

Observations on the Association's Statement on Teaching Evaluation

The following observations were approved for publication by the Association's Committee on Teaching, Research, and Publication in May 2005 on the occasion of the thirtieth anniversary of the adoption in 1975 of the Statement on Teaching Evaluation. These observations were adopted by the Association's Council in November 2005.

Introduction

The *Statement on Teaching Evaluation* remains sound policy, and its guidance "for arriving at fair judgments of a faculty member's teaching" continues to be invaluable.¹ The world of higher education has changed significantly, however, since the publication of the statement thirty years ago. The proportion of faculty appointments that are nontenure track, full time as well as part time, has grown to more than 60 percent of all faculty appointments; student evaluations of teaching are increasingly relied upon in decisions about renewal, tenure, promotion, and salary increases; new computer-based tools have been developed to administer, disseminate, and interpret student evaluations of teaching; and corporate forms of governance are threatening to become prevalent in higher education. In light of these new and important changes to higher education, we thought it desirable to bring forth these comments.

Observations

The *Statement on Teaching Evaluation* was published in 1975 "as a guide to proper teaching evaluation methods and their appropriate uses in personnel decisions." It recommends that colleges and universities and their academic departments have clear, written policies about expectations concerning teaching, and provide support for meeting those expectations. The statement emphasizes that descriptions of a professor's teaching and data about the teaching obtained from other sources must be accurate. It cautions that "the full dimensions of teaching should not be slighted in the desire to arrive at usable data and systematic practices." The statement further recommends that the faculty member being assessed have a meaningful role in the evaluation process, and it calls for faculty members to have the primary, although not the exclusive, role in evaluating an individual faculty member's performance as a teacher. Last, it urges that "factual data, student opinion, and colleague judgments should be central in the formal procedures for review, which should involve faculty discussion and vote."

Since the statement's publication, a growing body of scholarship, supplemented by extensive experience, has developed regarding effective teaching and learning strategies, the role of teaching centers in assisting faculty to enhance their teaching, and the evaluation of teaching by faculty peers and students. Both this scholarship and experience show that faculty members share pedagogical and evaluation materials with colleagues; that self-evaluation, assessment by teams of faculty members, and student evaluations provide a regular flow of data that facilitates continual improvement in teaching; and that technology has contributed new tools with which teachers may assess and improve their teaching and conduct student evaluations. In addition, the dissemination of evaluations of teaching can provide students with useful information with which to plan their course of study, and offers faculty peers and administrators a richer body of material on which to base judgments of professional merit.

Two key issues identified in the 1975 statement continue, however, to trouble the evaluation of teaching: how best to ensure that evaluations of teaching provide accurate information about the effectiveness of teaching and how to ensure that faculty have the primary responsibility for devising and implementing teaching evaluations. We will offer the following comments on these two issues, and will then turn to an issue of academic freedom.

The expanding use of, and reliance on, teaching evaluations since 1975 have given rise to an abundance of data about classroom performance. Although survey instruments, data collection, and methods for disseminating information vary within and across institutions, several common areas of concern exist with regard to the type, quality, and accuracy of the data collected. One concern involves the practice of relying solely on numerically based student evaluations. Although survey forms that call for numerical evaluations have the advantage of offering a common instrument that can provide comparative data, in many places they have become the dominant, or even sole, component of the evaluation of teaching and tend to displace less standardized and more individualized forms. We encourage the inclusion of a section that invites students to provide written comments relevant to a particular course and a particular instructor.

A second concern is that numerically based evaluations are sometimes misunderstood and therefore misinterpreted, resulting in erroneous conclusions about the absolute and relative merit of faculty members. The data employed are often ill-suited for the type of statistical analysis carried out, being neither continuous in nature nor useful for making the fine distinctions on which rewards are often based. For example, teachers with numerical ratings falling below a certain percentile obtained by the entire faculty, such as 50 or even 90, may be rated as inadequate. Such a use of relative position as an absolute measure of merit overlooks the possibility that the majority of faculty in some departments or institutions will be ranked as "superior," so that some or even all of those in the department or institution with evaluations below the mean may in fact be good teachers. In other departments, most faculty will be ranked as "poor," and some or even all of those ranked above the mean may be poor teachers.

Another problem is the belief that judgments can be made about the relative merit of faculty members despite response rates of students that may vary widely from one class to another. If the response rate by students in one class is higher than 90 percent while in another class of the same size the response rate is 50 percent, no comparison of the sets of responses is statistically valid.

Beyond these concerns about the interpretation of numerical data, a growing body of evidence suggests that student evaluations create pressures that work against educational rigor. Rather than exclusively measure teaching effectiveness, evaluations tend also to measure the influence of personal style, gender, and other matters extraneous to the quality of teaching. One possible way to offset such influences is to include questions directed at student self-evaluation in evaluations of teaching (for example, did the student dedicate sufficient time to course work?).

There is also evidence suggesting that the use of student evaluations in faculty personnel decisions may produce incentives to weaken or dilute course content and contribute to grade inflation. The pressure to do so can be acute for probationary faculty and especially for part- and full-time non-tenure-track faculty who are subject to nonreappointment at the discretion of the administration. These faculty members may come under pressure to give higher grades than students deserve to improve their teaching-evaluation scores. The pressure could be diminished in several ways: through evaluations with specific criteria for faculty members on renewable term appointments; through appointment and reappointment decisions based on criteria specific to the positions; and through opportunities for faculty on renewable term appointments, part or full time, to move into tenured or tenurable positions.²

The faculty should have primary responsibility for developing reliable methods for evaluating teaching, which should distinguish between questions that are appropriately answered by students and those that are appropriately addressed by peers. The numerical scores that students give instructors on evaluations should be interpreted with valid statistical methods, and, as noted above, a comments section should be included in the instrument.

A separate set of concerns arises as a result of shifts from collegial forms of academic governance, with the emphasis on consultation and participation, to corporate, top-down managerial styles. The corporate model has brought a tendency to judge the success of teaching and teachers in numerical terms, and by the redefinition of education as a commodity that universities are expected to sell and that students, as paying customers, are expected to consume. A potentially troubling effect of this approach to education is that it can create pressure

on faculty members to please students rather than to educate them. In this context, teaching evaluations may be seen less as a guide for improving teaching and more as an important marketing tool because they may enhance the image of the institution as responsive to student opinions. Online capabilities may then extend the use of teaching evaluations beyond their traditional scope and purpose. Thus it is not unknown for links to teaching evaluations to be provided in online course catalogues, or for class-by-class, year-by-year evaluation scores to be posted on the Web and thus made available to all faculty and students. Since these measures do not require permission of the instructor, faculty oversight of the evaluation process can be compromised. We urge faculty members to take an active role in formulating and adopting policies that specify the proper use and dissemination of student evaluations of teaching, including the purpose and scope of their online availability.

The relationship between faculty members and those in positions of institutional authority, including department chairs, has also changed. In collective-bargaining situations, consideration should be given to including contractual provisions to restrain the potential for administrators to extend their role in the evaluation process by requiring a single approach or evaluative method of the faculty within a department or school, or by mandating the inclusion of certain quantifiable features of evaluation, such as timetables for improvement and specified educational outcomes.

With regard to the role of administrators in the evaluation process, the *Statement on Teaching Evaluation* states, "Emphasis in evaluation should be upon obtaining first-hand evidence of teaching competence, which is most likely to be found among the faculty of a department or college and the students who receive instruction. Evaluation of teaching in which an administrator's judgment is the sole or determining factor is contrary to policies set forth in the *Statement on Government of Colleges and Universities*." We reaffirm the primacy of the faculty in the evaluation of teaching and recommend that institutional evaluation policies preserve the essential role of peer evaluation of courses and teaching performance.

The increasingly important role that technology has assumed in the instruction of virtually all disciplines has raised an issue of academic freedom: the widespread expectation that instructors use particular forms of technology in their courses. Evaluation questionnaires often ask about the use of technology in the learning experience (for example, how many assignments used Web-based resources, or how often did the instructor use the smart workstation for lecture demonstrations?). The inclusion of such questions seems to suggest that faculty members are expected to use these technologies in their classrooms, even though they may have no pedagogical reason to do so. Instructors should be free to determine the extent to which they employ technology in their classrooms according to their professional assessment of its benefits and costs to instruction in a particular subject matter.

The authors of the 1975 statement advocated the evaluation of teaching for the "development of the teacher and the enhancement of instruction." Instructors may accomplish these purposes in several ways. One is to make more extensive use of the teaching portfolio. Because the portfolio typically contains a statement of teaching philosophy, copies of syllabi, sample corrected work and other course materials, and summaries of student evaluations of the faculty member's teaching, it goes beyond exclusive reliance on student ratings in the evaluation of teaching. Teaching centers have also proved valuable. They provide faculty members with access to fellow teachers or mentors of recognized quality, to courses and workshops on pedagogical matters, to assistance in videotaping of classes, and to libraries with relevant literature. We encourage the use of classroom visitation by peers on a regular basis with advance notice of such visits. We recognize that the logistics of implementing peer evaluation, with at least two or three visits to each class for each evaluation, may be cumbersome. Recognition of the practical difficulties, however, does not relieve members of the profession of their obligation to ensure the quality of teaching, or their responsibility to evaluate teaching performance according to knowledge both of the subject matter and of teaching methods that students may not possess. With respect to guidelines for classroom visitation, we reiterate those outlined in the 1975 statement: "There must be an understanding among the visitors and the visited upon such matters as who does the visiting, how many visits are made, what visitors look for, what feedback is given to the visited, and what other use is made of the information."

In conclusion, institutions, departments, and faculty members should ensure that the evaluations of teaching promote and sustain excellence of teaching and education, that faculty be primarily responsible for devising systems of evaluation and monitoring their use, and that the development and implementation of teaching evaluation methods be consistent with principles of academic freedom and shared governance.

Notes

1. Since the publication of the 1975 "Statement on Teaching Evaluation," the Association has addressed the issue of college teaching in additional policy statements. Published statements in *Policy Documents and Reports*, 10th ed. (Washington, D.C.: AAUP, 2006), include "The Work of Faculty: Expectations, Priorities, and Rewards" (1993); "The Assignment of Course Grades and Student Appeals" (1997); and the "Statement on Faculty Workload with Interpretive Comments" (2000).

2. On these and related issues, see the report on "Contingent Appointments and the Academic Profession" in *Policy Documents and Reports*, 98–114.