

JOINT SENATE AND OFFICE OF ACADEMIC AFFAIRS
TASK FORCE ON COURSE AND TEACHING EVALUATION

FINAL REPORT

APRIL 21, 2005

Executive Summary

This final report contains seven recommendations on how the academic community can enhance its capabilities to assess and improve curriculum and instruction. The Task Force was charged with addressing a major question, among others, of whether there should be a university-wide requirement for student evaluations of teaching. The Task Force members unanimously agree that a university-wide course evaluation requirement and system be adopted. This document addresses that question and others with relevant recommendations and rationale for how this can be achieved. The Task Force Report is divided into the following sections:

- I. Introduction
- II. The Process
- III. Task Force Recommendations
- IV. Soliciting Input from the Academic and Student Communities
- V. Implementation Plan
- VI. Conclusion
- VII. Appendices

Section I: Introduction

The Senate Executive Committee and the Provost jointly charged this Task Force (a) with determining whether the campus should adopt a University-wide requirement for student evaluations of teaching, (b) what the criteria and process for implementation should be if the Task Force were to recommend a standard evaluation instrument, and (c) how the campus might use both a standard instrument and other forms of teaching evaluation to meet multiple needs. Appendix A lists the members of the Task Force.

After extensive study of processes and instruments used at peer institutions and reviewed in the literature, the Task Force does recommend that the campus adopt a University-wide student evaluation instrument for undergraduate and graduate courses. Research completed over the past two decades suggests that student evaluations are psychometrically sound and positively related to a number of indicators of teaching effectiveness. Research clearly shows that students' evaluations of teaching, regardless of how the questions are worded, are an evaluation of the instructor. In other words, the variance in student evaluations of teaching is predominantly at the instructor level. There are, however, different aspects of the instructor's approach that can be evaluated. The universal form recommended by the Task Force contains items that refer to the instructor's effectiveness and process, student outcomes, and student questions.

The Task Force further recommends that the instrument contain a set of universal questions, as well as optional ones added by colleges, departments or programs, or instructors. A universal set

of questions will improve the consistency of course evaluations. The Task Force recommends that the Internet be used, both to administer the instrument and to make some of the information available to students. Finally, the Task Force suggests additional modes of teaching evaluation and the ways in which the campus might use these multiple indicators of teaching effectiveness.

A summary of the research collected and used by the Task Force members is found in Appendix B. The following sections elaborate the recommendations of the Task Force.

Section II: The Process

The Task Force began its work in the Summer of 2003 and met through the Fall of 2004. The Task Force took a broad and comprehensive approach to dealing with its charge. The members of the Task Force reviewed relevant literature on student evaluation of teaching; examined evaluations currently being used at UM; reviewed evaluation instruments used at peer institutions and examined the evaluation policies and procedures at these institutions; previewed the College of Computer, Mathematical, and Physical Sciences Web CT course evaluation pilot; and spoke to informed sources on campus.

Section III: Task Force Recommendations

The Task Force proposes seven recommendations in response to its charge from the Provost and the Senate Executive Committee:

Recommendation 1: For the purposes of aiding the improvement of the quality of teaching and to assure consistency and comparability of information obtained through student evaluations of teaching, we recommend that there be a university-wide requirement for student evaluations in all undergraduate and graduate courses.

The University Senate at its May 3, 2004 meeting acted upon this first recommendation and adopted the following resolution: "We recommend that there be a university-wide requirement for student evaluations in all undergraduate and graduate courses." Appendix C contains the full Senate resolution. The Task Force was then directed to return to the other remaining questions presented in the original charge and to provide its recommendations in a full report. The following recommendations pertain to the requirement of evaluations in all undergraduate and graduate courses:

- A. The Task Force recommends that the campus adopt a universal set of questions that examines: (a) the process of teaching, (b) the outcomes attributable to the specific course, and (c) provides information for students that can be used for course selection. Appendix D contains the recommended universal course evaluation questions.
- B. The Task Force members recommend that a set of universal questions be used across campus (see attached passed Senate resolution in Appendix C). This small set of universal questions would be used in all undergraduate and graduate level classes.

Colleges, departments and instructors can elect to add additional questions to these universal questions (Appendix D).

- C. The Task Force members recommend that decisions about exempting educational activities from the course evaluation process that do not conform to the standard definition of a course be left at the College and/or departmental level. Each Dean will establish criteria to determine which types of courses should be excluded from campus course evaluation requirements. These decisions are best made by those who have a more intimate knowledge of the nature of a non-standard educational activity. Another consideration for exclusion from this requirement might be course enrollment (e.g., small classes).
- D. Another major issue to be addressed in the student evaluation of teaching is the process of administering these evaluations. Currently, the process of administration varies dramatically across departments and colleges. Task Force members agreed that there should be a uniform administration process for student course evaluation. The Task Force members recommend that the universal questions described in Appendix D be administered using a web-based system. Web administration would lead to: 1) comparison data; 2) cost savings and consistency in administration, and efficiency in data collection, analysis and reporting; and 3) the opportunity to gather data across all colleges in compliance with the new Middle States standard on measuring student learning outcomes.
- E. Because of the differences across the various Colleges on campus, the Task Force members recommend that colleges collect and make available normative information. Further, because research indicates that class level is related to course evaluation, Colleges should maintain this normative information by course level (e.g., 100 & 200 level, 300 & 400 level, and graduate level). Colleges should decide if this normative information should be broken down to the department level. This normative information should, at a minimum, include means, standard deviations, and confidence intervals for each of the recommended universal questions.
- F. To allow flexibility for instructors, programs, departments, and colleges to assess important dimensions of teaching and learning not covered in the universally administered questions, the Task Force recommends that the course evaluation have optional questions available for specific users.

Recommendation 2: The Task Force recommends that the academic community educate students about the importance of providing formal feedback on course evaluations.

Students, early in their college careers, need to be aware that their feedback about their experiences in academic courses is important and valued. Part of this educational process also includes helping students to take responsibility for their communication. This is an essential component of the evaluation process given the anonymity of students' responses. The greatest value from evaluations will be realized by students seeking to objectively and fairly provide feedback. The anonymity of students' responses does not

justify comments that can be considered offensive speech. Faculty and instructors might consider adding a statement on their course syllabi about the importance of students' feedback and expectations for the quality of students' comments. Students must be held to high expectations by the university community for the quality of their written communication on course evaluations.

Recommendation 3: The Task Force recommends the following four purposes for university-wide student evaluations of teaching:

- a. Formative evaluation: to provide diagnostic feedback to faculty for the improvement of teaching.
- b. Summative evaluation: to provide one measure of teaching effectiveness for use in the APT and post tenure review processes and in annual productivity reviews.
- c. Informative evaluation: to provide information to students for their use in the selection of courses and instructors.
- d. Outcome evaluation: for the purposes of documenting student learning.

Based on a review of the empirical literature, the Task Force members agree with Marsh (2001) that students' evaluation of teaching "are primarily due to the instructor who teaches a class and not the particular class being taught." Therefore, from a measurement perspective, what can be measured is the teacher and not the course. (Appendix B).

Recommendation 4: It is recommended that advisory criteria and guidelines be developed that might be helpful to academic units as they design their separate course evaluation procedures. These guidelines would highlight exemplary models of evaluations, including both quantitative and qualitative measurements.

The literature clearly shows that teaching needs to be evaluated along multiple dimensions. In developing the recommended questions, the Task Force members considered both the process of teaching and the type of learning outcomes that could and should be examined. In addition, the Task Force members considered the type of information that students may find useful in considering course and instructor selection. Our approach was to develop questions that specifically addressed each of these areas of student evaluation. This proposed list is not exhaustive and is intended to be succinct for data management purposes while allowing room for colleges and departments to add additional questions. It is recommended that customization of evaluation questions occur at three levels: university-wide, college-wide, and at the individual course level. The College of Behavioral and Social Sciences (BSOS) has an exemplary model for course evaluation customization.

In deciding on the teaching process questions, Task Force members focused on dimensions of teaching that were empirically identified and confirmed through factor analytic studies. These dimensions include: Instructor Enthusiasm, Organization and Clarity, Group Interaction, Individual Rapport, Breadth of Coverage, Examination and Grading Procedures, Assignments and Readings and Workload and Difficulty (Marsh, 2001). There are of course other aspects of teaching that could be considered and numerous evaluation questions that

have been used. A comprehensive list of additional evaluation questions has been compiled by BSOS.

In determining the student learning outcomes questions, the Task Force members examined the "University of Maryland Mission Statement" and "Building on Excellence: The Next Steps" to identify student learning goals that could be tied directly to evaluation items. The principle was to tie learning outcome goals to specific questions on the student assessment questionnaire. Because the recommended instrument is for universal use, the goals identified and the related questions are global in nature. Programs, departments, and colleges may want to identify more specific student learning goals and outcome questions directly tied to local needs. The student evaluation questionnaire developed by the School of Engineering is a good example of how student learning outcomes identified in accreditation materials can be tied to specific questions on student course evaluations. The data collected on the learning outcomes questions can be used, in part, to satisfy the new Middle States accreditation requirement on documenting student learning outcomes.

Recommendation 5: To execute a universal evaluation questionnaire, the campus should move to a web-based course evaluation system for efficiency and cost savings purposes. The Task Force members also recommend that the University administration explore mechanisms to link course evaluation to grade release as a way to improve response rates. The computing and programming capacity may need to be enhanced to administer web-based evaluations and to make the necessary administrative links to systems such as registration.

Web-based forms have the advantage of simplifying the collection and analysis of student evaluations. Response rates to web-based course evaluations have typically been moderate to very low. Some departments have used extra credit as an incentive for students to complete evaluations, yet the response rates in some cases remain relatively low. Regardless, the Task Force does not endorse this strategy.

Because several Colleges are currently using or plan to implement web-based administrations, the response rate issue is a critical one that demands a campus-level response. There is clearly interest in moving in the direction of web-based administration. We recommend that the Office of Information Technology (OIT) and the Office of Institutional Research and Planning (OIRP) work together to design centralized data collection and dissemination processes. It is further recommended that OIRP maintain the data once it has been collected and be responsible for implementing mechanisms to secure the evaluation data and limit access to authorized University users (students for the student questions, faculty and their respective College administration personnel for the non-student questions).

Our research suggests that the web-based student evaluation of teaching systems that have very good response rates (approaching 100%) also link course evaluation to other administrative systems. For example, course evaluations can be linked to delayed course registration or delayed grade release. In these systems, students cannot register for classes or receive their grades until they visit the course evaluation web site (students can opt to not evaluate a class or classes by visiting the evaluation web site and electing not to evaluate).

Currently, the response rate of web-based course evaluations in CMPS is low (less than 40%) without using a tie to withholding grades for students who do not complete their evaluations.

Recommendation 6: The Task Force recommends making some portion of course evaluation results available to students.

The Task Force members propose that a portion of the results of the student evaluation of teaching be made available to students. The information that is obtained could help students to better select courses in which they are likely to succeed. This also will create the incentive necessary for students to complete the evaluations online, which will add to their validity. The majority of our aspirational peer institutions have policies permitting the limited sharing of results with students (UNC-Chapel Hill, UI-Urbana-Champaign, and U. Michigan).

The Task Force members developed questions that provide useful information to students in course and instructor selection. The results of these questions should be made available to students. Students would not necessarily have access to the results of questions designed for other purposes (See questions 1-8 in Appendix D). An evaluation summary from individual courses will be made available, but course comparison data will not be publicized. The data from the student questions would be made available to students.

OIT and OIRP will design systems to provide access to the results of student questions on a centralized administration-maintained web site and will secure evaluation data. Information from other sections of the evaluation instrument would be maintained internally for use in instructional improvement and evaluation.

Recommendation 7: For summative evaluation purposes (the APT and post tenure review processes and in annual productivity reviews), the Task Force members recommend that student evaluation of teaching be one of multiple measures of teaching effectiveness.

Other measures of teaching effectiveness could include peer-review of teaching, review of syllabi and teaching methods. For formative evaluation (diagnostic feedback to faculty for the improvement of teaching), attaining evaluation at the end of the semester is not ideal for course modification and teaching improvement. Mid-semester or more frequent evaluations are ideal for purposes of improvement of instruction.

The Task Force members recommend that the university make available an easily administered, empirically-supported student evaluation of teaching that instructors could use at any point during the semester. For example, Penn State University has a web-based copy of Marsh's (1982) Students' Evaluation of Educational Quality (SEEQ) instrument (this instrument is free) that instructors can ask students to fill out at any point during the semester. Results of the SEEQ ratings are available to the instructor before the next class period.

IV: Soliciting Input from the Academic and Student Communities

Members of the Task Force represent a small segment of the campus community. Therefore, the recommendations contained in this report should be vetted widely before a formal Senate vote on possible adoption is addressed. We propose that the recommendations contained herein be made available through a Web site where input from students and faculty can be solicited. The use of web-based feedback for the proposed university-wide evaluation questions and process will ensure greater inclusivity of the academic and student communities. In addition, the recommendations contained in this report could be transmitted by the Provost through the Council of Deans and UPAC to solicit broader administrative input.

V. Implementation Plan

The Task Force recommends the following implementation schedule to the Provost and the Senate Executive Committee (SEC).

April 2005

The Task Force solicits feedback from the academic and student communities on the proposed universal evaluation questions. The Task Force recommends using a web-based process to collect feedback and hosting an open forum.

April 2005

The Task Force incorporates feedback and presents its final report to the SEC.

Spring Semester 2005-Summer 2005

The Provost charges a small Implementation Committee of OIT, OIRP, and representatives from the Deans' Council to examine course evaluation data collection methods and how to best secure the data. This group will formulate reporting methods for the course evaluation data. For example, scores could be reported in bands (4.5-5.0) as one possible method.

Fall Semester 2005

Colleges prepare faculty and instructors for course evaluation process

Spring Semester 2006

Pilot new course evaluation system with select colleges

Fall Semester 2006

Implement universal evaluation system in all colleges

VI. Conclusion

The Task Force members applaud the University Senate in its bold measure to adopt a resolution requiring student evaluations of all undergraduate and graduate courses. Expanding on this commitment to include student access to evaluation data and a universal set of evaluation questions will strengthen the University's commitment to excellence in teaching and learning.

Appendix A

JOINT TASK FORCE ON COURSE AND TEACHING EVALUATIONS

Revised Committee List 10-26-04

Professor Dennis Kivlighan, Task Force Chair

Professor & Chair

Counseling and Personnel Services

Professor David Bigio

Associate Professor

Mechanical Engineering

Professor Paul Hanges

Professor

Psychology

Dr. Nance Lucas

Special Assistant to the Provost

Office of the Senior Vice President for Academic Affairs
and Provost

Professor Phyllis A. Peres

Associate Professor & Associate Dean

Undergraduate Studies

Jordan Rothberg

Senate Executive Committee Student Representative

Darren Schneider

Graduate Student

School of Public Affairs

Professor Nancy Struna

Professor

American Studies

Dr. Kaci Thompson

Director, Undergraduate Research and Internship Programs

College of Life Sciences

Professor Scott Wolpert

Professor & Associate Dean of Undergraduate Studies/CMPS

2130 Mitchell Building

Appendix B

Research on the Evaluation of Teaching

Background

Cashin (1990) estimated that there have been more than 1,300 articles and books about student ratings of teaching, including such topics as methods of evaluation, reliability, validity, and sources of bias. In general, input from students is recognized as an essential component of a comprehensive system for evaluating teaching effectiveness (Cohen, 1982; Harris, 1982). The vast majority of those who evaluate teaching use quantitative student ratings (Braskamp, Caulley, & Cosin, 1979; Centra, 1980; Sledin, 1978; Shapiro, 1990). For example, in Centra (1980) more than 80 percent of those universities surveyed used student ratings.

Reliability of student ratings

Student ratings have acceptable levels of reliability. More specifically, students tend to agree about their description of the course instructor. While correlation between any two students in the same class is typically in the .20s (Marsh, 1987), Marsh and Roche (1997) indicate that the reliability of the average rating an instructor quickly reaches acceptable levels as the number of students in the class increases (i.e., .60 for 5 students, .74 for 10 students, .90 for 25 students). Even in relatively small classes, the reliability of the average course ratings meets traditional standards for reliability (Nunnally & Bernstein, 1992). For small to moderate sized courses (i.e., 25 or greater) the reliability of the average student ratings compares very favorably with the best objective tests (March & Roche, 1997).

The stability of student ratings appears acceptable with the test-retest reliability ranging between the mid .60s to the low .80s for time intervals of a few days to a few months (Pritchard, Watson, Kelly, & Paquin, 1998). Further, when end-of-course ratings are compared to ratings of the same instructor one year later, correlations are in the .80s (Guthrie, 1954; Overall & Marsh, 1980). There is even some evidence indicating that different students in the same course by an instructor will agree in their rating of the instructor over a seven-year period (Hanges, Schneider, & Niles, 1990).

In terms of internal consistency reliability, student ratings have ranged from .70s to the .90s (Doyle, 1975; Hanges, Schneider, & Niles, 1990; Marsh, 1987, Runco & Thurston, 1987; Shingles, 1977). In summary, the data suggest that overall, student evaluations show good reliability.

Validity: Comparison of student ratings with other measures

Research suggests that student ratings are multidimensional and that effective evaluation instruments should tap multiple dimensions of teaching effectiveness (Marsh, 2001). Research also indicates that there are sizable relationships between student ratings and other indicators of effective teaching. For example, Cohen (1987) conducted a meta-analysis of multi-section validity studies (i.e., multiple sections of the same course taught by different teachers) comparing

student ratings of teacher effectiveness to student learning on standardized final exams. He found sizable correlations between certain dimensions of student ratings and student achievement (e.g., 0.52 for interaction, .50 for instructor skill, .49 for overall course). Other researchers have examined the relationship between student ratings and instructor self-ratings. Feldman's (1989) meta-analysis found average correlations between .15 and .42 for average student ratings on specific dimensions of teacher effectiveness and instructor self-ratings. In their review of the student ratings literature, Pritchard et al. (1998) reported that the correlations between student ratings and instructor self-ratings ranged from .30 to .40.

Potential Biases

Many variables have been examined as possible sources of bias in student ratings. When reviewing the literature on bias, it is important to interpret the findings carefully. Just because two variables exhibit a positive correlation (e.g., class size and average ratings of instructor rapport) should not be taken as automatic evidence of bias. A reasonable argument can be made that these variables should be related and do not reflect bias.

According to Pritchard et al. (1998), the following factors do not appear to affect average student ratings:

- instructor and student gender;
- student age;
- student academic level (e.g., freshman, sophomore, etc.)
- time survey administration (i.e., middle or end of semester; time of day)

According to Pritchard et al. (1998), the following factors do appear to bias student ratings:

- lack of anonymity for raters;
- instructor's presence during ratings;
- informing students that ratings are used for personnel decisions.

Finally, according to Pritchard et al. (1998), the following factors are related to student ratings but it is unclear whether these relationships should be considered evidence of bias:

- student interest in the course;
- instructor age and teaching experience;
- class size and course level;
- grades in course.

Marsh (1987) reviewed several studies concerning the relationship between numerous background characteristics (e.g., prior student interest in the course, class size, etc) and student evaluations. In two of the studies reviewed by Marsh (1987), 16 background characteristics accounted for approximately 13% of the variance in the average student ratings. Of these background characteristics, only four accounted most of this explained variance. Specifically, higher average student ratings were correlated with:

- higher prior student interest in the course,
- higher expected grades,
- higher levels of workload/difficulty,
- higher percentage of students taking the course for general interest only.

Further analyses revealed that, of these four background variables, prior student interest had the strongest impact on student ratings.

Conclusion

The research suggests that students' ratings of teaching are psychometrically sound. Student ratings have acceptable to good indices of inter-rater reliability, test-retest stability and internal consistency reliability. In terms of construct validity, research suggests that student evaluations of teaching are multidimensional and that evaluation instruments should be designed to measure separate components of teaching effectiveness (Marsh, 2001). In addition course evaluations are related to a number of indicators of teaching effectiveness including: student achievement, faculty self-evaluations, and ratings of trained observers. Finally, student ratings tend to be relatively but not completely free of bias. The largest source of potential bias comes from students' prior interest in the subject area.

Appendix C

University Senate Resolution

A RESOLUTION ON THE USE OF A UNIVERSAL SET OF QUESTIONS TO BE USED IN STUDENTS' EVALUATIONS OF TEACHING

University Senate of the University of Maryland, College Park

BE IT RESOLVED that the University Senate, on behalf of the University of Maryland community adopts a set of questions (see attached) to be used in the university-wide requirement for student evaluations in all undergraduate and graduate courses.

The University Senate at its May 3, 2004 meeting adopted the following resolution: "we recommend that there be a university-wide requirement for student evaluations in all undergraduate and graduate courses." The present resolution builds on the resolution adopted May 3, 2004. A universal set of questions allows for consistency and comparability in the evaluation process.

A RESOLUTION TO MAKE AVAILABLE TO STUDENTS THE INFORMATION CONTAINED IN THE 'STUDENT QUESTIONS' SECTION OF THE UNIVERSAL SET OF QUESTIONS USED IN STUDENTS' EVALUATIONS OF TEACHING

University Senate of the University of Maryland, College Park

BE IT RESOLVED that the University Senate (through the various Colleges), on behalf of the University of Maryland community will make available to student's courses evaluation information contained in the "Student Questions" section of the universal set of questions used in students' evaluations of teaching.

Following the student evaluation practices of peer institutions, the Task Force members recommend that information contained in the "Student Questions" section of the teaching evaluation instrument be made available to students. Information from other sections of the evaluation instrument would be maintained internally for use in instructional improvement and evaluation. Colleges should maintain, in an easily accessible format (e.g., on a College Web page) the results of the "Student Questions" section of the teaching evaluation instrument

Appendix D

Proposed Universal Questions for Students Evaluation of Teaching Instrument

All questions (except background information) answered on a 6-point scale: n/a – not applicable:

1 2 3 4 5 6
SD – Strongly Disagree; D – Disagree; N – Neutral; A – Agree; SA – Strongly Agree; N/A

1. Overall, this instructor was an effective teacher.
2. Overall, I gained a great deal from this course.
3. The instructor treated all students with respect.
4. The instructor provided me with helpful feedback on my performance.
5. The instructor gave a clear presentation of the material.
6. My ability to think critically (i.e., analyze, interpret and evaluate information) has improved as a result of taking this course.
7. My ability to understand ideas and concepts has improved as a result of taking this course.
8. My ability to communicate effectively has improved as a result of taking this course.

Student Questions

9. The instructor handled questions well.
10. The instructor used examples that had relevance to the subject matter.
11. The instructor used class time well.
12. This course was designed to keep me engaged in learning.
13. The workload was appropriate for what I gained from this course.

Background Information

1. What do you expect your final grade in this course to be? (A+, A, A-, B+, B, B-, C+, C, C-, D, F, Pass, Fail)
2. How much effort did you put into the course? (None, Little, Average, Above Average)
3. For you this course is? (Core Requirement, Major Requirement, Minor Requirement, Elective)
4. Which of the following reasons were important in your decision to take this course? (Requirement, Interest, Convenient Time, Instructor's Reputation, Took Instructor Before)



UNIVERSITY OF MARYLAND

OFFICE OF THE SENIOR VICE PRESIDENT
FOR ACADEMIC AFFAIRS AND PROVOST

02-03-39

1119 Main Administration Building
College Park, Maryland 20742-5031
301.405.5252 TEL

August 26, 2005

Prof. Adele Berlin, Chair
Campus Senate
Marie Mount Hall
CAMPUS

Dear Adele:

At the request of the Senate, I asked my senior staff to explore ways in which campus-wide on-line teaching evaluations could be implemented in a way that ensures high student response rates. Their recommendations to me are enclosed in the attached memo.

I believe that these recommendations, based upon a program already successfully implemented at Northwestern University, meet this objective in a way that ensures that students assume a significant share of the responsibility for the success of the teaching evaluation program. I therefore am fully supportive of these recommendations.

Please let me know if there is any further assistance I can provide as the Senate considers this important issue.

Sincerely yours,

William W. Destler
Senior Vice President for
Academic Affairs and Provost

