Better Teaching Through Better Evaluation: A Guide for Faculty and Institutions

Susan Kahn

Follow this and additional works at: http://digitalcommons.unl.edu/podimproveacad

Part of the Higher Education Administration Commons


This Article is brought to you for free and open access by the Professional and Organizational Development Network in Higher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in To Improve the Academy by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Better Teaching Through Better Evaluation: A Guide for Faculty and Institutions

Susan Kahn
University of Wisconsin System

This paper surveys current literature and thinking on teaching evaluation in higher education. It is intended to help faculty, administrators, departments and institutions think through the main issues that need to be considered in developing a teaching evaluation plan. It is organized around these issues, which include definitions of good teaching, formative and summative evaluation of teaching, sources of evaluation information, use of evaluation to improve teaching, and features of effective evaluation programs. Along with discussion of these issues, it provides examples and models of successful evaluation approaches and includes a list of suggested readings for readers interested in learning more about particular aspects of teaching evaluation.

The quality of undergraduate teaching in American higher education has become the focus of intense discussion and debate in the past few years, both in and out of the academy. State legislatures, governing boards, and the broader public have demanded increased accountability from higher education institutions. At the same time, university faculty members and administrators have voiced deepening concern about the need for better ways to prepare a diverse student body to meet the changing needs of our society and economy. Fueled from without and within, campuses across the country have begun to
reassess their priorities and, in many cases, to strengthen their commitment to the undergraduate teaching mission.

As institutions rededicate themselves to enhancing teaching effectiveness, they are recognizing that efforts to improve teaching and learning must go hand in hand with efforts to improve the evaluation of teaching. Evaluation that yields meaningful, useful information about teaching has two important purposes: identifying areas of needed improvement and development, both for individual faculty members and across departments and institutions; and providing a basis for rewarding strong teaching performance in personnel and salary decisions. Serving both these purposes is essential to the success of any initiative to improve individual faculty members’ teaching and to encourage renewed commitment to undergraduate teaching at all levels of the institution.

This paper surveys recent literature and other work on the topic of evaluating teaching in higher education. It is intended to serve as a preliminary guide to the topic for faculty, administrators, departments and institutions interested in learning more about the current thinking on teaching evaluation. It is predicated on the new paradigms of scholarship developed by Boyer and others, who view teaching as a demanding and serious intellectual pursuit on a par with traditional research, and it is organized around the main issues that need to be considered in the development of teaching evaluation processes. Along with a brief introduction to these issues, it describes examples and models of successful evaluation approaches and identifies additional resources for readers wishing to pursue specific aspects of evaluation.

The main issues discussed here include:

- Definitions of Good Teaching
- Formative and Summative Evaluation of Teaching
- Sources of Evaluation Information
- Use of Evaluation to Improve Teaching
- Evaluation Programs That Work
Definitions of Good Teaching

Logically, it seems that any plan for evaluating teaching should begin with a clear definition of what good teaching is. But, in this case, clear definitions are hard to come by; every teacher knows that effective teaching and learning can occur in many different ways and that there is no single definition of good teaching. Well-conceived evaluation strategies avoid simplistic, prescriptive conceptions of good teaching. They are highly contextual, relying on criteria developed by faculty themselves and geared to the specific teaching and learning goals of their particular departments, programs, and institutions (Seldin, 1992).

In developing these criteria, several studies and approaches can serve as good starting points. Chickering and Gamson’s (1987) “Seven Principles for Good Practice in Undergraduate Education,” for example, applies across a broad range of disciplines. The authors provide seven criteria for good teaching: “good practice encourages student-faculty contact”; “good practice encourages cooperation among students”; “good practice encourages active learning”; “good practice gives prompt feedback”; “good practice emphasizes time on task”; “good practice communicates high expectations”; and “good practice respects diverse talents and ways of learning” (pp. 3-6). The principles are a concise and thoughtful synthesis of much of the recent research and new thinking about what works in the contemporary college classroom. (For some important caveats on the principles and their application, however, see Creed, 1993).

The recent monograph on The Teaching Portfolio published by the American Association for Higher Education (AAHE) takes a different approach. This report proposes that teaching includes four “core tasks”: course planning and preparation; actual teaching; evaluating student learning and providing feedback; and keeping up with the professional field in areas related to one’s teaching. It suggests that these tasks might provide a useful framework for formulating criteria for teaching evaluation (Edgerton, Hutchings, & Quinlan, 1991).

Other researchers and commentators emphasize the highly contextual nature of effective teaching. Lee Shulman’s (1989) “Toward a Pedagogy of Substance,” for example, defines good teaching as the
capacity to transform the specific concepts of a particular discipline or subject into terms that can be understood by a particular group of students. William Cerbin (1992b) proposes a related approach called "learning-centered evaluation," which evaluates teaching effectiveness in the context of the learning goals of a specific course. The approach focuses on the relationship among teaching objectives, actual teaching practices, and student learning outcomes—"the complexities of daily teaching and learning"—and uses this focus to capture "the real action in the classroom....the relationship between teaching and learning." In a recent article, Cerbin (1992a, p. 8) suggests strategies for accomplishing this and for translating the results into real teaching and learning improvements.

The literature on effective teaching is vast and is growing rapidly; it includes many ideas and approaches that can help in planning teaching evaluation strategies. But the crucial ingredient in developing successful evaluation strategies, according to much of this literature, is faculty participation and leadership. Also important are clear, written criteria, developed by faculty and communicated to those being evaluated. Criteria should take into account the complexity and variety of teaching and learning and be appropriate to the context and the purposes of the evaluation. The different purposes of evaluation are discussed in greater detail below.

**Formative vs. Summative Evaluation**

The information resulting from teaching evaluation is most often used for one of two purposes: to assess the effectiveness of specific teaching practices and identify areas for improvement or development; or to arrive at broad judgments of teaching effectiveness that allow for comparisons among faculty members and that can be used to make personnel decisions. Evaluation conducted for the first purpose is called "formative evaluation"; evaluation conducted for the second is called "summative evaluation." Both types of evaluation are legitimate and important to enhancing and maintaining instructional quality. As Maryellen Weimer (1987) explains: "Formative evaluations must target appropriate areas of change. Summative assessment must reflect the impact of those changes" (p. 10).
The procedures for conducting an evaluation and the type and amount of information collected depend on the purpose of the evaluation. For example, since formative evaluation aims to identify particular areas for improvement, evaluation procedures are usually designed to collect more detailed information than procedures used for summative purposes. The timing of an evaluation may also reflect its purpose. For instance, student evaluation for formative purposes might be carried out early or mid-semester so as to allow for mid-course changes, while student evaluation for summative purposes normally occurs toward the end of the term. Summative evaluation procedures also tend to be relatively standardized, while formative evaluation procedures usually allow for more individual faculty control and choice.

Most experts on teaching evaluation agree that formative and summative evaluation procedures and information should be strictly separated. They argue that formative evaluation data must be entirely confidential and that using such data for summative purposes discourages faculty participation in formative activities. Other researchers counter that the amount of time and work involved in developing and administering two distinct teaching evaluation systems is simply impracticable for many institutions. They suggest seeking creative ways to combine the two. For example, student evaluation forms might be designed to elicit both formative and summative data; the formative data could then be provided to the instructor only.

Each institution must resolve these issues for itself, keeping in mind that evaluation procedures, practices, and forms should reflect the purposes of the evaluation. Some additional examples of formative and summative approaches are given in the sections below. Several of the suggested readings listed at the end of this paper also include discussions of the differences between formative and summative evaluation and effective approaches to each.

Sources of Evaluation Information

Too often, "teaching evaluation" is equated with "student evaluation." While student evaluations are essential to assessing teaching, they do not give us a full picture of teaching effectiveness and should
always be used in combination with information from other sources. In *Scholarship Reconsidered*, Boyer (1990) suggests that evaluation information be collected from at least three sources: student assessment, peer assessment and self-assessment (Boyer, 1990). This section discusses each of these sources, and includes a brief discussion of the teaching portfolio, a promising, multi-faceted approach to capturing the complexity and diversity of effective undergraduate teaching.

**Student Evaluation**

Student evaluations are by far the most widely used approach to teaching evaluation. A 1991 survey showed that about 75 percent of colleges and universities use student evaluation to assess faculty teaching effectiveness (Seldin, 1992). Other studies have found that most faculty members favor involving students in teaching evaluation. A majority of faculty members also believe that their own teaching has improved as a result of student input (Boyer, 1990).

A wealth of information and literature—though not all of it in agreement—exists on effective design and use of student evaluations. But procedures and forms are often poorly designed and administered. For example, students are not always given adequate preparation for their roles in teaching evaluation. They may not understand that their opinions are important and will be taken seriously. The purpose of the evaluation, who will see the results, and how the results will be used should be explained to them, perhaps as part of a class session or of freshman orientation (Boyer, 1990; Sewall, 1992).

In addition, student evaluation forms and questionnaires must be carefully designed, preferably in consultation with a specialist in evaluation or faculty development. Their length and the questions included should reflect the purpose of the evaluation and the types of information sought (Sewall, 1992). Peter Seldin suggests that questionnaires intended for formative use include 20 to 30 diagnostic questions focused on specific teaching behaviors, while forms intended for summative use might include four or five questions on overall performance. Whatever the purpose, forms should include some open-ended questions and space for comments; instructors can
Better Teaching Through Better Evaluation

often learn the most from student comments about the reasons for their numerical ratings and the areas where change is needed (Seldin, 1992).

Some innovative approaches to formative student evaluation have emerged in the past few years. The most widely known are probably the “classroom assessment” techniques developed by Patricia Cross and Thomas Angelo (1988)—quick, simple exercises that enable instructors to gauge, in an ongoing way, how well students are learning material or skills, what they are having difficulty with, which teaching approaches are effective and which are not. A related strategy, the “student management teams” used at the University of Colorado, the University of Wisconsin System, and elsewhere, uses a small group of student volunteers to collect feedback from other class members and meet regularly with the instructor over the course of the semester (Nuhfer, 1992). The approach has parallels with Continuous Quality Improvement practices. Cerbin’s (1992b) “learning-centered evaluation” also calls for a mid-course formative review that focuses on how well specific teaching techniques are supporting the specific learning goals of a course. For example, to what extent are class discussions helping students learn to integrate and synthesize complex ideas?

These approaches yield more detailed, immediate and useful information than does the traditional student evaluation model of a single, generic questionnaire filled out by students at the end of the semester. The new approaches are also good pedagogy in that they allow for mid-course adjustments when needed, and give students a greater feeling of ownership and involvement in the course. In this sense, they are really a form of active learning as well as a type of formative evaluation of teaching effectiveness.

**Peer Review**

Most faculty members are ambivalent about peer review of teaching (Seldin, 1992). Only 25 percent of colleges and universities make regular use of classroom visits by colleagues as a method of evaluating teaching. In cases where colleagues do review their peers, what is actually “peer reviewed” is usually information and evidence submit-
ted by others, such as student evaluations, rather than direct observations of one another's teaching (Edgerton et al., 1991).

Yet there are aspects of teaching that one's colleagues are uniquely qualified to judge: how appropriately a course or class is organized; how well important content and concepts are represented; whether topics are integrated effectively; whether examples are relevant; whether classes are taught at an appropriate level of difficulty; whether the instructor is presenting the most current information about a field; whether assignments and tests are consistent with the teaching and learning goals of a course, and so on (Boyer, 1990; Edgerton et al., 1991; Seldin, 1992). For this reason, Boyer and others recommend a serious and systematic approach to the evaluation of teaching by one's colleagues. Such an approach might include regular classroom visits, with observations focusing on faculty-established criteria, as well as peer review of the "samples" or "products" of teaching, such as representative syllabi and examinations, videotapes of classroom teaching, or perhaps examples of student work at the beginning and end of a course. Ultimately, the aim is to foster a culture that encourages faculty to move freely in and out of one another's classrooms, both to learn from and constructively critique each other (Boyer, 1990; Edgerton et al., 1991; Seldin, 1992; Van Horne, 1992).

The research on peer review of teaching suggests that it has been more successful as an approach to formative than to summative evaluation. There are several successful models for formative peer review of undergraduate teaching, such as the "Teaching Improvement Process" (TIP) used in the University of Kentucky Community College System and elsewhere (Holmes, 1992). In another model, used by the University of Wisconsin-La Crosse's Foreign Language Department, junior and senior faculty pair up and exchange extensive classroom visits. The Communication Department at the University of Wisconsin-Parkside, which has a competency-based undergraduate major, recently adopted a competency-based approach to teaching evaluation that is used for both peer review and self-evaluation (Rusterholz & Logsdon, 1992). These models and others are showing how effective peer review can be in improving teaching, stimulating discussion, and transforming the teaching culture within departments and institutions.
When peer review is used for summative purposes, it should be preceded by substantial faculty discussion and carefully planned and carried out. Classroom observations of an instructor should be conducted by several colleagues, not just one (Seldin, 1992). Written appraisals, whether for formative or summative purposes, are most useful when they discuss specific teaching practices and make specific suggestions for change (Weimer, 1987). Finally, Boyer (1990) and others recommend that, in addition to teaching and its direct products or results, articles and conference presentations related to teaching should be peer-reviewed and given credit in tenure, promotion and merit decisions.

**Self-Evaluation**

Some departments and institutions have found self-evaluation helpful for both formative and summative purposes. A common self-evaluation strategy is to ask faculty members to prepare philosophical statements about their teaching, including discussion of their teaching goals and of how their teaching practices support these goals. These statements might also discuss teaching strengths, plans for improvement, and contributions to the teaching needs of the department or institution (Boyer, 1990; Edgerton et al., 1991; Seldin, 1991; Seldin & Annis, 1992).

For summative purposes, such narratives may provide a helpful framework for interpreting evaluation information from other sources. For formative purposes, preparing such statements may be a developmental process in itself, since it requires instructors to reflect on what they are doing in the classroom and why. To the extent that instructors develop more self-awareness and, in turn, can more clearly communicate their teaching goals to students, self-evaluation contributes directly to better teaching.

Self-evaluation/philosophical statements about teaching are frequently associated with teaching portfolios, an approach to teaching evaluation that many institutions are currently testing. The next section of this paper focuses on this promising evaluation strategy.
The Teaching Portfolio

A teaching portfolio is "a collection of materials documenting teaching performance" (Seldin & Annis, 1991-92, p. 6a), a kind of "extended teaching resume" (Edgerton et al., 1991, p. 3). It represents a multi-faceted approach to teaching evaluation that uses material from several sources to explore the various dimensions of teaching. At its best, it documents an instructor’s overall approach to teaching, bringing together specific evidence of instructional strategies and effectiveness in a way that captures teaching’s intellectual substance and complexity.

Portfolios have many advantages over traditional approaches to teaching evaluation. For instance, faculty members typically compile their own portfolios. The portfolio approach thus shifts much of the responsibility for the evaluation process into the hands of those being evaluated; evaluation becomes less something that is done to faculty. Assembling a portfolio is a developmental experience in itself, since it requires instructors to reflect on and rethink their teaching goals and strategies. In addition, portfolios have great versatility and may be geared to a variety of purposes: conducting formative and summative evaluation of instructional effectiveness; screening applications for teaching positions; evaluating candidates for outstanding teaching awards, and so on (Edgerton et al., 1991; Seldin, 1991; Seldin & Annis, 1992).

Researchers have suggested several approaches to assembling and organizing teaching portfolios. Peter Seldin recommends that portfolios include material drawn from three broad areas: material from oneself, such as representative syllabi and examinations, reflective statements on teaching philosophy and strategies, descriptions of efforts to improve or innovate; material from others, such as results of student and peer evaluations, documentation of teaching awards or other recognition of excellent teaching; and the products of one’s teaching, such as student essays or scores on standardized tests (Seldin, 1991; Seldin & Annis, 1991-92; Seldin & Annis, 1992).

The AAHE monograph (Edgerton et al., 1991) on the teaching portfolio recommends organizing portfolio entries around the four core tasks of teaching: course planning; actual teaching; evaluating
students; and keeping up with teaching developments in one's field. In the first category, course planning, a portfolio might include successive syllabi from a course that has evolved over several years. A sample of "actual teaching" might include a short videotape or detailed observations of a particular class session by peers or students. To illustrate one's approach to evaluating students, one might include a student paper or examination with instructor's grades and comments. Portfolio entries for the fourth category, maintaining currency in one's field with respect to teaching, might note teaching-related conferences or sessions attended and discuss how their ideas were incorporated into a particular course. For each category, the AAHE monograph suggests including an actual work sample—such as a syllabus, examination, or a videotape—along with a reflective statement, which might comment on the rationale for the approach used and for changes made over the years, or on what worked well, what did not, and why, or on other aspects of the portfolio entry.

All this can seem daunting, but Seldin (1991) and others emphasize that an effective portfolio is selective, not exhaustive, in the materials it includes. The point is to suggest the scope and quality of one's teaching through careful selection of representative materials. He recommends that the body of the portfolio be no more than four to six pages long, and that this material be supplemented by appendices that include empirical evidence, work samples, and other data supporting the assertions made in the body (Seldin & Annis, 1991-92; Seldin & Annis, 1992). The AAHE monograph (Edgerton et al., 1991) similarly proposes that portfolios be kept as "lean and lively" as possible; it notes that "a selected, limited number of sample entries can be highly revealing," and encourages a view of the portfolio as a set of examples of "best work," not as a compilation of "all work" (pp. 4, 11).

The contents and organization of a portfolio also should reflect its purpose. Portfolios assembled for formative purposes, for instance, might focus on a particular course: its objectives and the methods used to support these objectives, innovations attempted, student achievement, and lessons learned. Portfolios compiled for summative purposes would likely include a core set of mandatory items—such as summaries of student evaluations, syllabi, statements of teaching
philosophy and efforts to improve teaching—in order to allow for comparisons between and among portfolios (Edgerton et al., 1991; Seldin, 1991; Seldin & Annis, 1991-92; Seldin & Annis, 1992).

Finally, experience suggests that portfolios are best prepared with the help of a colleague (Seldin, 1991; Seldin & Annis, 1991-92; Seldin & Annis, 1992). This colleague might be a peer "partner" also working on a portfolio, a senior faculty mentor, or a faculty development specialist. Such collaborations provide an outside perspective that strengthens the portfolio and supports the developmental aspects of portfolios by stimulating discussion and reflection on teaching. In fact, by encouraging faculty to take on new roles in "the documentation, observation and review of teaching," the processes of preparing and evaluating portfolios can foster "the creation of a culture in which thoughtful discourse about teaching becomes the norm" (Edgerton et al., 1991, p. 4). In this sense, both the substance of portfolios and the work and discussion surrounding them can help to encourage the development of a community of teacher-scholars within departments, programs, colleges, and institutions.

Use of Evaluation to Improve Teaching

The main purpose of teaching evaluation is to improve teaching. Evaluation policies and processes should be designed and carried out with that purpose in mind. This section focuses on ways to use evaluation processes and results to bring about real instructional improvements.

Formative evaluation activities work best when participation in them is voluntary, non-threatening, and collaborative—that is, when faculty do not feel that evaluation is being done to them. Some evaluation experts suggest putting all formative activities under the control of the individual faculty member (Weimer, 1987). At the same time, departments and institutions should encourage collaboration among faculty, in pairs or small groups, on formative evaluation and teaching improvement; many of the evaluation activities discussed earlier—construction of teaching portfolios, formative student evaluation approaches such as classroom assessment and learning-centered evaluation, and formative peer review, for example—lend themselves
well to collaboration and discussion. In addition, formative evaluation and improvement efforts are more likely to become part of a department or institution's culture if participation in them is recognized and rewarded in formal personnel decisions (Boyer, 1990; Edgerton et al., 1991; Van Horne, 1992; Weimer, 1987).

Several studies have shown that evaluation is much more likely to lead to improvement when it is followed up by consultation with a faculty/instructional development specialist or someone with similar expertise, such as a senior faculty mentor who is an excellent teacher. Such consultations can help faculty interpret evaluation results and decide on specific improvement plans. Most experts suggest focusing improvement efforts on one or two manageable goals at a time—integrating active learning activities into a particular unit of a course, for example, or redesigning writing assignments in a course to require students to do more creative or analytical thinking (Weimer, 1987).

Most important, formative evaluation activities should always emphasize development, input, and feedback, not overall judgments of teaching effectiveness or comparisons among faculty members. These kinds of judgments are not the purpose of formative evaluation and can discourage faculty members from participating (Weimer, 1987).

In the case of summative evaluation of teaching, on the other hand, overall judgments and comparisons of teaching effectiveness are important goals. To ensure that judgments are as fair and accurate as possible, summative evaluation processes should be well-designed and well-documented and should rely on multiple sources of information (Boyer, 1990; Edgerton et al., 1991; Seldin, 1992). Procedures, forms, and criteria should be designed with the institution or department's mission, needs, culture, and values, as well as validity and reliability, in mind. Faculty should be involved as much as possible in developing procedures, and all faculty members should be given a full written description of the evaluation program (Seldin, 1992).

Perhaps most important, there should be clear connections between summative evaluation results and rewards such as promotion, tenure, and merit pay. Faculty committed to teaching excellence deserve recognition and rewards; when departments and institutions
provide them, "they extend to faculty powerful reasons to make that commitment" (Weimer, 1987, p. 11).

Evaluation Programs that Work

Faculty acceptance is the foundation of any successful teaching evaluation program. It is faculty who must carry out the evaluations, interpret the results, plan for improvements, and make recommendations on tenure, promotion, and other personnel matters. The most effective evaluation programs—not just bureaucratic exercises, but real efforts to enhance teaching and learning—are thus designed and controlled by the faculty. Strong administrative backing, including resources for professional development to complement and follow up on evaluation activities, is also essential to making teaching excellence a true institutional priority (Seldin, 1992).

At the same time, faculty members and administrators interested in improving teaching evaluation need to recognize that reforms cannot be instituted overnight. Current evaluation systems are often little more than pro forma exercises, largely ignored by most faculty members; others operate in punitive, demoralizing ways, providing little support for genuine improvement. Many faculty members are understandably apprehensive or skeptical about the prospect of "more" teaching evaluation. The development of better ways to evaluate teaching must thus go hand in hand with the development of ways to encourage ongoing discussion about teaching and its improvement as a routine departmental and institutional activity.

For these and other reasons, it may be best to "start small" in planning initiatives to improve teaching evaluation—perhaps with a pilot program involving a few outstanding instructors in developing teaching portfolios or using classroom assessment techniques for formative purposes. These kinds of activities combine evaluation with instructional improvement and development, and lend themselves to collaboration and discussion. As such, they can provide a good basis for convincing faculty to see them as valuable and to consider more ambitious efforts.

In the end, effective evaluation and improvement efforts depend on the development of institutional and departmental cultures that
value undergraduate teaching. Providing resources to support teaching improvement and evaluation, encouraging faculty collaboration on formative evaluation and improvement activities, building a system that rewards commitment to teaching excellence, and encouraging a view of teaching as a form of scholarship worthy of serious consideration and discussion can all help support the growth of such a culture. While change is likely to be slow and the process often frustrating, it is still a worthwhile goal to pursue: faculty, administrators, and students alike stand to benefit from its achievement.

References


Holmes, S. (1992). *The teaching improvement process (TIP).* Workshop presented at *Evaluating Teaching: More Than a Grade,* conference of the University of Wisconsin System Undergraduate Teaching Improvement Council, Madison, WI.


Sewall, T.J. (1992). *Student evaluation.* Paper presented at *Evaluating Teaching: More Than a Grade,* conference of the University of Wisconsin System Undergraduate Teaching Improvement Council, Madison, WI.


### Other Suggested Reading


