*The Committee to Assess the Evaluation of Teaching was established in the spring of 2001 jointly by Provost Robert Barchi and then Senate Chair Larry Gross (*Almanac *March 27, 2001).*
See www.upenn.edu/almanac/v47/n27/senate.html.

Report of the

Joint Faculty Senate/Provost's Office
Committee to Assess the Evaluation of Teaching

Summer 2002

Prologue

As the charge to the committee indicates *(reprinted below)*, we were directed to conduct an assessment of the current instruments and methods used by the various schools to evaluate faculty teaching. We have done so and have made a number of concrete recommendations which we believe will improve this process.

However, and by way of preface, we also want to emphasize that evaluation, no matter how precise, can only measure how good teaching is at a given time; in itself, evaluation can do little to create good teaching. Furthermore, we believe that we have already achieved most of the gains in teaching quality that can be expected from evaluations, no matter how well conceived or how widely used they are. If Penn is serious about increasing excellence in teaching across the University, at all levels, then we must realign our institutional structures accordingly. This realignment will require more substantial faculty participation than have our previous efforts.

In fact, Penn's priorities and practices, like those at other research-intensive universities, often subordinate teaching to other faculty responsibilities, in particular of course to research. To be sure, as the teaching evaluations demonstrate, a great deal of excellent teaching takes place here in a wide range of settings, from classrooms and laboratories to field stations and individual faculty offices. It is clear from this evidence that most faculty approach their teaching obligations with care, dedication and even devotion. The increasingly intense competition for the various teaching awards given each year throughout the University is but one indication of this fact.

Nonetheless, there remains a significant distance between our pedagogical ambitions and our accomplishments, a distance created by the relentless demands of research and the rewards attached to that activity. The disjunction between teaching and research is inadvertently but quite accurately captured in our local language: we speak of our "teaching load," but we refer to research and scholarship as "doing our own work."

Consistent with this disjunction is the widely held belief among faculty that the semester or year of leave, in which one does not teach at all, is our best, most productive time, even though we are not performing the one task that separates us from members of a research institute or a think tank. Semesters of teaching small graduate seminars are typically considered next best, followed by advanced undergraduate courses in one's specialty. Many faculty try to avoid teaching general or introductory courses, especially in large lecture sections. Salaries and raises are largely tied to research productivity and to outside offers. Teaching too often appears to be the residual in departmental or school planning, while areas of research specialization command the bulk of our collective attention.

If this construction of our campus imperatives seems reductive or extreme, it undoubtedly conveys something of the uneasy reality. And if we aspire to nurture a "culture of teaching" at Penn, then our faculty must accept a larger share of responsibility for developing that culture at all levels: in the training of graduate students, in the mentoring of junior faculty colleagues, in the assignment of courses among senior faculty.

Such a cultural change will only take place if the senior faculty leads it. If this change does take place, the gap between our pedagogical ambitions and our accomplishments will rapidly close.

We make an affirmative reference in the report to the utility of the Center for Teaching and Learning, and we are unanimous in our endorsement of the Center's work. We also applaud the emphasis on teaching in the new Strategic Plan. We urge all of the Deans, as they take up the specific recommendations in this report, to situate those detailed discussions in the broadest context. As Donald Kennedy, former president of Stanford, has written in a recent book on higher education: "Responsibility to students is at the very core of the university's mission and of the faculty's academic duty." We as faculty need to do all we can to carry out that duty with the same energy, imagination, and zeal that we invariably bring to our scholarship.

With this in mind, we offer below some specific recommendations concerning the construction and administration of teaching evaluations. It should be kept in mind that these evaluations are used by three different audiences, each for different purposes: by students when they select classes; by faculty seeking to improve their own teaching; and by administration in evaluating other faculty members for salary raises, tenure and promotion. The comments offered in the paragraphs that follow are intended to improve the evaluation process for all of these purposes.

Committee Charge

The Committee to Assess the Evaluation of Teaching was established in the spring of 2001 jointly by Provost Robert Barchi and then-Senate chair Professor Larry Gross. The committee was given the following charge:

"Over the past few decades the University of Pennsylvania has steadily heightened the level of scrutiny of undergraduate and graduate teaching by our faculty. However, many of the mechanisms that are used to generate data for use in such assessments have evolved haphazardly, often by incorporating into official use course evaluation forms initiated by undergraduate student bodies (such as SCUE, the Student Committee for Undergraduate Education) for their own purposes. As these systems have been woven through the fabric of official University procedures for evaluating faculty performance, it is appropriate that we undertake a review of current teaching evaluation mechanisms with the goal of identifying the most appropriate means for each school to undertake this important responsibility. While it seems impossible to seek, or achieve, a single evaluation system for all schools and programs, there should be minimum standards of fairness and quality that protect all of our faculty and further the education of all of our students.

"In April, 1998, the Subcommittee on Teaching Evaluations of the Faculty Senate Committee on Administration recommended the establishment of 'a committee representing the twelve schools to evaluate the current course/faculty evaluation process.' We expect that the current Committee will draw upon and benefit from this Subcommittee report."

Assessing the Evaluation of Teaching

Recommendations

I. Mid-Semester Feedback Forms

Mid-Semester Feedback Forms are useful for faculty who wish to improve their teaching, since these forms provide prompt feedback to the instructor concerning problems students may have with a course, together with suggestions for improvement and helpful information on successful aspects of the course. Following the recommendation of an earlier committee on teaching which reported in March 1991, a large and growing number of faculty have been using the optional Mid-Semester Feedback form (Appendix 1 includes copies of three different versions of this form).

We recommend that all faculty consider using a Mid-Semester Feedback Form, in both undergraduate and graduate classes. The form is strictly for the faculty member's individual use, and is not forwarded to the department chair or any other administrator. In Wharton and the College, systems are already in place to encourage faculty members to use Mid-Semester Feedback Forms each semester. To increase the use of these forms, we propose that each undergraduate school send an e-mail message to all faculty who are teaching undergraduate courses each semester two weeks before the middle of the semester with (1) a brief paragraph explaining the advantages of using Mid-Semester Feedback Forms, (2) the forms available as attachments or from a website, and (3) information on resources available for a faculty member who wishes to improve his or her teaching. Obviously, in order to include the last component, it will be important to identify such resources.

The College Mid-Semester Teaching Feedback Questionnaire Version B contains a large number of multiple-choice rating items, together with space for comments on each of these. To make this form more useful for large courses, the College experimented this past semester with a technique to mount the Mid-Semester Teaching Feedback Questionnaires on Blackboard course websites. This technique restricted access to students enrolled in the course and automatically tabulated all multiple choice and check-box responses. It turned out not to be possible to guarantee *absolute* anonymity of students' responses using the technology at hand, which many students and faculty regard as essential. Those who used this application nonetheless found that it provided a very efficient means for students to submit feedback and for faculty to digest it. The College will continue to seek a cost-effective means of securing complete anonymity.

II. Course Evaluation Form

Teaching evaluation forms have been used at Penn for over 35 years. Over that period, the instrument in use has been revised several times to improve the usefulness of the data it provides. With regard to the collection of quantitative data, we acknowledge that there is no such thing as a perfectly valid survey instrument, but we support the use of instruments which provide reasonably accurate information about teaching quality.

We propose that the undergraduate schools, as well as some graduate schools, continue to use end-of-semester Course Evaluation Forms to provide information for administrative purposes, information for student course choice, and to a limited extent for teachers to use to improve their teaching. Appendix 2 provides copies of two forms currently in use in undergraduate courses; the first is used in SAS, SEAS and Nursing, the second in Wharton.

Proposed Revisions for the Course Evaluation Forms

We propose modest revisions of the current Course Evaluation Forms which would continue the philosophy of having a relatively brief form with questions that would apply broadly to most or all courses in order to foster a high response rate and provide comparable data across a wide range of courses. Our suggestions for revisions would apply to both the SAS/SEAS/Nursing form and the Wharton form which are very similar (Appendix 2). The suggestions are listed in Appendix 2 in the order in which the items appear on the form. Our suggestions for retention, revision, or addition of items were based on (1) the usefulness of items for students, faculty and/or administrators and (2) research evidence that certain items are particularly strong predictors of student learning. [1]

Administration of Course Evaluation Forms

Given the importance of the purposes for which these evaluations are used, we believe it is important to encourage as thoughtful responses as possible, specifically with regard to the comments students make. We therefore recommend the following best practices in administering course evaluation forms in class at the end of each semester:

1. Ask students to complete the form at the beginning of class, rather than at the end. Students are more likely to take the time to make thoughtful comments if they do not have the option of simply filling out the form and leaving class early.
2. Before distributing the form and leaving the room, remind students of the importance of the evaluation process: that this is their way of giving feedback on this important aspect of their college experience, and therefore their evaluations should be as complete and thoughtful as possible.

Presentation of Results from Course Evaluation Forms

Reports summarizing the results of student ratings on the Course Evaluation Forms should include histograms of the distribution of responses for the first two items ("overall quality of the instructor" and "overall quality of the course"). These histograms should be labeled with the actual response categories, not numbers, so it is clear for example that an average rating of 2 corresponds to "good". When histograms are not provided, the percentage of "poor" ratings should be included with the mean in official reports of results from Course Evaluation Forms.

We propose that each year a letter be distributed to Deans, Department Chairs, and faculty in the Schools which use these forms. This letter would include (1) a brief summary of the evidence supporting the validity of student ratings of instruction and (2) a brief summary of the evidence that average course ratings vary by characteristics such as course level, size, disciplinary category, and whether the course was an elective, with a quantitative estimate of the magnitude of the difference in average scores related to each of these dimensions (controlling for the other dimensions and based on analyses of data from Penn). The version of the letter received by Deans and Department Chairs would encourage them to take this information into account in interpreting teaching ratings in dossiers for faculty retention and promotion, and would also encourage them to utilize various additional types of information (such as student letters and evaluation of teaching materials) to evaluate a faculty member's teaching for these purposes.

III. Teaching Evaluation for the Purpose of Faculty Tenure and Promotion

Teaching evaluation is used by students to select classes, and by faculty members to improve their own teaching methods. The third purpose, however—as part of the dossier of a candidate for tenure or promotion—has a much greater effect on a faculty member's future. It is therefore crucial that judgments about a candidate's teaching abilities and performance be based on as broad a range of information as possible. Specifically, information from a variety of sources should be used in order to provide a well-rounded picture of the candidate's abilities and accomplishments. Self-authored teaching statements, for example, can provide important information, as can input from peers. Our goal is to give senior faculty and administrators making decisions about tenure and promotion as much relevant information as possible.

We therefore propose that the dossier for every candidate for tenure or promotion be required to include a section on teaching. This section should include aggregate course evaluation data drawn from the course evaluation forms in individual classes (see section II, above). In addition, it should include one or more of the following items:
- a self-authored statement of the candidate's teaching philosophy and description of accomplishments, plus copies of syllabi or other course documents which the candidate believes accurately represent his or her teaching style; these could be annotated by another person knowledgeable in the same field;
- letters of reference:
  - from peers when appropriate (i.e., other faculty members with whom the candidate has team-taught, or a course director for a class of which the candidate has taught a particular section)
  - from students
  - from Teaching Assistants, where relevant

We propose that student reference letters be recommended as one element within the teaching section of a candidate's dossier for tenure or promotion. In order to ensure that the letters present a well-rounded picture of a candidate's teaching, they must be solicited under controlled circumstances. We recommend the following as best practices for this purpose:

a. For faculty members who have taught comparatively few students, letters should be solicited from all students. For faculty members who have taught a great many students (for example, in large introductory lecture classes), letters should be solicited from a selected group of students to illustrate the range of classes (small, large, graduate, undergraduate, etc.) which the candidate has taught over the course of at least two years.

b. When letters are not solicited from every student the faculty member has taught, students to be solicited should be selected in a random manner (for example, every tenth person on a class list).

c. The letter of solicitation should specify what information is to be included in the reference letter (Appendix 3).

d. The solicitation may be made electronically, and e-mail responses are acceptable, but anonymous letters should not be accepted.

e. The solicitation letter should explicitly note that letters will be kept confidential from the candidate.

If a School or department chooses not to use student reference letters, it may substitute a transcript of the comments from the instructor's course evaluation forms for use in the teaching section of the candidate's dossier.

IV. Teaching Improvement

Teaching evaluations often affect a faculty member's career; and although the validity of student evaluations is often debated, the responses are frequently helpful for individuals who seek to improve their teaching. Faculty who want to improve need the resources to accomplish this goal. We recommend, therefore, that the University provide a variety of department-based and university-wide opportunities to enhance teaching skills. Within departments, research has shown that discipline-based mentoring by senior faculty members provides substantial benefit, provided that the mentors are, themselves, trained to offer pedagogical support.

Department-based programs, however, are not enough. We recommend that the University dedicate resources to a Center for Teaching and Learning, taking advantage of the current program housed in SAS. Employing extensive faculty input, the Center would bring together other existing resources including the Learning Resource Center and the IT assistance the Library provides to faculty. The collaboration of these resources would create opportunities for a community of learning involving both faculty and students. A comprehensive Center for Teaching and Learning would also provide print and non-print resources (including current websites) designed to help faculty members improve their teaching skills.

V. A Note on the Use of Web-based Evaluation Forms

The Committee considered whether the course evaluation process should be moved entirely to the web. We concluded that course evaluation will almost certainly be done electronically within the next five to ten years, since electronic evaluations are flexible and capable of handling both open and forced-choice items. However, a number of technical issues await resolution, and we do not recommend that Penn be in the vanguard of this particular technological movement. We would prefer to learn from the process at other institutions, and then adapt as necessary for the Penn environment.

Some Concerns

Whether a system is purchased or custom-built, two of the looming issues are cost and processes for development/implementation. Within the latter are many types of issues that need to be addressed. A partial list of the issues includes the following:

• *Technical* – What system will be used? What language/software will be used? Where is the server located? What system can handle the load of thousands of students responding in a short amount of time? Who inputs the data to be evaluated?

• *Shared resources* – Will each school create a parallel system or will one system be jointly shared (and supported) by multiple users?

• *Compliance* – What incentives can or will be used to keep student participation high?

• *Reporting and Analyses* —The data architecture for the existing evaluation process involving SAS, SEAS, Nursing, and Wharton has already been developed for the University's Data Warehouse and is in the testing stage; it is expected to be ready in time for the Fall 2002 evaluation process. This could be the foundation for storage of future evaluations and creation of standard reports. Nevertheless, questions remain, for example, who has access to the data? Who actually produces the report?

We want to emphasize that many of these issues, and similar ones, have been addressed by groups involved in creating systems for sharing, analyzing and reporting the current paper evaluations. The expertise of individuals in these groups should be utilized as a transition is made to an electronic system. Any new system could build on the existing systems and take into account the need to link to historical data.

Since movement to some type of electronic system is inevitable, we recommend that a working committee be named now to begin the planning process. It is essential that future planning take into account historical data and build on the existing system so that a transition to a new system be as seamless as possible. Specifically, we recommend that individuals intimately involved in developing procedures and policies for the University's Data Warehouse also be involved in future planning groups.

Endnote

[1] A large body of research evidence indicates that student ratings on this type of form are correlated with student learning (as assessed by grades on common exams in multisection courses). In addition, more limited evidence indicates that more favorable student ratings of teaching are associated with "deeper study strategies" (more attempt to understand and to integrate material and less memorization) and with more subsequent coursework and activity in the field taught. (This research evidence is summarized in *Effective Teaching in Higher Education: Research and Practice*, edited by Raymond P. Perry and John C. Smart, 1997, Agathon Press.)

Committee to Assess the Evaluation of Teaching

David Pope, *co-chair,* Professor of Engineering
Peter Conn, *co-chair,* Professor of English and Deputy Provost
Jacob Cytryn, undergraduate student
Anita Gelburd, Assistant to the Deputy Provost, *ex officio*
Larry Gladney, Associate Professor of Physics
Robert Hornik, Professor of Communications
Arlene Houldin, Associate Professor of Nursing
Lindsey Mathews, undergraduate student
Paul McDermott, Professor of Education, and Psychology in Psychiatry
William McManus, Director of Institutional Research and Information Systems, School of Arts and Sciences, *ex officio*
Philip Nichols, Associate Professor of Legal Studies (on leave AY 2001-2002)
Kent Peterman, Director of Academic Affairs, College of Arts and Sciences, *ex officio*
Larry Robbins, Director, SAS Center for Teaching and Learning, *ex officio*
Judy Shea, Associate Professor of Medicine, Director of Evaluation and Assessment of the Academic Program, School of Medicine
Deborah Bolton Stagg, Director of Institutional Research, Wharton School, *ex officio*
Alan Strudler, Associate Professor of Legal Studies
Archana Vemulapalli, graduate student
Ingrid Waldron, Professor of Biology

Editor's Note: *See next page for Appendices*

Appendices

In the interest of conserving space, only a section of Appendix 2 is included here.
To see the full appendices, contact Mary Esterheld in the Deputy Provost's Office, 122 College Hall.

Appendix 1: Mid-Semester Feedback Forms

There are three Mid-Semester Feedback Forms currently in use:
- Wharton Midterm Course Feedback Form
- College Mid-Semester Teaching Feedback Questionnaire - Version A
- College Mid-Semester Teaching Feedback Questionnaire - Version B

Appendix 2: Revisions to Course Evaluation Form

In addition to the proposed revisions to the course evalation form and the sample form, both of which appear below, this appendix also includes the two course evaluation forms currently in use.

In order to accommodate two additional items and a statement of purpose not included currently in the SAS/SEAS/Nursing form, we propose that the top part of the form be condensed, for example by decreasing the space for the top heading and by having the course identification information be represented by 5 rows of bubbles (with twice as many columns) instead of 10 rows of bubbles. These changes will be important to preserve adequate space for comments at the bottom.

To improve the usefulness of the home school information, we propose the following categories:

College/SAS
Engineering
Nursing
Wharton
CGS
GSE
GSFA

We propose no changes in class level, expected grade or major vs. general requirement vs. elective items.

To more accurately reflect the distribution of students' GPAs, we propose to change the cumulative GPA categories to:

3.7-4.0
3.4-3.6
3.1-3.3
2.0-3.0
< 2.0

To inform students concerning the purposes served by the information they provide, we propose that a modified version of the statement currently included in the Wharton Course Evaluation Form be included, as follows. "Your responses on these forms are used for various purposes, including decisions concerning faculty reappointment, promotions, tenure, and teaching awards."

For the left column of items, we propose retaining the poor/fair/good/very good/excellent rating scale and four of the existing items, but modifying two items, adding two and making one substitution, resulting in the following list:

1. Overall quality of the instructor
2. Overall quality of the course
3. Instructor's ability to communicate the subject matter
4. Instructor's ability to respond to students' questions
5. Instructor's ability to stimulate student interest
6. Instructor's accessibility
7. Course organization
8. Value of assignments (including homework and/or papers)
9. Amount learned from this course (including knowledge, concepts, skills and/or thinking ability)

To accommodate the added items in the left column, the "additional questions" in the SAS/SEAS/Nursing form might need to be moved to the right column. Aside from renumbering the items in the right column, the only other change we propose for these items is a revision of the last item, including a rewording as follows:

If you are aware of cheating in this course, please fill in the circle, and describe the type and extent of cheating.

This would be followed by only one circle and an open space for the requested description (at least in the SAS/SEAS/Nursing form which has space for this).

Finally, we propose that the revised course evaluation form be pilot tested for one semester before general adoption.

Appendix 3:
Sample Solicitation Letter for Student Reference Letter

Appendix 2: Proposed Course Evaluation Form

YOUR HOME SCHOOL	YOUR CLASS LEVEL	YOUR EXPECTED GRADE	THIS COURSE IS	YOUR CUM GPA
COLLEGE/SAS ○ CGS ○	FRESHMAN ○ POST-BAC ○	A ○ PASS/FAIL ○	IN MAJOR ○	3.7-4.0 ○
ENGINEERING ○ GSE ○	SOPHOMORE ○ GRAD/PROF ○	B ○ AUDIT ○	GENERAL REQ ○	3.4-3.6 ○
NURSING ○ GSFA ○	JUNIOR ○ OTHER ○	C ○	ELECTIVE ○	3.1-3.3 ○
WHARTON ○	SENIOR ○	D ○		2.0-3.0 ○
		F ○		<2.0 ○

PLEASE RATE QUESTIONS 1-9 ON A SCALE OF: P=POOR, F=FAIR, G=GOOD, VG=VERY GOOD, E=EXCELLENT.
PLEASE LEAVE THE QUESTION BLANK IF YOU LACK INFORMATION ON A PARTICULAR CHARACTERISTIC.
YOUR RESPONSE ON THESE FORMS ARE USED FOR VARIOUS PURPOSES, INCLUDING DECSIONS CONCERNING FACULTY REAPPOINTMENT, PROMOTIONS, TENURE, AND TEACHING AWARDS.

Please rate the	P	F	G	VG	E			1	2	3	4	5
1. Overall quality of the instructor	○	○	○	○	○		10. Please rate the difficulty of this course (1=easy to 5=difficult)	○	○	○	○	○
2. Overall quality of the course	○	○	○	○	○							
3. Instructor's ability to communicate the subject matter	○	○	○	○	○		11. Please rate the amount of work required for this course. (1=very little to 5=very much)	○	○	○	○	○
4. Instructor's ability to respond to student's questions	○	○	○	○	○		12. Would you recommend this course to a major? (1=no to 5=strongly)	○	○	○	○	○
5. Instructor's ability to stimulate student's interest	○	○	○	○	○		13. Would you recommend this course to a non-major? (1=no to 5=strongly)	○	○	○	○	○
6. Instructor's accessibility	○	○	○	○	○		14. If you are aware of cheating in this course, please fill in the circle and describe the type and extent of cheating.			○		
7. Course organization	○	○	○	○	○							
8. Value of assignments (including homework and/or papers)	○	○	○	○	○							
9. Amount learned from this course (including knowledge, concepts, skills and/or thinking ability)	○	○	○	○	○							

Use this space to answer any additional questions provided by the instructor

	P	F	G	VG	E
Additional Question 1	○	○	○	○	○
Additional Question 2	○	○	○	○	○

COMMENTS ENCOURAGED BELOW