ONLINE COURSE EVALUATIONS

REPORT #1 OF THE TASK FORCE ON ONLINE EVALUATIONS & PLACEMENT EXAMINATIONS

The University of Michigan

March 28, 2007

An online copy of this report is available for download at the following site:

www.provost.umich.edu/reports/online_course_evaluations.pdf

TABLE OF CONTENTS

TASK FORCE MEMBERS

Gary Herrin, Task Force Chair, Associate Dean for Undergraduate Education and Professor, College of Engineering

Maria De Lourdes Dorantes, Interim Director of the Elementary Spanish Program, Lecturer IV, Department of Romance Languages and Literatures

Lisa Emery, Senior Business Systems Analyst, Michigan Administrative Information Systems (MAIS)

Annemarie Hindman, Center for Research on Learning and Teaching and Rackham Graduate Student

James A. Kulik, Director and Research Scientist, Office of Evaluations and Examinations

Helene Neu, Director of the Elementary French Program, Lecturer IV, Department of Romance Languages and Literatures

Diana Perpich, Usability Support and Evaluation Lab (USE-LAB) Consultant, Digital Media Commons

Michael Thouless, Professor, Materials Science & Mechanical Engineering

Gretchen Weir, Assistant Vice Provost for Academic Affairs

PREFACE

WE PRESENT THIS REPORT to the Provost of the University, the Senior Vice Provost for Academic Affairs, the deans of the schools and colleges of the University, and the University community for their consideration.

The report is our response to a proposal to replace the University's existing system of course evaluations with a Web-based approach. The proposal raised questions with implications for students, staff, and faculty. Should the University move to online ratings? Can the University avoid problems commonly associated with such systems? How can the University plan for further changes in its rating system?

To answer such questions, the Provost and the Senior Vice Provost set up a task force on online evaluations and examinations. The mandate of the task force was to develop recommendations in two areas:

> The committee is to consider how placement exams, now administered during new student orientation or proctored individually, and teaching evaluations, currently collected on paper, can be made more efficient and less costly by being administered online. The task force is not being asked to create new tools but rather to "web-ify" the existing tools. The committee is not responsible for implementing electronic systems.

This report, which focuses on online evaluations, is the first of two that we are issuing in response to this mandate.

To reach our recommendations, we reviewed evidence from pilot studies of online ratings carried out at the University, field studies carried out at other institutions, and discussions between individual task force members and faculty, students, administrators, and support staff. Task force members wrote background papers on the issues raised by the evidence, and we discussed the issues extensively before reaching our conclusions.

We came to realize that collecting student ratings of teaching is truly a community effort. The success of a student rating system depends on the efforts of many individuals: teachers who design the questionnaires, distribute them, and interpret summary reports; students who fill out the forms; administrators who review and interpret results; and departmental staff members who handle orders, questionnaires, and reports. The smooth functioning of an evaluation system also requires the good will and support of University organizations, including the Faculty Senate, the Graduate Student Organization, the Lecturers' Employees Organization, and the Michigan Student Assembly. All are stakeholders in course evaluation, and all are affected by changes in the University's evaluation system.

Stakeholders sometimes see the rating system from a limited perspective—a soldier's view of an army. Our position on the task force gave us a broader view, and we attempt in this report to share this view with other stakeholders. Our report therefore

describes in some detail the evidence we reviewed, our reasoning about the issues, and the recommendations that flowed from our investigations and deliberations.

We are grateful for the opportunity that we had to reflect on issues that affect so many in the University community. We hope that our report will lead to further reflection and to action.

Gary Herrin (chair),
Maria Lourdes Dorantes, Lisa Emery,
Annemarie Hindman, James Kulik, Helene Neu,
Diana Perpich, Michael Thouless, Gretchen Weir

EXECUTIVE SUMMARY

COLLECTING STUDENT RATINGS on the Web seems to be an idea whose time has come. At least a dozen universities have developed school-wide programs of online ratings during the past decade, and more than fifty other schools are now collecting online ratings in some departments or courses. In addition, survey results suggest that nearly half the schools in the country are thinking about establishing online rating systems.

Compared to paper-based rating systems, online systems promise lower costs, timelier feedback for teachers, more anonymity for students, better safeguards against tampering, and more flexibility in questionnaire and report design. But field evaluations at major universities show that there are risks in collecting ratings online: (a) outside access to sensitive data may be easier; (b) response rates may be lower with online questionnaires; and (c) ratings may be less favorable.

Most University of Michigan teachers now collect ratings on paper questionnaires printed by the Teaching Questionnaire (TQ) program of the Office of Evaluations and Examinations (E&E). In pilot tests carried out in 2002 and 2004, however, TQ ratings were collected on the Web in several University departments. In these tests, performance of the Web-based TQ system fell slightly below current standards. Response rates were somewhat lower and ratings were slightly less favorable with Web-based data collection.

The central recommendation of this task force is that the University should implement a paperless, online rating system that reproduces, insofar as possible, features of the current TQ system. We also recommend the inclusion of communication and training as essential components of the new TQ system. The new system will work well only if students, teachers, and support staff understand it and know how to use it.

Other recommendations in this report address common concerns about online rating systems. Foremost among the concerns is the potential for outside access to sensitive data. For example, access to student comments is easier in online systems because online systems store student comments electronically, and electronic storage reduces the cost of Freedom of Information Act (FOIA) access to student comments. We are therefore recommending that comments not be retained in central electronic files after they have been distributed to instructors and designated individuals within departments and colleges. We are recommending that quantitative rating results continue to be archived in a central database.

In addition, lists of respondents are not kept in paper-based rating systems, but such lists must be kept in online systems to ensure that only registered students rate a class and that each student provides only one rating. We recommend that student identifiers be stripped from ratings and comments as soon as possible during data collection and that lists of respondents be deleted from the system at the end of the evaluation period. These lists should never be available to deans, department chairs, faculty, or staff members.

A second important concern in online rating systems is low response rates. Although it may be possible to improve response rates by using incentives and sanctions, we recommend against their use. Instead, we recommend the use of targeted announcements and frequent follow-up reminders for students during the period when evaluations are being collected. In addition, campus units that handle University orientation programs (e.g., the Center for Research on Learning and Teaching and the Office of New Student Programs) can help faculty, graduate instructors and students understand the importance of course evaluation at the University and the role that they play in such evaluations.

The third problem addressed in our recommendations is that online ratings are less favorable, on the average, than ratings made in classrooms. Our recommendation is that online results and in-class rating results be kept completely separate. Statistics such as medians and quartile points printed on online reports should be based only on results from online questionnaires. Furthermore, to reduce confusion between results of online and in-class questionnaires, E&E should discontinue its use of paper questionnaires. Setting up a two-track system of both paper and online evaluations would introduce even more complexity into the collection of data and interpretation of results.

These recommendations are based on our assessment of current risks and benefits of online rating systems. We relied on information now available to us but realize that the knowledge base for decision-making will grow rapidly after the University implements an online system. Our final recommendation is that the University monitor the performance of the new system carefully and make changes to the system when needed.

PART 1. THE CHALLENGE

COLLECTING STUDENT RATINGS of teaching is a well-established tradition in American colleges. It would be hard to find a major American college or university that does not use student ratings in the evaluation of teaching.[1] It would also be hard to find another topic in higher education that has been as extensively studied.[2]

A new idea shook this well-established area at the end of the century. The idea was to abandon the paper questionnaires on which ratings had been collected since the 1920s and to collect ratings instead via the Web. Students were registering for their college classes, viewing their course schedules, ordering their books, and paying their bills on the Web. Why should they have to rate their courses on paper questionnaires? Collecting ratings online seemed to be an idea whose time had come.

1.1 PROMISE AND PROBLEMS

It is easy to imagine a Web-based system of student ratings in which electronic communication replaces paper communication. Teachers and departments design and order their course evaluation questionnaires online instead of on paper. Students fill out questionnaires on websites rather than on scantrons. Teachers and departments receive rating summaries electronically rather than by campus mail. Colleges do not have to print hundreds of order forms, thousands of instructor reports, and hundreds of thousands of paper questionnaires each year.

One of the important benefits of a web-based rating system is efficiency. Web-based systems eliminate the time-consuming clerical work of traditional systems, including the printing, mailing, sorting, and scanning of hundreds of thousands of paper forms. Consequently, Web-based systems can be inexpensive to run. Further, because online reports can be generated as soon as the ratings have been collected for a class, teachers do not have to wait weeks for a paper summary to arrive.

Beyond efficiency, web-based rating systems promise more security for users. When students keyboard their comments about courses, they do not have to worry that teachers will guess their identities from handwriting clues. In addition, teachers and administrators can feel more confident about data handling. Many people have to handle paper evaluation forms (e.g., teachers, students who collect forms and deliver them to department offices, secretaries, campus mail personnel, evaluation center personnel), sometimes with little monitoring of those who handle the forms. Online systems provide fewer opportunities for tampering and loss of data.

An online format also opens up new possibilities in questionnaire and report design. For example, an online format frees questionnaire designers from the constraints of the printed page. Items do not have to fit on a single line or two; questionnaires do not have to fit on a single sheet; and everything does not have to appear in black-and-white. Response formats can vary, and the time of data collection can also be set to suit

individual courses. Similarly, online reports can be individualized to provide information to instructors. They can contain web links to appropriate comparison statistics or to specific sites that provide information about specific teaching problems.

However, some potential complications regarding online evaluations must be considered as well. One concern involves response rates. Students who are handed a paper questionnaire by their instructor on the final day of class and asked to fill it out are, in many respects, a captive audience. In contrast, those invited to fill out online questionnaires are potentially free to ignore requests for online ratings. This raises the question of how many students will actually submit online ratings. Another issue concerns the validity of online ratings. Students are free to fill out online questionnaires whenever they want, wherever they want, and with whomever they want. But do students fill out questionnaires responsibly under such uncontrolled conditions? Are online rating results valid?

In addition, electronic storage can change the ease of public access to sensitive data. For example, outside organizations wishing to access University documents using the Freedom-of-Information Act (FOIA) have not in the past had blanket access to comments written on paper questionnaires at the University in Michigan, in part because of the high cost in both time and money of filing requests for such comments. An organization filing a FOIA request would have to cover the cost of retrieving the original comment sheets from thousands of teachers and also the cost of copying hundreds of thousands of sheets. Costs go down dramatically when comments are made and stored electronically and centrally.

Thus, online rating systems in theory provide substantial benefits, but they also create possible risks. It is hard for college decision-makers to assess the benefits and risks without empirical evidence. Fortunately, field studies carried out in recent years provide concrete evidence of some of the benefits and risks.

1.2 FIRST EVALUATION STUDIES

Northwestern University and Brigham Young University were the first major universities to implement online rating systems. Northwestern began collecting online evaluations in the spring semester of 2000,[3] and Brigham Young University did so in the fall semester of 2002.[4] The two universities also carried out early assessments of the utility and practicality of online systems.

The Northwestern and Brigham Young researchers empirically confirmed some of the theoretical advantages of online rating systems. Brigham Young analysts, for example, reported that costs (including cost of development) associated with online systems were far less than the costs of traditional rating systems. Specifically, they calculated the total annual cost of the online system as $186,617, or $0.47 per evaluation form, compared to an annual cost of a paper-based system of $436,838, or $1.09 per evaluation form. In addition, Northwestern University researchers showed that students were more likely to make written comments on online questionnaires than on paper ones. More than three times as many students wrote online comments, and in total these students wrote five times as much as did the students who filled out paper questionnaires.

But Northwestern and Brigham Young researchers also found problems with online rating systems. In two analyses at Northwestern, online ratings were 0.25 points (on a 6-point scale) lower than in-class ratings. Thus, online ratings were less favorable than ratings on paper questionnaires. In addition, at both Northwestern University and Brigham Young University, online response rates were about 20 percentage points lower than response rates in classes using paper-based systems. Researchers at Brigham Young suggested, however, that teachers could influence response rates in online systems. The researchers found that response rates were higher when teachers made teaching evaluations a course assignment (whether or not the teachers gave points for completion of the assignment) than they were when teachers simply encouraged students to complete evaluations or did not mention teaching evaluations in their courses.

These pilot studies prepared the way for the development of full-scale online systems at Brigham Young and Northwestern. Both schools started out with small pilot studies of online ratings, then moved to mixed systems that included both online and paper questionnaires, and finally adopted systems that collect online ratings exclusively. Today, the systems at Brigham Young and Northwestern serve as models for other universities and colleges.

A Brigham Young website[5] currently lists 16 universities that collect online evaluations campus-wide: Bates College, Brigham Young University, Carnegie Mellon University, Georgia Institute of Technology, Hong Kong University of Science and Technology, Laval University, Northwestern University, Polytechnic University of Brooklyn, Smith College of Social Work, Tel Aviv University, University of Idaho, University of North Texas Health Science Center, University of Virginia, Wellesley College, Whitman College, and Yale University. The website lists 43 schools that collect online evaluations in one or more departments but not campus-wide, and it also lists 25 other schools that collect online evaluations in one or more courses but not department- or campus-wide.

Survey research indicates that online rating systems are even more widespread.[6] One survey found that 17 percent of schools were collecting online ratings by the end of 2002, another 10 per cent were planning to initiate online ratings in 2003, and another 18 per cent were reviewing options for collecting ratings online. In other words, nearly half of the surveyed schools were collecting online ratings or considering collecting them in 2003.

Thus, lessons can be drawn from the experiences of other schools with online ratings. However, few of the schools that now use online ratings are as large and diverse as the University of Michigan. And few of the schools had to replace a paper rating system as elaborate and individualized as the University of Michigan's system. The challenges at the University of Michigan are especially great.

1.3 STUDENT RATINGS AT THE UNIVERSITY OF MICHIGAN

When teachers think about course evaluation at the University of Michigan today, they usually think of the Teaching Questionnaire system (or TQ) administered by the Office of Evaluations and Examinations (E&E). This rating system gives faculty members a way to collect student opinions on questionnaires designed to fit their courses. Teachers and

departments design their questionnaires by selecting questions from a large, organized catalog. E&E enters the orders it receives from instructors and departments into a database management system, which custom-prints machine-readable questionnaires for each teacher. After students fill out the questionnaires, E&E scans the sheets, tabulates results, and returns printed reports to teachers and colleges.

The TQ program was developed at the Center for Research on Learning and Teaching (CRLT) in 1976 under the direction of CRLT Research Scientist James Kulik. The program was originally known as the Instructor-Designed Questionnaire (or IDQ) program, and instructors originally referred to the teaching questionnaires as IDQs (for Instructor-Designed Questionnaires) or simply as CRLTs. Both names eventually became anachronisms. After carrying out extensive computer reprogramming and making other revisions of the system, CRLT in 1996 gave the IDQ program its current name. In 1998, when CRLT Research Scientist Kulik became E&E's director, E&E became the new home for the TQ program.

The number of users of the TQ system has grown steadily for three decades. Last year, nearly 17,000 classes used TQs. The questionnaires were used school-wide in 15 of the University's 19 schools and colleges—all units except Art and Design, Business Administration, Dentistry, and the Medical School. TQs were also used in three schools on the Flint campus. In all, E&E printed nearly a half-million questionnaires last year.

The TQ system has grown not only in size but also in influence. When the TQ system was originally developed, it was used for a single purpose: to give faculty members information that they could use to improve their courses. Today, administrators consider ratings in annual reviews and in promotion and tenure cases; committees look at TQs when making teaching awards and preparing materials for accreditation reviews; faculty members use TQs in improving teaching and in mentoring and hiring GSIs; graduate students (and sometimes faculty members) include TQ results with job applications; and students consult TQ results when making course selections. Today, the TQ system truly serves many functions at the University.

1.4 PILOT STUDIES AT THE UNIVERSITY OF MICHIGAN

E&E carried out two pilot studies to determine whether the TQ system could be put online.[7] The first study examined ratings in two introductory-level courses in the undergraduate engineering program during the fall term of 2002. The second study examined ratings in 70 graduate-level courses offered in the economics, political science, and sociology departments during the winter term of 2005.

At the end of the term, in an effort to maximize response rates in the online evaluation conditions, E&E sent out an e-mail asking students to complete the ratings on the Web and then sent out e-mail follow-ups to those who did not respond to the initial request. The first e-mail reminder went out four days after the original request, and the second reminder went out two days later. E-mail reminders worked very well in the study with beginning engineering students. Response rates were nearly identical in the two conditions of the study: 74% for the online classes and 75% for the on-paper condition. E-mail reminders were not as effective in the study of graduate-level courses.

The average response rate was 65% in the online condition and 80% in the on-paper condition.

A number of factors might have produced the different findings in the two studies. The target populations in the two studies were different (i.e., beginning students vs. graduate students), and the subject matter in the two studies also differed (i.e., engineering vs. the social sciences). In addition, the undergraduate students in the first study were asked to fill out only one course evaluation form online, whereas the graduate students in the second study, who were enrolled in several graduate-level courses, were asked to fill out as many as six, seven, eight, or nine evaluation forms online. Some of the graduate students may also have been unsure of their responsibilities, because in some cases, the graduate students were enrolled in cross-listed courses under a number different from the one listed on the questionnaire. Whatever produced the difference in results in the two studies, the important point to note is that response rate was adequate, if not ideal, with online data collection: 74% in the first study and 65% in the second.

In both studies, ratings were less favorable on online questionnaires than on paper questionnaires. Online ratings averaged about 0.15 points lower than on-paper ratings, a small but statistically significant difference. In addition, a follow-up analysis suggested that the online ratings are as reliable as on-paper ratings. The analysis was carried out only in the engineering study, where it was possible to establish a criterion measure of teacher effectiveness for each of the teachers based on their ratings in the same class in other terms. The correlation between online ratings and the composite criterion was as high as the correlation between on-paper ratings and the criterion. For example, for the question on the overall excellence of the course, the correlation with the composite criterion was .74 for online ratings and .63 for on-paper ratings.

The likelihood of students writing comments on their rating forms was the same for online and on-paper conditions. A total of 61% of the students in the online condition and 63% of students in the on-paper condition wrote comments. The difference is small, and it is not statistically significant. The length of comments was also very similar for the two conditions. The average length of a written comment in the online condition was 52 words; the average length in the on-paper condition was 54 words. Again, the difference between conditions was small and not statistically significant.

The pilot tests at Michigan thus produced results that were encouraging but not definitive. The results suggested that Michigan's online system would have to be designed with great care. On the basis of these results, E&E proposed to the Senior Vice Provost for Academic Affairs that TQ be converted to an online rating system. The Vice Provost approved the start of the project, and in June 2006 E&E began working with MAIS to develop an online TQ system. The Vice Provost also appointed a task force on online placement examinations, and he asked the task force to develop recommendations that would ensure the successful collection of online ratings at Michigan.

PART 2. RECOMMENDATIONS

ALTHOUGH PILOT STUDIES have demonstrated the feasibility of collecting online ratings at the University, the studies do not tell us what to do next. Should the University move to an online rating system? If so, can it avoid problems often associated with such systems? For example, can the University maintain current response rates with a switch to online questionnaires? And how should the University plan for future developments in rating systems? We have reflected on the issues, and we are now making recommendations for action.

2.1 MOVING TQ ONLINE

After three decades, E&E's paper-based TQ system still meets many University needs. Teachers use it for course improvement; departments use it for promotion and hiring; students use it for course selection; and committees use it for award decisions. Printed order forms, printed summary reports, and hundreds of thousands of paper questionnaires flow back and forth each term between teachers and E&E with few problems. What does the University have to gain from switching to an online rating system?

Schools with online rating systems have listed a number of possible benefits. Compared to paper-based systems, online systems promise lower costs, timelier feedback for teachers, more anonymity for students, better safeguards against tampering, and more flexibility in questionnaire presentation. Five years after their introduction, online systems are thriving at the schools that pioneered in their development.

Although we are in favor of moving TQ online, we believe that an online TQ system should preserve all the important features of the current system. We envision a TQ system in which teachers and departments will still design forms, students will fill out familiar-looking questionnaires, and teachers and departments will receive the same kind of reports that they currently do. The main distinguishing mark of the new system will be the use of electronic communication as a replacement for paper communication.

E&E, MAIS, and CTools developers are already working to bring the current TQ system online, and they are in a good position to complete the job. E&E has guided the development of TQ from its birth in 1976 to its present form. MAIS has developed systems for handling the University's most sensitive financial, personnel, and academic records. CTools has provided secure communication links between teachers and students in thousands of University courses.

Recommendation #1: *E&E, working with MAIS and CTools developers, should implement a paperless, online version of the current TQ system. The online TQ system should, insofar as possible, incorporate the features of the current system while eliminating paper order forms, questionnaires, and reports.*

Writing new software is a major part of the task ahead, but more than software is needed for an online system to work properly. Teachers, administrators, and support staff need to know how online systems work and how to use the new software. Placing online orders, for example, is different from submitting paper orders. In the current TQ system, teachers and departments submit orders by campus mail, e-mail, fax, or phone, and E&E personnel enter the orders into its database. In the new online system, teachers and departments will enter their orders directly into a secure University database. E&E will no longer be a backstop. Teachers and departments will have to follow precise rules in placing orders.

E&E is the steward of the course rating data, and it should have the major responsibility for communicating basic information about the new TQ system to the University community. But MAIS also has an important role to play in communication and training. MAIS has many years of experience in developing workshops and tutorials on entry and retrieval of data from its databases. MAIS should use its expertise to write materials for training department support staff in entering data into and retrieving data from the new TQ database.

Recommendation #2: *E&E should communicate basic information about the online TQ system to the University community, using traditional print outlets, electronic communications (such as e-mail and website postings), and personal communications. MAIS should provide training for department staff members who will be responsible for entering departmental data into the new system.*

2.2 OVERCOMING COMMON PROBLEMS

Designers of online rating systems face major challenges. Our discussions focused on three of these: (a) outside access to sensitive data may be easier; (b) response rates may be lower with online questionnaires; and (c) ratings may be less favorable. We are recommending specific ways of meeting these challenges.

2.2.1 Easier Access to Sensitive Data

Students do not put their names or identifying numbers on paper questionnaires, and they usually feel comfortable with the assurance that their responses are anonymous. With online questionnaires, the situation is different. Students must log in to a website to provide their ratings, and to log in, they must provide their names or other identifying information. Names or other identifying information is needed so that the computer can check eligibility to make ratings. While data are being collected from a student, the student name is attached to student responses.

We consider it essential, however, that a student's ratings and comments be decoupled from the student's name or ID as soon as possible after the student logs off. The Family Educational Right to Privacy Act (FERPA) would prevent outside access to the name of a student who provided ratings, but students need more than this protection.

It should be impossible for instructors, department chairs, deans, or staff members from finding out the names of student who made specific ratings or comments. Decoupling student ratings and comments from student names and identifiers is the best way to ensure the anonymity of ratings and comments.

Although decoupling can be carried out immediately after a student logs off, the system has to create a list of students who have and have not responded, and this list must be maintained for the full evaluation period. The list is necessary for at least two reasons. First, the system must know which students have already submitted ratings. Without a list of respondents, the system could not ensure that each enrolled student submitted only one set of ratings for a course. Second, the system must send follow-up reminders to students who delay submitting their evaluations. Without a respondent list, follow-up reminders would have to go out to both responders and non-responders.

It is important that this list be destroyed at the end of the evaluation period. Instructors, department chairs, deans, and staff members should have no way of knowing which students responded to a course evaluation request and which did not. CTools and MAIS have experience in protecting the anonymity of respondents, and they should be able to guarantee anonymity by decoupling student names from their responses and destroying lists of respondents as soon as the system can.

Recommendation #3. *CTools and MAIS developers should develop mechanisms for stripping student identifiers from ratings and comments as soon as possible during data collection and for deleting lists of student respondents at the end of the evaluation period. It must be impossible for deans, department chairs, faculty, and staff members to identify the students who provided evaluations and to associate specific ratings and comments with specific students.*

Michigan law requires the University to release employee information requested under the Freedom-of-Information Act (FOIA) as long as the information does not fall into a protected category. Teaching evaluation results to do not fall into a protected category, and so the University has to release such results—both quantitative ratings and student comments—to anyone who is willing to reimburse the University for preparing the data for release. A switch from paper to electronic storage of information would change the cost of filling data requests, and it might result in easier access to evaluation results through FOIA requests.

Both the Michigan Student Assembly (MSA) and the commercial organization Pick-a-Prof have already obtained access to quantitative TQ rating results through FOIA requests, and quantitative TQ results now appear on both MSA and Pick-A-Prof websites. However, E&E has not had to release copies of student comments to these groups, and student comments about University teachers and courses do not appear on the Web.

Three factors have protected E&E from having to release student comments. First, students write their comments on the back of TQ forms, and the University would have to prepare typed or Xeroxed copies of the comments before releasing them in response to a FOIA request. With a half-million evaluation sheets collected each year, the cost of copying the comment sheets would be significant. Second, student comments sometimes contain information from which student identity can be inferred. To comply with the

Family Educational Right to Privacy Act (FERPA), the University would have to remove all information that might identify the writer of a comment before releasing the comments. The group making the FOIA request would have to cover the cost of redaction. Third, copies of written student comments are not kept in a central repository. E&E instead returns the sheets containing the comments to instructors and departments after tabulating the rating data. To respond to a blanket FOIA request for student comments, the University would have to contact all the individuals (thousands each semester) who may have the handwritten comments in their possession. Again, the cost of rounding up and copying all the comments would be high.

Together, these three factors make the cost of FOIA requests for student comments prohibitive for most groups. But the situation might be different if students entered their comments directly into electronic files. The cost of copying electronic files would be negligible in an online system, and the cost of tracking down comments would also be negligible if the electronic files containing the comments were retained centrally. The only real cost to a FOIA requester would be the University's charge for redacting the comments.

We considered two options for handling comments. One option is for MAIS to destroy all files containing student comments as soon as summary reports are prepared. The second option is to retain copies of the comments in MAIS warehouse tables. We are recommending the first option, because it would prevent FOIA requesters from obtaining access to student comments through E&E or MAIS. Instructors and departments, however, would have the same FOIA responsibilities that they currently have.

Recommendation #4. *E&E and MAIS should not retain copies of student comments in a data warehouse after the comments have been forwarded to instructors. E&E and MAIS should maintain an archive of quantitative data only.*

If E&E and MAIS do not retain copies of student comments, who should receive copies of the comments? Currently, E&E returns copies of comments to departments for distribution to instructors. Some departments make copies of the student comments for department files before distributing the comments; others do not. Some GSI supervisors review student comments about GSIs before distributing them; other supervisors do not.

In the proposed online system, instructors will receive the same reports that they currently receive, but the reports will be in electronic rather than printed form. The reports will contain: (a) a summary of the quantitative rating data; and (b) images of rating sheets that contain the comments made by each student. Departments will also receive summaries of quantitative ratings, but should they also receive copies of the comments? There are three alternatives: (a) only instructors receive copies of the comment sheets; (b) department chairs also receive copies; and (c) department chairs receive copies of the comment sheets but only if department chairs have made prior arrangement to receive these comments.

We believe that the third option comes closest to present practice. It means that E&E and MAIS would have to develop formal ways for departments to indicate their

preferred ways of handling student comments. The FOIA responsibilities of departments and instructors would remain the same as they currently are. Instructors and departments that retain copies of student comments would have to release copies of these comments—after redaction to remove student identifying information—in response to a FOIA request.

Recommendation #5. *Deans and department chairs, in consultation with their faculties, should determine who will receive copies of rating results and who will receive copies of student comments. E&E and MAIS should collect this information from departments and colleges and use it in routing comments to instructors and departments.*

2.2.2 Lower Response Rates

Most schools that collect online ratings consider response rates to be their biggest challenge. In early evaluation studies, response rates were about 20 percentage points lower in online systems than in paper-based ones. Response rates for online questionnaires averaged around 60%; response rates to paper questionnaires averaged around 80%.

Schools have tried to improve on such results by using rewards and sanctions—the carrot and the stick. Some have set up simple reward programs that give extra points or raffle tickets to fill out students who fill out evaluations. Other schools have used more sophisticated rewards. For example, students who complete online ratings at Northwestern University receive access to the ratings and comments about Northwestern courses made by their fellow students. Still other schools use sanctions to increase response rates. One approach is to withhold a student's grade until the student submits his or her online evaluations. Although this sanction may seem severe, it appears to be very effective. Schools that use this approach have reported response rates approaching 100%.

We prefer approaches that use neither rewards nor sanctions but instead stress direct and frequent communication with students. E&E can play a role in this communication by reminding students of their evaluation responsibilities in initial email requests and in follow-up reminders to those who do not respond promptly to initial requests. E&E used this procedure in its pilot test of online evaluation in Engineering, where it worked very well. Response rates were 74% for the students filling out online evaluations and 75% for those filling out in-class evaluations. Results were less promising, however, in a second, less well-controlled pilot test carried out in graduate LS&A courses.

Recommendation #6. *To promote a high response rate to online questionnaires, E&E should send announcements and follow-up reminders to students during the period when evaluations are being collected. E&E should also monitor response rates and adjust its strategy for promoting response rates if necessary.*

E&E reminders alone may not solve the response-rate problem. Common sense and research suggest that student attitudes are a critical determinants of response rates and that teachers can affect attitudes toward ratings. Research has shown, for example, that response rates go up when completion of an online questionnaire is listed as a course assignment, even when no points are awarded for completing the assignment. Thus, when teachers treat completion of evaluations as an essential course element and students accept this as a course responsibility, response rates are high.

The University is fortunate in having outstanding programs for orienting faculty and students to the campus and its culture. The Center for Research on Learning and Teaching (CRLT) conducts orientations for new University teachers and also offers other development opportunities for faculty and GSIs. The Office of New Student Programs (ONSP) orients all incoming students to the campus. The programs of both units can, and should in our view, include information about the role that course evaluation plays at the University and the role that students and teachers play in course evaluation.

Recommendation #7. *CRLT should include in its orientation and development programs information about faculty responsibilities with online evaluation systems. ONSP should include in its orientation programs information about student responsibilities with such systems.*

2.2.3 Less Favorable Ratings

In each of the four departments in which E&E carried out pilot tests, online ratings were lower than ratings collected on paper questionnaires. The difference between ratings collected under the two conditions averaged between 0.1 and 0.2 points lower on a 5-point scale. When some teachers use online questionnaires and others use paper questionnaires to collect ratings, those using the online questionnaires are at a small disadvantage.

When everyone uses the same method of data collection, this bias disappears. Ratings may fall slightly with a switch to online data collection, but they should fall equally for all. Relative standings should not change. TQ summary reports contain comparative statistics that makes relative comparisons feasible, and the TQ interpretive guide stresses the importance of interpreting results in relative terms. The relative comparisons will be just as accurate in an online system as in a paper-based system, providing that the comparative statistics for the online system are based on data from online questionnaires alone.

It is easy to imagine a two-track evaluation system that provides online questionnaires for some teachers and paper questionnaires for others. A two-track system would increase instructor and department choices. But this kind of system would also introduce more complexity into the collection of data and interpretation of results. It would be necessary to set up separate database systems with separate norms, and it would be necessary to keep results from the two systems distinct.

To reduce confusion between online and in-class ratings, we recommend that the University discontinue its use of the paper-based TQ system as soon as an online system is up and running. The online system should be a replacement for, not a supplement to, the current TQ system. In addition, we recommend that E&E calculate the comparison statistics for online reports from online questionnaire data only. Mixing online ratings and in-class ratings would be unfair to users of the online rating system.

Recommendation #8. *To reduce confusion between results of online and in-class questionnaires, E&E should discontinue its use of paper questionnaires and print statistics (i.e., medians and quartile points) on online reports that are based on results from online questionnaires only.*

2.3 PLANNING FOR THE FUTURE

We are making these recommendations based on our assessment of current risks and benefits of online ratings. Our assessment uses only information available to us right now. The knowledge base for decision-making will grow rapidly after the University implements online ratings. We will then have a clearer idea of what works and what does not—what needs tweaking and what needs re-thinking. This taskforce will not be meeting when the new information comes in, and it will be unable to review the information. But the University should take steps to make sure that the information is reviewed and acted on.

It is also certain that communication technologies will change in the future. For example, the day may not be far away when every student in every UM classroom has a wireless connection to the Internet via laptop, phone, or some other device. When that day arrives, it will no longer be necessary to distinguish between online and in-class evaluations. Students will be able to fill out online evaluations in class, or in-class evaluations on line. Special procedures for encouraging high response rates may become irrelevant. And the difference in favorability of online and in-class ratings will no longer be an issue. The University also needs to make sure that it is monitoring developments in technology and online rating systems and capitalizing on them.

We believe that the University needs to appoint an advisory committee to guide the development of the online TQ system during the next crucial years. The task of the advisory committee would be to monitor performance of the system, review technology developments that can affect collection of online data, and to make recommendations for adjustment and change.

Recommendation #9: *The Provost's Office should appoint an advisory committee to guide the development of an online TQ system during its first years of implementation. The committee's task would be to monitor performance of the online rating system, review technology developments that can affect collection of online data, and to make recommendations for adjustment and change.*

* * * *

We look forward to further discussions of the merits of what we have recommended, and we intend to participate fully in those discussions.

NOTES

1 First collected during the 1920s, student ratings were reportedly collected in 29 per cent of colleges in 1973, 68 per cent in 1983, and 86 per cent in 1993. See Peter Seldin, "The use and abuse of student ratings of professors," *The Chronicle of Higher Education*, July 21, 1993, p. A40. A survey in 1999 of the 200 "most wired" colleges in the country located only six schools that did not require the collection of student ratings. See K. Hmielski, "Barriers to online evaluation: Surveying the nation's top 200 most wired colleges" (Troy, NY: Interactive and Distance Education Assessment Laboratory, Rensselaer Polytechnic Institute, 2000).

2 Rating experts Michael Theall and Jennifer Franklin in 1990 concluded that research on student ratings was the largest single research area in postsecondary education. See Michael Theall and Jennifer Franklin, "Student ratings in the context of complex evaluation systems," in M. Theall and J. Franklin (eds.), *Student Ratings of Instruction: Issues for Improving Practice*, New Directions for Teaching and Learning, no. 43 (San Francisco: Jossey-Bass, 1990).

3 Hardy, N. (2003) Online ratings: Fact and fiction. In D. L. Sorenson and T. D. Johnson (Eds.), *New Directions for Teaching and Learning, no. 96.* San Francisco: Jossey-Bass. Pp.31-38.

4 Johnson, T. D. (2003) Online student ratings: Will students respond? In D. L. Sorenson and T. D. Johnson (Eds.), *New Directions for Teaching and Learning, no. 96.* San Francisco: Jossey-Bass. Pp. 49-60.

5 Bothell, T. W., and Henderson, T. (2003) Do Online Ratings of Instruction Make $ense? In D. L. Sorenson and T. D. Johnson (Eds.), *New Directions for Teaching and Learning, no. 96.* San Francisco: Jossey-Bass. Pp. 69-80.

6 Hoffman, K. W. (2003) Online Course Evaluation and Reporting in Higher Education. In D. L. Sorenson and T. D. Johnson (Eds.), *New Directions for Teaching and Learning, no. 96.* San Francisco: Jossey-Bass. Pp. 25-29.

7 Kulik, J. A. *Online Collection of Student Evaluations of Teaching.* Ann Arbor, MI: Office of Evaluations and Examinations, Dec. 2005. (Unpublished report).

APPENDIX: BACKGROUND PAPERS BY TASK FORCE MEMBERS

Online Collection of Student Evaluations of Teaching. James A. Kulik, Office of Evaluations and Examinations, December 5, 2005

Changes in Student Ratings of Teaching at the University of Michigan During 25 Years, James A. Kulik, Office of Evaluations and Examinations, August 29, 2006

Collecting Student Evaluations Online: A Proposal, James A. Kulik, Office of Evaluations and Examinations, September 25, 2006

Evaluation Processes, Lisa Emery, Michigan Administrative Information Services, April 16, 2007