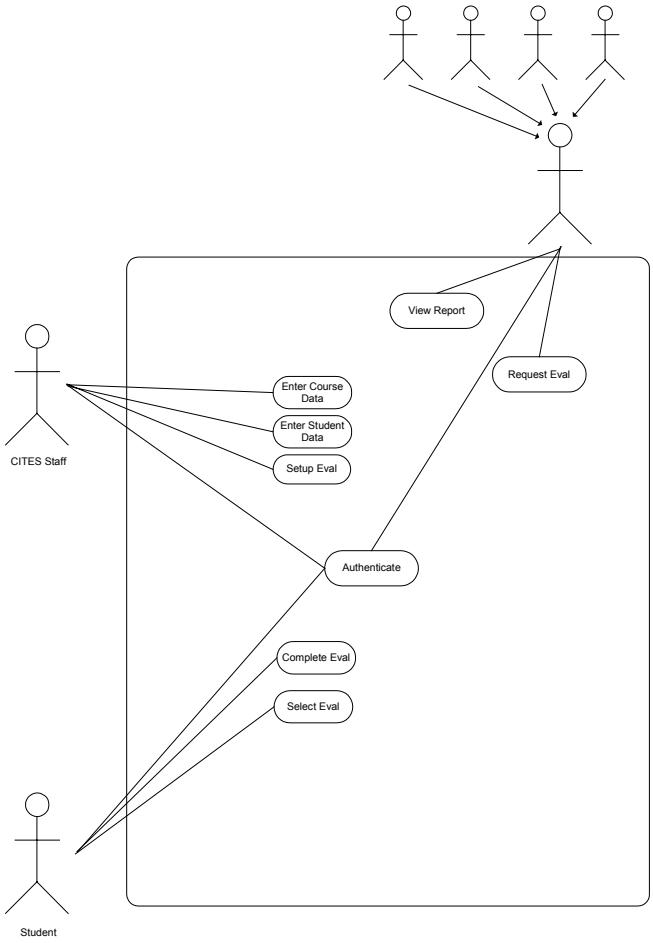
Because the web-based course evaluation development process was internal, where the primary actor in the actual construction of the system was not only a school employee but a functioning member of the committee developing requirements, the development process was more informal and iterative than might otherwise be the case. Prototypes were developed and tested with actual users and additional features were added as they were identified. Much of the knowledge necessary to perform the work was already available

either via understanding of school process, systems and user aptitudes gained through several previous years of work experience in the school or via informal conversations and brief e-mail exchanges.

In a large sense much of the information gathering piece of the process was accomplished through committee meetings. The make-up of the committee was representative enough that inferences could be made about how the current system worked from the standpoint of evaluation data users. Details of the actual administration of the paper evaluations were learned through interviews and e-mails with the staff person most directly involved with the administration of the evaluations. Informal conversations with people on the periphery of the process were also useful, as was examination of written, albeit limited, documentation.

Once information had been gathered from the various sources on the current system, the first goal was to formally identify the users of the system. There are three main groups. Students use the system to identify evaluations that they are responsible for completing and to select and complete particular evaluations. Faculty and school administrators access the system to retrieve results for completed evaluations. It is helpful to subdivide this group because of the tiered access structure (see the discussion of the current system). The specializations reflect the access policy: teaching or clinical faculty comprise one sub-group; those functioning as the course coordinator a second. And then there are administrators: faculty division chairs and the Associate Dean for Academic Affairs. The third type of user that interacts with the system are the system administrators. Their tasks include setting up evaluations, associating students with particular evaluations and collecting

results for report generation one the evaluations are complete. The following use case diagram gives a visual representation of the model used to guide the development process.



Once the analysis of users and their uses of the system was complete, a system backed by a relational database seemed the obvious choice. A database would provide the storage capabilities that file cabinets in the paper-and-pencil system offered. Further the database, through queries capabilities, would provide the flexibility to reconfigure data into reports spanning class offerings over time and the ratings of professor contributions to an number of courses, semester-to-semester. As mentioned before, this kind of reporting was achievable in the old system, but only at the expense of substantial work hours.

Another factor influencing the choice of a database system was the access that CITES staff (who were responsible for the development of the system) had to an in house Cold Fusion server. Cold Fusion is a proprietary development environment produced by Macromedia. Essentially a scripting language offering the same kind of functionality that language/development environments such as PHP, Microsoft's ASP, Perl, Python and Ruby offer, Cold Fusion has built-in interfacing mechanisms to a variety of database management systems and the capability to handle CGI-type input from web-based forms. The CITES staff facility also had several years of experience developing thin the Cold Fusion environment reducing any need for staff training or reliance on external out-sourcing of development tasks.

Additionally, it seemed a requirement that system user interfaces should be web-based. As mentioned students at the Nursing School are increasingly comfortable getting syllabi information, submitting assignments and communicating with fellow students and professors on the web. Many courses at nursing have a website and a great majority of these are hosted on a campus server running the courseware package Blackboard. Aside from expanded and built-in functionality, Blackboard gives course web sites a similar template providing the students with navigational familiarity. Blackboard also -- and this became significant -- has a function for administering tests and surveys. Additionally the laptop requirement for incoming students –implemented in 2000 for the University – means that as of Fall 2002 all of the incoming students (students begin the Nursing undergraduate program at the start of their junior year) have their own personal computers which can access the campus network (and thus the web) by plugging in at any of a multitude of points on campus or by using the wireless network.

Aside from providing an environment where students should presumably feel comfortable because of their abundant experience, use of Blackboard also provided a vehicle for authentication, one of the most challenging aspects of the new system's design. One of the fundamental goals set out for the online system was anonymity. It was felt by the committee that the web-based system should at least provide students with as much confidentiality as the pencil-and-paper system. Since the evaluations are distributed to the class as a whole and returned within a specific class period and since the forms contain no identifying information other than a code distinguishing this classes evaluations from another class, tracking an evaluation from a particular student is impossible. Handwriting, as pointed out above, is however potentially recognizable in the written comments which are copied and returned to the professor, who might be able to identify a student based on penmanship.

But offering evaluations over the Internet -- with it's basis in open and free exchange of information -- necessitated that students enter some identifying information in order to bar non-students, or perhaps more likely, students not enrolled in the particular course. The option of passing out or communicating to a class of students a particular code was discussed, but was soon abandoned as too clumsy. Students might easily lose or forget the code, and, in general, it seemed to add another potential source of human error. Other identifying methods were discussed such as the student's university unique identifier (known colloquially as the PID), or the exam code, which is distributed to students by the school so that exam results can be publicly posted for students. But these, too, were judged to increase the complexity for students or, in the case of the PID, be an identifier that was not

completely secure since it is by design a number used to uniquely identify students in a number of University data stores.

Any system that students would have to acclimate themselves to would be likely to increase the non-response bias already identified by the literature as the biggest pitfall of an online system. All of the various login code solutions appeared to increase the complexity for students. All were items that are either entirely new or used sporadically enough to think that students might have to go and search for and/or do some nontrivial memory recall in order to be able to enter the system. On the other hand, Blackboard is based on a university wide authentication system driven by Kerberos. Students have one identifier that allows them access to a number of university computer resources from storage space on central campus file servers to library electronic databases. Students are used to logging in to the Blackboard system -- they do so, presumably, daily -- with this identifier ID and password. That reason, coupled with most students's assumed comfort and familiarity with the look and feel of Blackboard's interface and navigational elements made Blackboard seem like the logical hosting environment for the evaluation system.

Actual data collection (i.e. students actually taking the evaluations) was less of a problem. While the Blackboard system provides a template and many teaching tools and communication features, it also is flexible enough that it allows course designers to provide their own HTML content. Laid out in frames, custom designed pages can be rendered in the main content window, while left side and top edge navigational elements remain set. The content window can even point to and render data that exists on a separate server. Using then the built form elements of HTML, we were could construct two evaluation templates that matched the paper version, substituting radio buttons for the Likert scaled questions

previous answered by "bubbling in" Scantron sheets and textarea field for the open ended questions.

In fact, further investigation of the Blackboard environment revealed that the software's built-in survey tool could actually accomplish this task quite well and potentially save the development time involved in writing Cold Fusion scripts to handle and aggregate answers that students submitted on the HTML forms. This also had the added inducement of making submitted data even more "untrackable" and anonymous, since it would be processed and aggregated essentially within a black box (i.e. a machine and computer code which we did not have access to or knowledge of).

The decision to use Blackboard's survey tool, however, did not come without a new set of challenges. In short, since Blackboard is designed as a courseware tool, the users who have access to a particular site are assumed to be either students or instructors. (There are some other user roles defined but their capabilities differ only marginally from either of these two roles.) A student given access to the site in order to complete an evaluation would – by nature of his or her role as a Blackboard "student" – have access to all items in the site, namely all the surveys available at the site, be it a survey they were required to complete or not. In the interest of making the user's experience as uncomplicated as possible, rather than expecting students to fish through a list of many evaluations for many classes to find those required of them, a HTML form was created to serve as a gateway, filtering evaluations so that students would see only the evaluations for which they were responsible.

On entering the evaluation site, Blackboard initially directs students to this form. It consists of introductory text that stressing the importance of the evaluation process and the

seriousness with which faculty and administration consider the results of evaluations. That is followed by a text input box where students enter an identifier -- the student PID was decided upon -- and then submit the form. The form, in actuality, submits the PID to a script on the Cold Fusion survey maintained at the School of Nursing. The script queries the database table that associates the many-to-many relationship between the student entity and the evaluation entity. In short the table links a particular student to the particular evaluations for which they are responsible. These results then get returned to the student with a list of evaluations hyperlinked to the URL of the actual evaluation form on the Blackboard server.

A second challenge was the "harvesting" of the aggregated survey/evaluation results from Blackboard at the end of the survey period. Blackboard provides a running results page -- available to only those with "instructor" status in the site -- showing percentages of those selecting each response for each item and a listing of the whether a free form answer was given to those questions, and, if so, what that answer was. However, reports that faculty and administration receive in the current system also contain means and standard deviations. Additionally, the system needed to provide the capability of archiving data and generating reports that would give aggregate data across semester and cumulative data for any specific faculty member. To gather the "raw" data that would enable calculation of these figures and reporting capabilities, a script was written in Cold Fusion that would parse the result pages generated by Blackboard and store raw answer counts in a results table in the database. Basically, the process works as follows, the HTML generated results pages from the Blackboard server are saved as HTML files and then "fed" through the parse script. Because the results pages are themselves "machine" generated their HTML code is predicable enough that, through a series of regular expressions in the Cold Fusion script, the raw data can be

extracted from surrounding tags, stored in an array, and then the array is traversed and the data are fed into the database via an ODBC connection.

The database underlying these scripts is represented graphically in the Entity-Relationship diagram included in the appendices. It consists of four main entities: students, faculty, courses, and evaluations. There is also a fifth, more minor entity: adjunct faculty. It captures those faculty from other health science schools at the University who participate in nursing courses as guest lectures, but who are never actually employed by the school as faculty.

Students are described by a name, a university assigned unique identifier known as a PID, an e-mail address and a Kerberos ID known as an ONYEN (an acronym for: only name you'll ever need). Similarly faculty are identified by PID, name and their faculty division. The faculty entity is "specialized" into overlapping sub-entities reflecting the different roles that an instance of the faculty can play in the system: course coordinator and/or report viewer (i.e. one who views results of evaluations). Courses are identified with a combination primary key of course number and course section. They also have the attributes of title, semester, course coordinator.

The evaluation entity is the centerpiece of the system. It has the attributes: ID, Blackboard ID, URL (both used to identify the survey in the Blackboard system), title and the associated course. The evaluation entity is also specialized in three sub-types. Primarily this is done to keep database tables in normalized form, but also helps in creation of queries for report generation. The entity is the subject of two key relationships: the viewing relationship that the report user entity participates in and the tertiary relationship that a

course, a student and an evaluation participate in when a course requires an evaluation of a student. Both of these relationships are many-to-many in cardinality and thus comprise individual tables in the database. These are the tables that ultimately control access, allowing faculty access to appropriate result data and students to appropriate evaluation forms.

Piloting the System

The first test of system occurred fairly informally with several courses taught in summer 2002 by one of the committee members and one large course taught by an adjunct faculty member. The committee's concern -- given reports from the literature and the experience of another school at the University who had implemented an online system -- was whether or not students, freed from the more implicit incentive of devoted class time to completion on an evaluation, would actually bother to fill out and evaluation. The response rates were encouraging in all but one course. Although effort was made in this test to gather feedback on the system, several students volunteered very positive responses and all faculty seemed both pleased with the administration and with the relative quickness that they received results. The faculty member of the one course that received a low response rate attributed it to two factors: this was the final class before students graduated from an intensive program and his lack of emphasis on the evaluation.

Measures of System Effectiveness

Survey of Student Attitudes toward online system

A more formal pilot test was run in October of 2002, halfway through the Fall 2002 semester. Four large second year undergraduate classes were asked to evaluate their clinical faculty member at the conclusion of their first clinical rotation of the semester, which ends at the halfway point. This time more formal efforts were made to gather feedback. Not only

were response rates monitored, but because all of these courses are curriculum requirements, not electives, comparisons between previous year data could be made. Additionally, a week after the end of the evaluation period students were asked via an e-mail to respond to a brief survey about the online system.

The survey was web-based and designed to take students no more than a minute, as was advertised in the e-mail solicitation, assuming that a shorter survey would maximize response rates to an item that students would have little incentive to complete. E-mails included a link to the survey page; with each URL carrying an encoded identifier to prevent duplicate or spurious responses. On the initial screen the reasons for the survey were explained -- reiterating and in some cases duplicating the messages of the e-mail. The next screen asked students whether they had in fact completed an evaluation and depending on that their answer -- yes or no -- they were directed to a page containing four statements relating to their experience with the evaluation or system or, conversely, their four reasons for non-response distilled from findings in the literature. Each group was asked to rate their level of agreement -- from strongly disagree to strongly agree -- on a five point Likert-scale to each of the statements. The table gives the number of participants, breaks that down into those answering yes and no and for each group gives the aggregate percentages of each response.

Number of respondents: 67 out of 129					
Yes = 60	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Q1 (n: 60) **The online evaluation(s) were easier to complete than the paper evaluation(s).**					
	1.67%	0.00%	6.67%	28.33%	63.33%

Q2 (n: 59) **I found that I had more time to reflect upon and write my answers because I could do the evaluation(s) outside of class.**					
	1.69%	6.78%	6.78%	35.59%	49.15%
Q3 (n: 60) **I was able to complete the online evaluation(s) faster than the paper evaluations.**					
	0.00%	0.00%	20.00%	28.33%	51.67%
Q4 (n: 60) **I felt more anonymous submitting my evaluation(s) online than I have, in the past, on paper.**					
	5.00%	13.33%	43.33%	15.00%	23.33%
No = 7					
Q1 (n: 7) **Because I had to complete the evaluations outside of class, on my own time, I did not have time to complete the evaluation(s).**					
	14.29%	71.43%	14.29%	0.00%	0.00%
Q2 (n: 7) **I did not feel that the online evaluation system was as confidential and anonymous as was claimed.**					
	0.00%	28.57%	28.57%	42.86%	0.00%
Q3 (n: 7) **I am not comfortable using the Internet -- and computers in general -- so I did not complete the evaluation(s).**					
	42.86%	42.86%	14.29%	0.00%	0.00%
Q4 (n: 7) **I feel that doing evaluations -- no matter what format, paper-and-pencil or online -- is a waste of my time.**					
	57.14%	14.29%	14.29%	0.00%	14.29%

The answers from those responding to the evaluation show a very strong nearly completely non-negative response in terms of the ease of use. While such attitudes might be expected from the responder group, the preponderance of agreement with the two statements seems potentially significant. Perhaps, as suggested by both Tomsic and Thorpe, that increasing familiarity and exposure to technology is working to reduce non-response bias. Many of students who took these evaluations would have just begun high school (or

thereabouts) when use of the web had reached critical mass. In not in only a few years, the matriculating student will have been of elementary school age when that saturation began.

But such extrapolations are debatable. Less questionable is the fact that ease of use stands as the clear preference among the statements. Students seem nearly as positive about the potential of giving more reflective answers using the online system. The result of the fourth item, comparing the confidentiality of responses, is in contrast rather interesting. Written instruction and statements about the online system continually tout the lack of access of faculty to the Blackboard site that "houses" the evaluations but still the responses here seem to indicate a certain level of uncertainty. However, it may well be a case of "protesting too much." Students were confronted with such claims so frequently that they've applied a skeptical level of early-21st century cynicism to the assertions. On the other hand, this response pattern could reflect participant's lack on confidentiality in the paper system.

Drawing conclusions from the non-responding group is likely even more perilous given the very small (by comparison) sample size. Probably the only conclusion that can safely be drawn is that none of the statements offered resonated very deeply with non-responders as a reason for their non-response. Antons, Dilla, and Fults (1997) notes, by way of conclusion, that a core of students may exist in any evaluation system who are stubbornly resistant to any prods toward participation in such a ratings system, and we may be witnessing that resistance in the 7 non-responders here.

Response rate comparison with previous year

As was mentioned additional data was gathered from comparisons with evaluations for the same classes from the previous year where evaluations were administered on paper.

The initial observation was that the data from the previous year was highly disorganized with substantial variation in response rates. Some classes showed full response (i.e. everyone given an evaluation turned one in) to some classes where the response rate was 0% indicating either lost data or an evaluation that had in fact not been administered. On top of the rather disconcerting implications of this discovery were the perhaps more troubling discrepancies between actual students in certain clinical sections and evaluation return rates that exceeded records for the number of students in the section. An attempt is made to summarize this comparison in the table below.

	2001 % responding	2002 % responding
NURS 074	34%	65%
NURS 073	103%	81%
NURS 071	76%	66%
NURS 075	44%	72%

Even if one disregards the chaotic, highly suspect state of the 2001 archive, there is indication that, from an aggregate standpoint, the return rates for the online system pilot are similar enough that further investigation of system is warranted. In addition, there is likely to be a period of adjustment when switching the evaluation mode from paper to online, just as there is for many systems. As Cummings intimates, faculty may need to become well versed

enough in the process to provide the proper motivation. Indeed, techniques for generating responses to online surveys are still evolving (Couper, 1999).

Word count comparisons with previous year

Word counts for the open ended questions were also compared. The NURS 073 class was chosen for this comparison because the same four faculty members served as clinical faculty in both years. On average the answer length was up substantially for the online evaluations. The question asking students to comment on the professor's positives had an average response of 25.7 words on the paper evaluation. The length grew to 32.6 words on the online forms. Similarly responses to the negative question was 2.3 words in 2001 and 7.2 in 2002. A number of factors may account for this rise. For one, the paper form allows about an inch of writing space for each question. Students are free to write on the back of forms, but in general seems to confine them selves to this space, which, depending on the size of their handwriting serves as a constraint. On the Web form the text area is five lines long, but the comments can be -- by way of the built-in scrolling mechanism of the form element -- unlimited in length. However, it is possible that the longer responses bare out findings of the literature that online surveys, because they can be completed outside of class and allow more time for student reflection, solicit richer, more considered responses.

Conclusions

Given the burgeoning use of computers and Internet technologies in higher education, moving administration student evaluations of course and teaching from pencil-and-paper format to an online system would seem a logical evolution of a system that has become a fundamental tool for improving instruction and component for administrative personnel decisions. While early implementations of online student evaluation systems

seemed to avoid the pitfalls that can plague the general online survey, non-response bias was seen to be a significant barrier to a paper to electronic transition. However, there have been indications that response rates might improve once use of computers and the Internet becomes more common both in the higher education environment and in the course of a student's experience before reaching college.

But the factors causing reduced response rates when transitioning from paper to electronic survey formats are far from definitive, as are effective methods for stimulating response. While the web offers capabilities to survey designers that far exceed those of pencil-and-paper, the guiding principle used in development of the online system at the School of Nursing leaned toward a simple interface presented in a navigational and authentication environment that students would find comfortable and familiar. Response rates for the pilot of this system, which were fairly promising, may bare out the assertion of the most recent studies that non-response bias is a decreasing problem. But certainly, the strong positive attitude expressed by respondents to the post-pilot survey indicates that evaluations must be perceived as less cumbersome to use than paper evaluations and quicker to complete.

Bibliography

Antons, C., Dilla, B., Fultz, M. (1997). Assessing Student Attitudes: Computer versus Pencil-and-Paper Administration. Paper presented at the Association for Institutional Research Annual Forum, May 1997. [ERIC Document Reproduction Service ED 410 866].

Asiu, B., Antons, C., Fultz, M. (1998). Undergraduate Perceptions of Survey Participation: Improving Response Rates and Validity. Paper presented at the Association for Institutional Research Annual Forum, May 1998. [ERIC Document Reproduction Service ED 422 805].

Bosnjak, M., Tuten, T. (2001). Classifying Response Behaviors in Web-based Surveys. *Journal of Computer Mediated Communication*, (6).

Cashin, W., (1995). Student Ratings of Teaching: The Research Revisited. Paper published by the Center for Faculty Evaluation and Development at Kansas State University. [Retrieved from World Wide Web at http://www.idea.ksu.edu/papers/pdf/Idea_Paper_32.pdf]

Couper, M (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, (64), pp. 464-494.

Cummings, R. and Ballantyne, C (1999). Student Feedback on Teaching: Online ! On Target? Paper presented at the Australasian Society Annual Conference, October 1999. [Retrieved from World Wide Web at http://cleo.murdoch.edu.au/evaluation/pubs/confs/aes99.html]

Dillman, D., Tortora, R., Conradt, J., Bowker, D. (1999). Influence of plain vs. fancy design on response rates for web surveys. [Retrieved from World Wide Web at http://survey.sesrc.wsu.edu/dillman/papers/asa98ppr.pdf].

Handwerk, P., Carson, C., and Blackwell, K. (2000). On-line vs. paper-and-pencil surveying of students: A case study. Paper presented at the 40th Annual Meeting of the Association of Institutional Research, May 2000. [ERIC Document Reproduction Service ED 446 512].

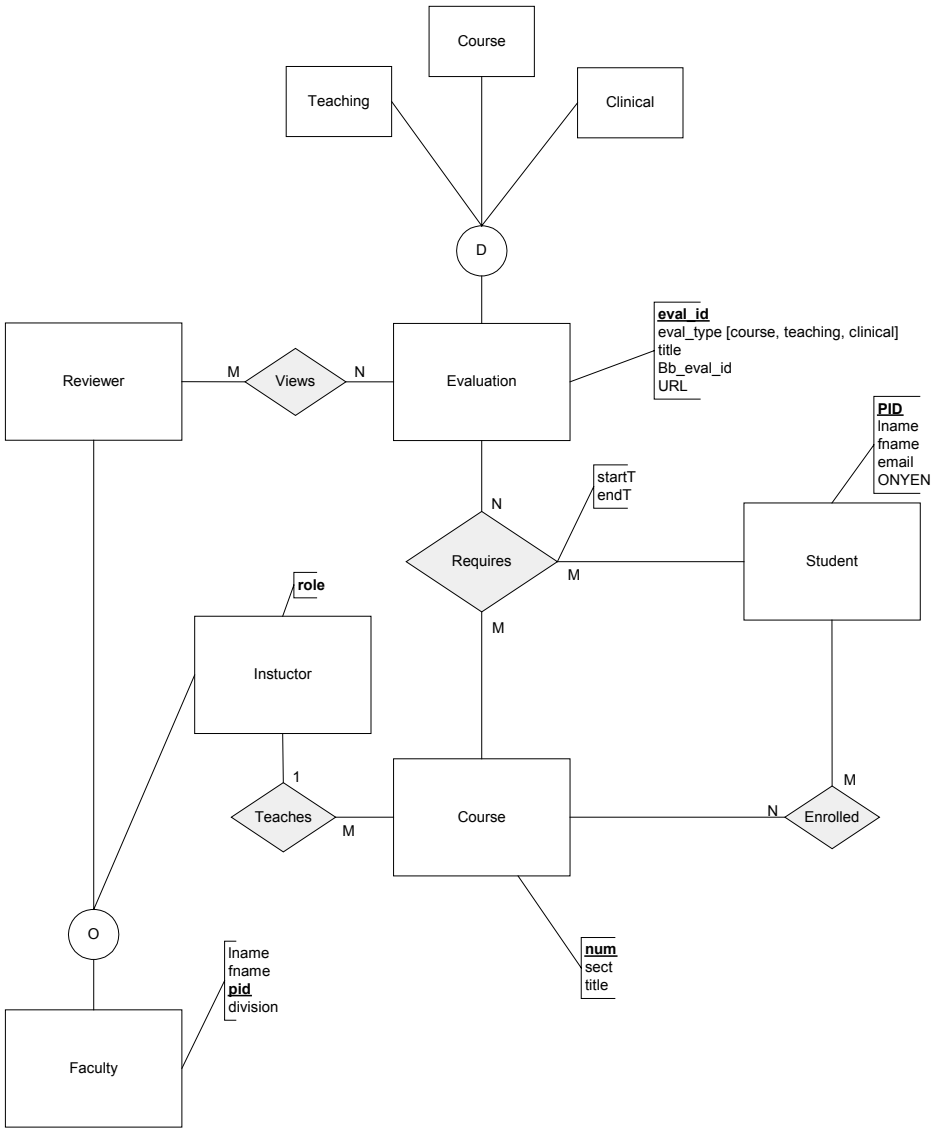
Hmieleski, K. (2000). Barriers to online evaluation: Surveying the nation's top 200 most wired colleges. Report prepared by the Interactive and Distance Education Assessment Laboratory at Rensselaer Polytechnic Institute, Troy. NY. [Retrieved from World Wide Web at http://idea.psych.rpi.edu/evaluation/barriers.doc].

Matz, M. (1999). Administration of web versus paper surveys: Mode effects and response rates. Masters Research Paper, University of North Carolina at Chapel Hill. [Retrieved from World Wide Web at http://ils.unc.edu/MSpapers/2555.pdf]

Thorpe, S. (2002). Online Student Evaluation of Instruction: An Investigation of Non-Response Bias. Paper presented at the 42nd Annual Forum of the Association of Institutional Research, June 2002.

Tomsic, M., Hendel, D., Matross, R. (2000). A world wide web response to student satisfaction surveys: Comparisons using paper and Internet formats. Paper presented at the 40th Annual Meeting of the Association of Institutional Research, May 2000. [ERIC Document Reproduction Service ED 445 633].

Entity-Relationship Diagram



Data Dictionary

Student		
PID	Char (9)	9 digit University assigned unique identifier
lname	Char (50)	Student surname
fname	Char (35)	Student given name
email	Char (50)	email address of student
added	Date	Date/Time stamp recording when student added to table

Course		
num	Char(3)	Course number – assigned by University Registrar Combination Primary Key with ***section***
section	Char(3)	Three digit section number, differentiating multiple offerings of course in same semester; combination Primary Key with ***num***
title	Char(50)	Course name

Course_instance		
cnum	Char(3)	Foreign key of course
csect	Char(3)	Foreign key of course
instructor_id	Char(9)	University unique identifier – FK of FacAdministration
semester	Char (4)	Four digit University assigned code for semester and semester year

Evaluation		
eval_id_num	AutoNumber	Primary key; sequentially incremented integer
Bb_eval_id	Char(8)	Identifier for survey on Blackboard system; assigned by Blackboard software upon creation of evaluation
title	Char(50)	Title for evaluation
url	Char(125)	URL of survey created in Blackboard

Eval_instance		
sPID	Char(9)	Student PID – FK of Student
eval_id_num	Long Integer	Foreign key of evaluation
cnum	Char(3)	Foreign key of course
csect	Char(3)	Foreign key of course
completed	Binary	Yes/No indication of whether student has completed eval or not

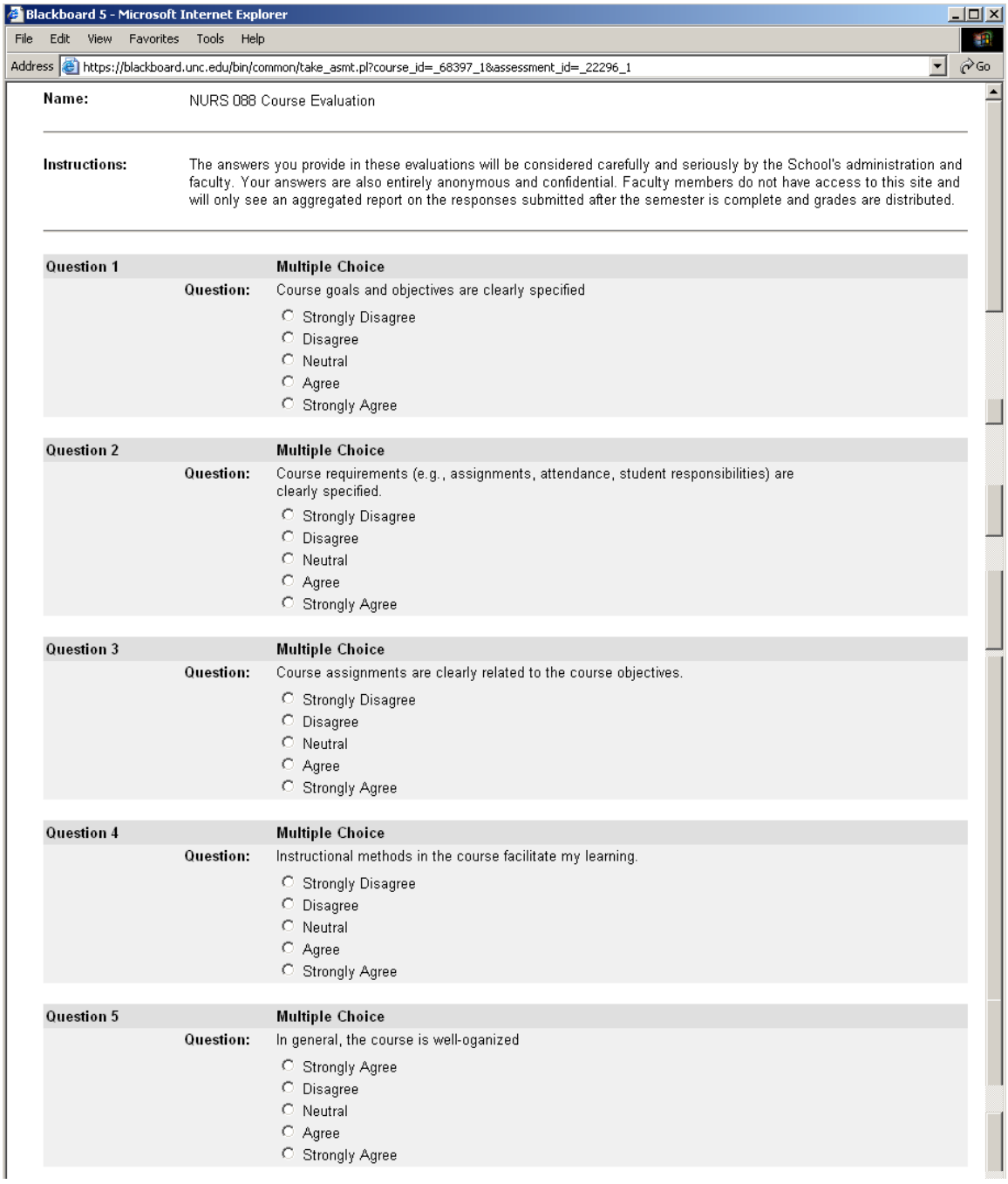
FacAdministrator		
PID	Char(9)	Primary Key – University assigned unique identifier
lname	Char(50)	Faculty member surname
fname	Char(35)	Faculty member given name
division	Char(3)	I, II, RSC, CLN – denoting faculty administrative division that faculty member belongs

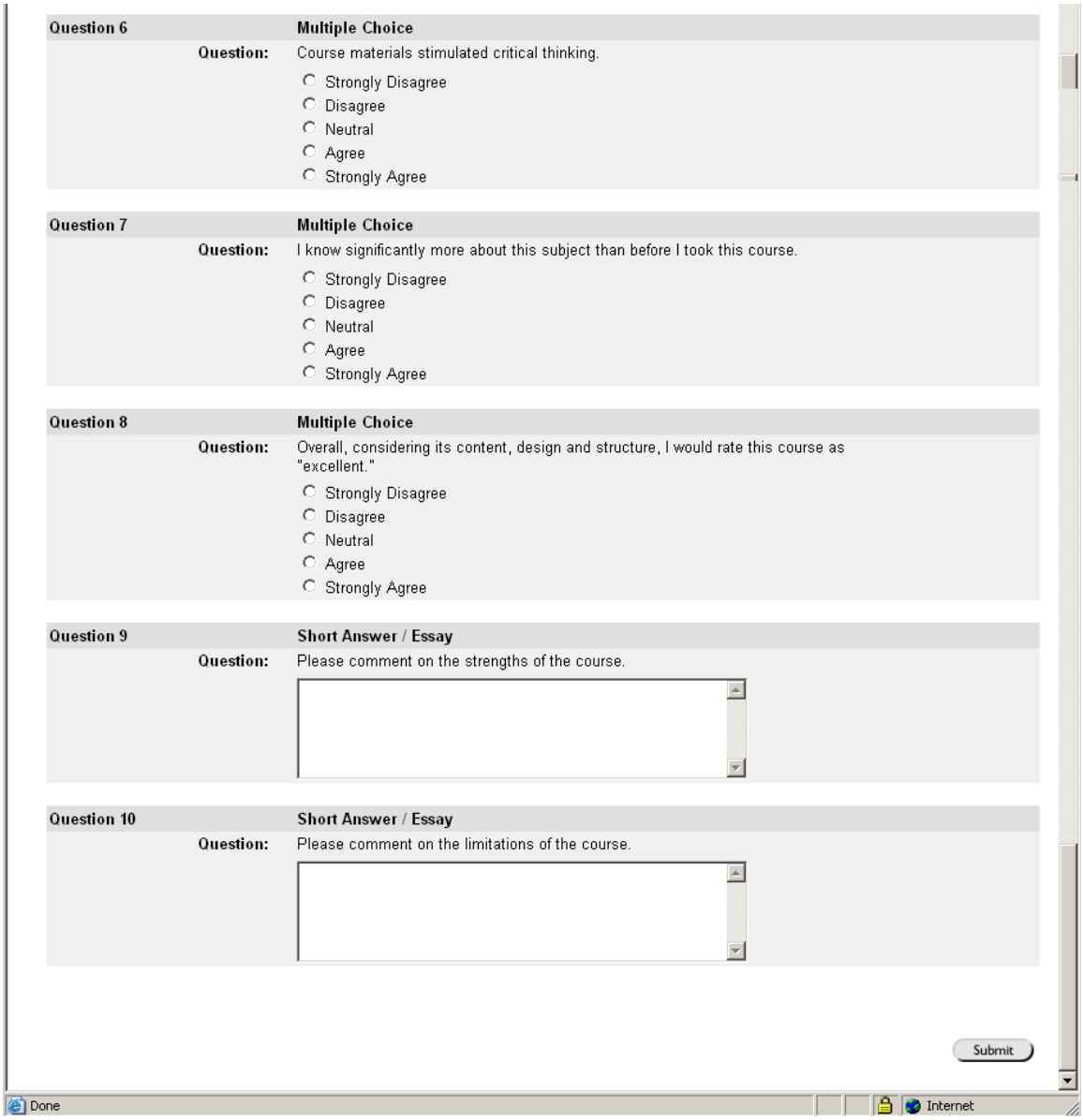
Adjunct		
PID	Char(9)	University unique identifier – FK in FacAdministrator
Affiliation	Char(40)	Department or organization that is adjunct faculty member's primary employer

Report_access		
PID	Char(9)	Foreign key of FacAdministrator
eval_id_num	Long Integer	Foreign key of evaluation

Screen Shot: Student Evaluation "Gateway" Form

Screen Shot: Course Evaluation HTML Form





Question 6 — **Multiple Choice**

Question: Course materials stimulated critical thinking.

- ○ Strongly Disagree
- ○ Disagree
- ○ Neutral
- ○ Agree
- ○ Strongly Agree

Question 7 — **Multiple Choice**

Question: I know significantly more about this subject than before I took this course.

- ○ Strongly Disagree
- ○ Disagree
- ○ Neutral
- ○ Agree
- ○ Strongly Agree

Question 8 — **Multiple Choice**

Question: Overall, considering its content, design and structure, I would rate this course as "excellent."

- ○ Strongly Disagree
- ○ Disagree
- ○ Neutral
- ○ Agree
- ○ Strongly Agree

Question 9 — **Short Answer / Essay**

Question: Please comment on the strengths of the course.

Question 10 — **Short Answer / Essay**

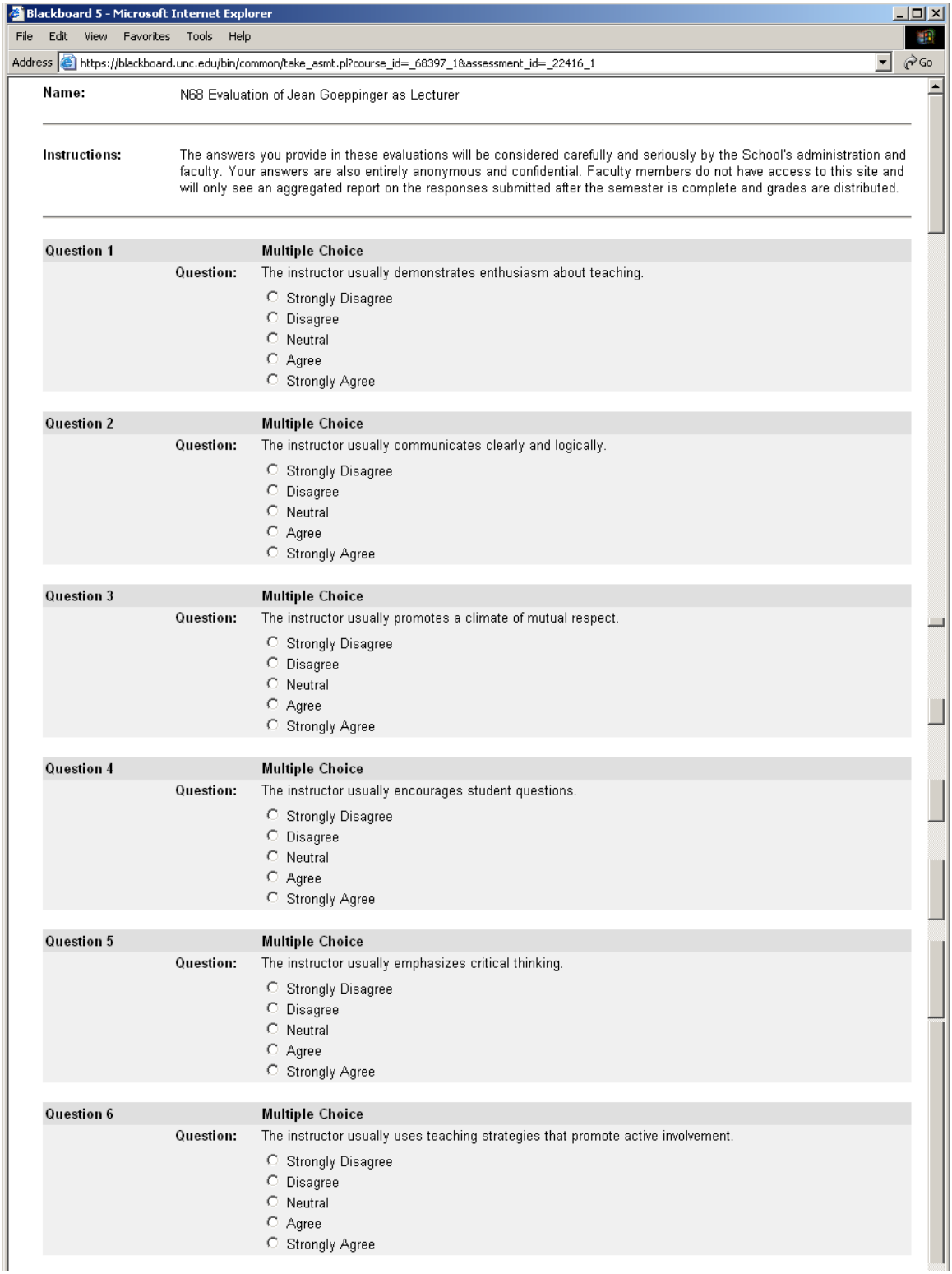
Question: Please comment on the limitations of the course.

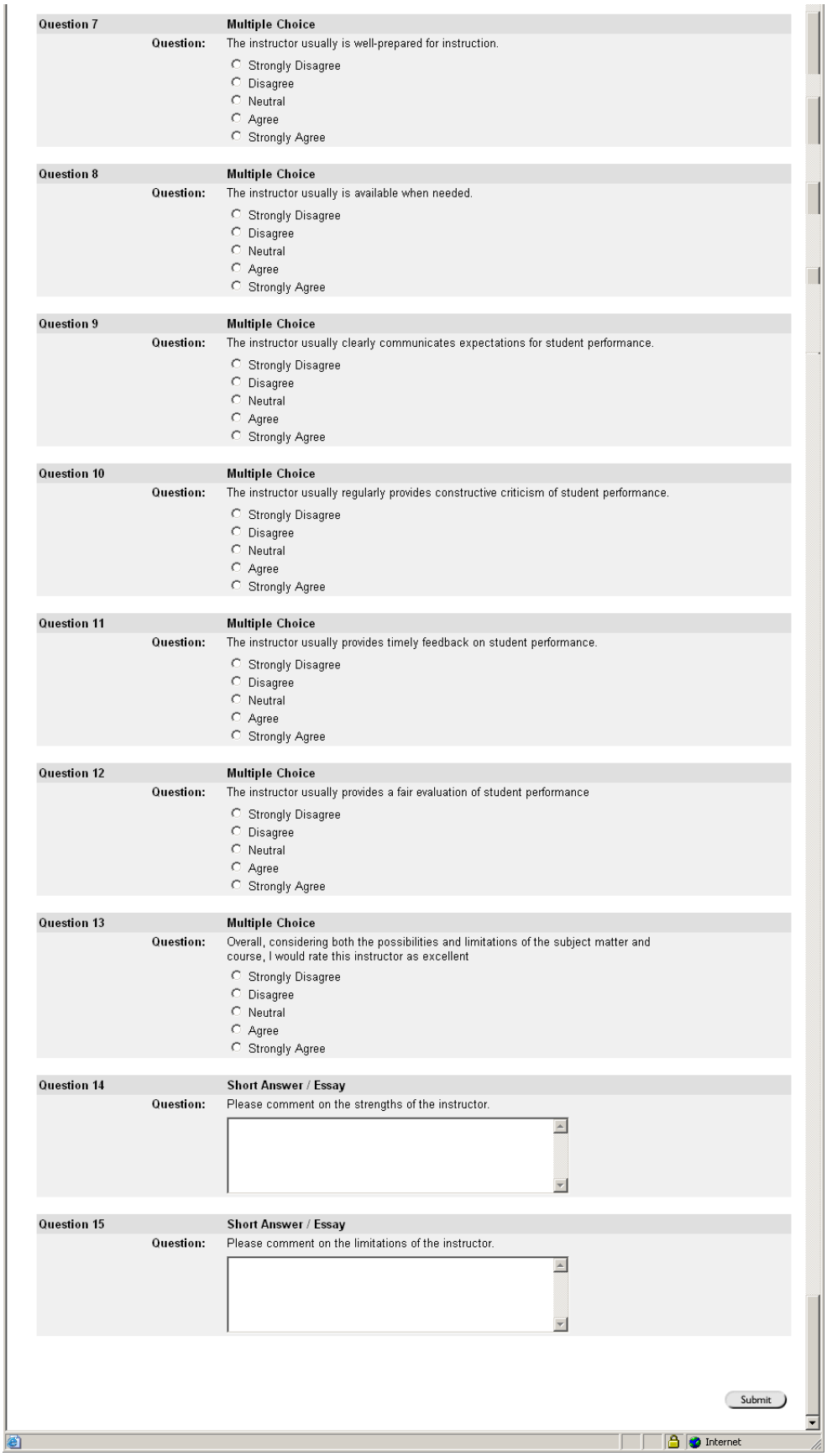
Submit

Done

Internet

Screen Shot: Teaching Evaluation HTML Form





Question 7 — **Multiple Choice**

Question: The instructor usually is well-prepared for instruction.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 8 — **Multiple Choice**

Question: The instructor usually is available when needed.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 9 — **Multiple Choice**

Question: The instructor usually clearly communicates expectations for student performance.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 10 — **Multiple Choice**

Question: The instructor usually regularly provides constructive criticism of student performance.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 11 — **Multiple Choice**

Question: The instructor usually provides timely feedback on student performance.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 12 — **Multiple Choice**

Question: The instructor usually provides a fair evaluation of student performance

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 13 — **Multiple Choice**

Question: Overall, considering both the possibilities and limitations of the subject matter and course, I would rate this instructor as excellent

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Question 14 — **Short Answer / Essay**

Question: Please comment on the strengths of the instructor.

Question 15 — **Short Answer / Essay**

Question: Please comment on the limitations of the instructor.

Submit

Internet