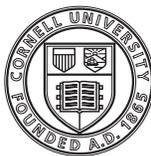


TEACHING EVALUATION HANDBOOK

Third Edition, 1997

David Way, Ph.D.
Cornell University



CENTER FOR TEACHING EXCELLENCE
420 Computing and Communications Center
cornellcte@cornell.edu
www.cte.cornell.edu
607-255-3990

Chapter I—Conceptual Overview

During the last 30 years research on teaching effectiveness has shifted away from behavioral and psychological approaches. This shift has grown out of the gradual recognition that teaching is a sufficiently complex human endeavor that its effectiveness cannot be accounted for in purely behavioral terms. The research traditions prevalent in psychology and education 30 years ago, which produced many experimental studies, proved less fruitful than was hoped because methodological constraints limited (and probably trivialized) the knowledge contributed. The focus on *behaviors* (rather than the richer and more powerful realm of teachers' *thinking*) and the oversimplification of teaching to a set of technical skills have been partly responsible for many faculty member's mistrust and skepticism about a meaningful and practical body of knowledge about teaching effectiveness in higher education.

Knowledge about what constitutes teaching effectiveness has therefore not had much influence on teaching practice in higher education, until recently. The practice of teaching in higher education very strongly reflects the educational system that prepares and qualifies its teachers during their graduate education. As research has enlarged and deepened the knowledge in the disciplines, the knowledge that informs the teaching of that substantive knowledge has been largely limited to modeling one generation of faculty to guide the next. Teachers in higher education are, for the most part, educated to be justifiable authorities on the subjects they teach, but only indirectly in how to teach those subjects. Exceptions include those faculty members who were fortunate enough to be exposed to individuals that stimulated a broader range of practice and experimentation and who were more likely to adopt proven innovations. The attention given to “active learning” (Angelo, '95, Halpern, '99 & Mazur, '97) seems to indicate that there is too much lecturing and not enough critical thinking going on in higher education classrooms. We must begin to do things differently if we desire different results.

Over the last 10 years, there has been a significant increase in the concern for the training of graduate teaching assistants, more attention has been placed on the relative importance of undergraduate education, and strong criticism has been levied at the degree to which research has become almost obsessively dominant. All of these factors have contributed to creating an atmosphere where the central administration at Cornell has supported an in-depth analysis of the tenure system and the way teaching is evaluated and rewarded.

Some Basic Definitions and Assumptions

In the literature on the evaluation of teaching there has been a tradition of distinguishing two forms of evaluation: *summative* evaluation—made for personnel decisions like tenure and promotion, and *formative* evaluation—conducted for the improvement of practice.¹ This tradition has maintained that these two evaluative practices be conducted separate from each other. There are strong arguments for this separation. One is that summative evaluation serves the purposes of administrators and is a public process, while formative evaluation serves the individual teacher and is therefore confidential. If a tenure candidate cannot trust that consultation sought out to help improve teaching practice is confidential, the candidate may not seek out that consultation because the implication available publicly is that his or her teaching needs improvement. Another argument for keeping formative and summative evaluation processes separate is the concern that the consultant may have a conflict of interest between serving the needs of the teacher and serving the needs of the administration.

However, there is a price to be paid for treating these two evaluation functions separately: the summative evaluation process may become too oriented towards comparing faculty with each other as a means of defining teaching effectiveness, while an individual's achievements in the improvement of teaching practice (the objective of formative evaluation) may become overlooked. When the primary means of evaluating an individual's teaching is based on comparing his or her mean scores on several indices with an aggregate set of scores computed by lumping all faculty in a department or school into a common formula, it is possible to lose sight of significant differences in improvement achieved on an individual level. Regardless of relative experience and skill, everyone has the potential to improve. In fact, a truly scholarly approach to teaching would imply that one is never finished learning about it, just as one never comes to fully understand a phenomena under investigation in research.

Another important argument for not separating summative and formative evaluations has to do with efficiency and cost. All the effort, care, and time expended on documenting and evaluating teaching can be used most efficiently if it serves two functions simultaneously rather than just one. It is extremely important that any added emphasis on the evaluation of

¹Scriven, M. "Summative Teacher Evaluation" *Handbook of Teacher Evaluation*, Millman, J. (ed), Sage Publications, 1981., pg. 244.

teaching for summative purposes be carried out as efficiently as possible so that it does not become a major burden to all those involved. This can be achieved if evaluating teaching within the tenure process is thought of as informing and supporting the improvement of teaching practice.

To achieve efficiency in the evaluation of teaching, specific procedures must be adopted and carried out by everyone involved. Procedural roles must be defined to avoid duplication of work and to ensure the highest standards are maintained. In addition, considerations of confidentiality—*what* is confidential and *when* during the tenure period—are important in defining roles. Time availability and evaluative expertise are other factors in determining efficiency. Summative evaluation should be conducted by peers. However faculty peers may not always have the expertise or time necessary to properly conduct formative evaluations of teaching.

It is possible to achieve evaluation efficiency without costing anyone in the process significantly more time. If we acknowledge that improvement of teaching practice is an expectation that all faculty, both tenured and untenured, must continually demonstrate, and that it is also an important value that faculty hold, just as they value and expect research quality and results to continually evolve, then there already exists a basic motivating force to encourage efforts at improvement on an individual, on-going basis. This places the burden of proof on each faculty member from the moment he or she is hired. Monitoring of instructional quality and effectiveness, strategizing and experimenting with activities aimed at improvement, and the documentation of those activities and their measured results are all responsibilities that can be expected of a tenure candidate starting from day one. These expectations should be communicated to faculty members when they are hired so they can begin preparing for them right away.

However, the individual faculty member is not without resources to help fulfill these goals and responsibilities. The fourth recommendation of the Dean of Faculty Committee's report (Appendix) states, "that all departments (or other appropriate unit) establish a standing committee on teaching. Such committee members would be responsible for overseeing peer evaluation of a tenure candidate's teaching." In this case "peer evaluation" is not limited to classroom observation but is meant to encompass the entire range of data gathered, from student and alumni letters, to course teaching materials and measures of student learning. It is the intention that this committee establish and maintain guidelines and criteria for the

evaluation of teaching, establish procedures to be followed, and standards to be set. Colleagues are available to assist the tenure candidate in making his or her case for quality teaching. Where faculty colleagues do not have the time or expertise to assist in these areas, instructional development services are available through the Office of Instructional Support.

The question of defining excellence in teaching

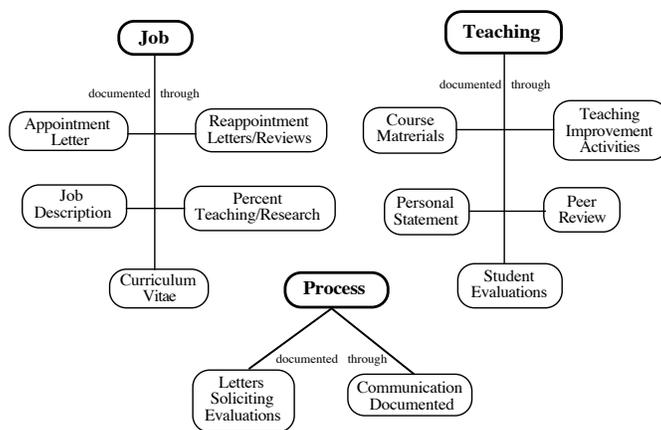
A frequently raised question both in the literature on the evaluation of teaching and in conversations with Cornell faculty members is “How can we define excellence in teaching?” It would seem this question must be answered before one can proceed with any kind of evaluation. The problem with this question is that it may not be answerable in *absolute* terms. A major reason there has not been a useful and practical definition of excellence in teaching is that teaching may be too broad a concept to be limited by a single definition. Teaching undergraduates will involve different criteria than the context of teaching graduate students. The criteria for excellence in teaching to be considered for promotion to full professor will necessarily be different from those for consideration at the associate professor level. Excellence in teaching will vary by discipline, course design and level of experience. A more useful way of thinking about excellence in teaching is in *relative* terms: to what degree has improvement in practice revealed an individual’s capacity for continual growth and development which is intrinsic worth to the department and college?

It will be far more difficult to agree upon, measure and evaluate an absolute definition of excellence in teaching than a relative one. It is possible to adequately evaluate teaching quality in terms of how much an individual has improved over time and still be fair between candidates when it is recognized that the level of standards that establish that relativity have been set upon the candidate’s appointment. The fact that the candidate has been hired to teach at this particular institution sets the level of standards by which he or she will be evaluated. The task now becomes one of determining how capable the individual is of improvement based upon a sufficiently broad range of criteria and data sources. Some people may not require a lot of improvement to function at an exemplary level, yet because of their particular capabilities, they may exceed established expectations. Others will show an even greater degree of improvement, but still not measure up to expectations; in which case they probably should not have been hired in the first place.

Documentation of Teaching within the tenure file

All colleges and schools have guidelines for evaluating faculty for tenure. These guidelines should include explicit instructions for gathering data for documenting teaching. It is extremely important that explicit criteria for evaluating teaching be established and communicated to the candidate upon hiring through these college guidelines. The relative weight of all criteria and data sources should be spelled out and periodically discussed among the faculty. The roles of everyone involved in a tenure case should be known and spelled out (chairs, *ad hoc* committee members, students, colleagues, dean, chief academic officer).

Tenure and promotion files should document contracts between the department, the chair and the candidate upon hiring through the inclusion of appointment and reappointment letters. A possible framework for documenting teaching appears in Figure 1 below:



Framework for Tenure File Contents

Figure 1

This framework focuses on three major areas of *documentation*, 1) Job—how was the position originally described during the search, how did the parameters of the position change once the position was filled? What was the teaching and content area background of the person hired? What relative weight was given to teaching responsibilities and did that weight (percentage of time commitment) change over time during annual reviews. 2) Process—was the evaluation process itself documented? Did reviewers, both students and peers, receive explicit instructions and criteria to help them carry out their evaluations? A simple letter specifying to reviewers the criteria to be used in evaluating classroom

performance, course design and materials makes everyone’s job easier—both internal and external reviewers and administrators reviewing the documentation: department chairs and deans. 3) Teaching—is there a balance in data between the various components of teaching and data sources? Does peer review adequately cover the range of teaching activities peers are capable of evaluating? Are student evaluations reported in a way that development of effectiveness over time can be determined? Were course materials reviewed? Does the documentation include a reflective personal statement by the candidate which explicates his or her efforts at improving teaching effectiveness?

A framework for evaluating teaching is included in Figure 2 below:

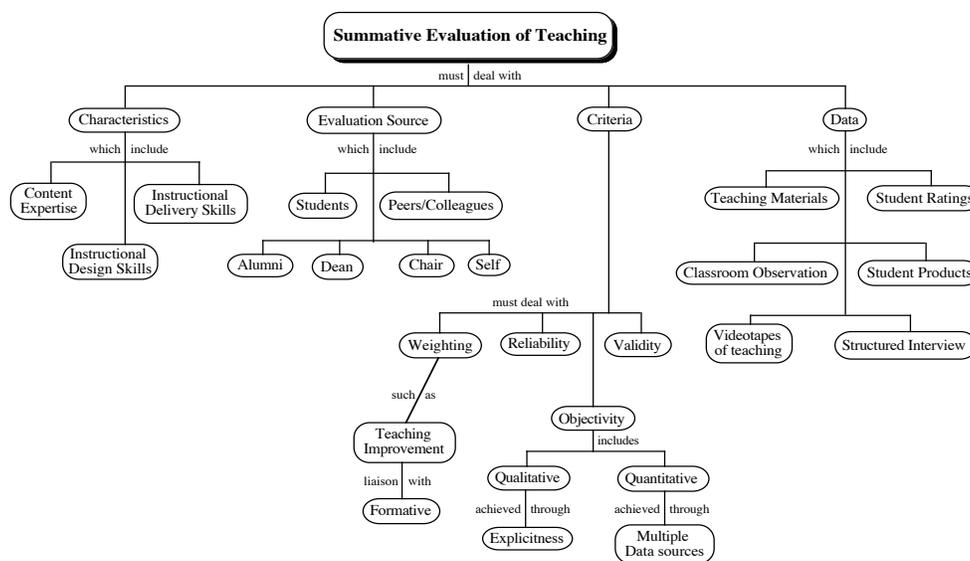


Figure 2

This framework focuses on four major categories of the summative *evaluation* process: the range of characteristics of teaching which are evaluated, the range of evaluation sources available to supply data, the criteria by which the data on teaching are evaluated, and the range of data types available to be evaluated. The first category includes three basic characteristics and responsibilities of teachers: content expertise, instructional design skills, and instructional delivery skills. Since these are all primary skills of teaching it is necessary that all be discretely evaluated. Content expertise is the most obvious link to the candidate’s educational and professional background and to his or her research expertise. Instructional design skills are necessary for effective course design, development and planning. Included in this are the skills necessary to effectively evaluate student learning as evidenced in examinations, paper and project assignments, and grading schemas.

Instructional delivery skills are those that are evident in the classroom and in interactions with students during office conferences and advising.

With this range of characteristics of teaching, an equivalent range of evaluation sources is necessary since no single evaluation source will be qualified to evaluate all characteristics of teaching. Content expertise must be evaluated by peers in the discipline and faculty colleagues in the department. Instructional design skills are best evaluated by peers both within the candidate's department and discipline and outside of it. A balanced *ad hoc* committee will include peers throughout the candidate's own institution and outside of it. Expertise from peers acclaimed not only for their knowledge of course design but also in educational practice are important in the makeup of who is involved with evaluating course design skills. Students have been proven to be effective evaluation sources for instructional delivery skills, and to a certain degree course design skills. The weighting attributed to their contributions of evaluating instructional and course design skills is an important consideration. The issue of weighting different types of teaching data is addressed more thoroughly in Chapter 4.

Some authors in the area of evaluation of teaching suggest that alumni are useful for providing a perspective on a candidate's teaching that no one else can. Students who have graduated and been in the work force for a year or two have the opportunity to reflect on how effective the teaching they received was from a practical point of view. However, alumni evaluators must be enlisted with caution. Alumni are more valid as evaluation sources if they are asked to evaluate an overall course and what they have learned as a result of taking that course, rather than relied upon to evaluate specific aspects of classroom performance after they have left the institution. A major disadvantage in soliciting alumni for teaching evaluations is they have low response rates.

The role deans and chairs can play in the evaluation process is of a qualitatively different nature than students or peers. A primary responsibility of deans and chairs is to ensure that the tenure file is complete, follows accepted college and departmental guidelines, includes a sufficiently broad range of data and that appropriate and explicit criteria have been used to evaluate the data. Additionally, administrators are concerned with the candidate's long-time worth to the department or college. Including a broader range of data on teaching in a tenure file will prove more cumbersome to administrators, peers and the candidate alike if certain guidelines and procedures are not adopted and adhered to in order to reduce that

data to a manageable form. The task is to reduce the data in a way that does not distort it. More will be said about data reduction in Chapter II.

The candidate is an important evaluation source, especially in terms of instructional development. Hard data by and of themselves cannot tell the complete story of an individual's teaching experience and development. It is not only advantageous but also valid for a candidate to supply some form of reflective, written statement which not only provides a more intimate view of what has transpired, but helps in interpreting abstract data, like numerical student evaluation scores.

A common complaint of the evaluation of teaching is that it is a subjective judgment, that objectivity is impossible. However, objectivity is possible through both qualitative and quantitative approaches. The *quality* of objectivity can be achieved by the development of explicit criteria for evaluating the data collected. To achieve *quantitative* objectivity, data should be collected from multiple sources (colleagues, students, advisees, graduate students, alumni) and in various forms (quantitative data from student questionnaires, peer evaluation, classroom observation, course materials, personal statements from the candidate, qualitative data from students, advisees and alums in the form of letters and samples of student products.)²

Reliability and validity are two criteria to be applied to all data provided for tenure and promotion decisions. For example, in the case of classroom visits by peers, if peer observers are untrained in the task, their observations may be less reliable³. A more complete discussion of reliability and validity as they relate to student and peer evaluation data will follow in Chapters III and IV.

Weighting is another important factor in the evaluation of teaching. Its consideration begins with the candidate's job description: what percentage of his or her time has been designated for teaching? Has this percentage changed over time? These factors will govern what overall weight should be given to teaching *per se*. Once that has been established other weighting decisions must be made. What relative weight will be given to evaluating

²Whitman, N. & Weiss, E. *Faculty Evaluation: The Use of Explicit Criteria for Promotion, Retention and Tenure*. AAHE-ERIC/Higher Education Research Report, No. 2, 1982, ERIC ED221148.

³Centra, J.A. "Colleagues as Raters of Classroom Instruction". *Journal of Higher Education*. 46, 1975, 327-337.

the candidate's instructional design skills? What weight will be given to the improvement of teaching practice for this candidate, based upon his or her previous experience and performance and the work load assigned? These matters will be addressed further in Chapter IV.

If quantitative objectivity is an important criteria for evaluating teaching, data must be representative of all dimensions of teaching: content expertise, instructional design skills, instructional delivery skills and the capacity for improvement of practice. Data on content expertise will be found in course materials: what readings were assigned, examples of exams, examples of graded papers and projects, classroom teaching plans, lecture notes and handouts. The focus on content expertise will vary, depending on the candidate's relationship with the course content: to what degree does it overlap with the candidate's field of expertise? The primary consideration in evaluating content expertise is: is the right stuff being taught? This can only be determined by looking at what was taught as evidenced in course materials, and to a certain degree through classroom observation.

Instructional design skills will also be evident in course materials: syllabi, assignments, schemas for evaluating what students have learned, handouts, non-print materials like computer software and the choice of textbook and course readings used. Measuring improvement in practice will require historical data gathered over time. This data must be comparable, for example, if a candidate has been observed in class within the first year and an observation report is included in the tenure file, equivalent observations must be provided for subsequent years to determine instructional development.

The range of data on teaching included in a tenure file and the way it is evaluated are matters which must be decided on the department and college level. This discussion has brought up some of the major issues that should be considered in setting departmental and college policy. To help make those decisions less cumbersome we will look more closely at the documentation and evaluation tasks. The next chapter will present a model—developed by the American Association of Higher Education in collaboration with some of the leading authors in the area of documentation of teaching—the Teaching Portfolio.

References

Angelo, T.A. (1995). Classroom Assessment for Critical Thinking. *Teaching of Psychology*, 22(1), 6-7.

Halpern, D.F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning*, No. 80, 69-74.

Mazur, Eric. (1997) *Peer Instruction —A User's Manual*, Prentice Hall: Upper Saddle River, N.J.