

## Preface

This handbook grew out of collaborative efforts between the Dean of Faculty's office during Professor Walter Lynn's tenure and a select committee of senior faculty from throughout the university in the early 1990's. The handbook was a recommendation in the committee's report, *Evaluation and Recognition of Teaching*. That report's several recommendations were discussed with the faculty at various meetings in the Cornell colleges during the spring of 1992, and finally presented to Provost Malden Nesheim and President Frank Rhodes who adopted them all. The report was based on an analysis of tenure files from throughout the university and interviews with the deans that were conducted in the early summer of 1991. The tenure files reviewed were chosen according to the following criteria: tenure had been granted within the previous three years, the candidate had at least a 45% responsibility to teach as indicated by the chair, and the broadest number of colleges and departments were represented in the sample.

## Introduction—Purpose of Handbook

This handbook is designed to serve as a guide to encourage a view of teaching practice and its evaluation that reflects the intellectual challenges and richness that are an integral part of it. It takes the view that teaching and research activities do not have to compete with each other, but constitute two integrated aspects of scholarly activity. Roald Hoffman, Emeritus Professor of Chemistry, has described the reciprocal roles of teaching and research as a “complex dance.”

The desire to teach others, enhanced by being *obliged* to teach others, leads to greater creativity in research. The rhetorical imperative operates to make a scientist or scholar examine widely the potential responses (objections?) of his or her audience. Teaching enlarges one's encounters with real audiences, therefore sharpens the imagined audience one engages in the inner dialogue in the course of research.<sup>1</sup>

The practices suggested in this handbook are meant to unburden faculty with unnecessary drudgery within the teaching evaluation process. Departments and colleges will continue to make their own decisions regarding the evaluation of teaching. This handbook is meant to support those decisions as a guide and resource.

---

<sup>1</sup>Hoffman, R. (1989, Summer). Research/teaching—A complex dance. In *Cornell Undergraduate Education*, ( 2), 2, Office of Instructional Support, Cornell University, Ithaca, NY, 1-3.

## **Using this Handbook**

This handbook has been organized to reflect the process of evaluating teaching for tenure and promotion, starting with the need to document the candidate's teaching and instructional development, and finishing with the evaluation of that documentation's data by peers and administrators. It has been written to be of use to deans, department chairs, faculty members who are serving on tenure review committees and new faculty, as well as departmental standing committees on teaching. Chapter I provides a conceptual overview that will be of particular interest to department chairs and newly hired faculty members. The major concepts relevant to the tasks of documentation and evaluation are explored, including the distinction between evaluation of teaching for tenure decisions and the evaluation of teaching for the improvement of practice, and how the two are related. This is followed by a discussion of what an explicit definition of excellence in teaching implies for the evaluation process. Chapter I finishes with a review of the issues relevant to establishing criteria for evaluating teaching and how these criteria can drive the process of documenting one's teaching.

Chapter II will be of particular usefulness to the newly hired faculty member, and also to department heads and unit-based teaching committees. It presents a model for documenting teaching by the tenure candidate—the Teaching Portfolio. As a model, the teaching portfolio can serve to set the boundaries of the documentation task, including establishing criteria for inclusiveness that do justice to the range of responsibilities inherent in teaching, yet set limits on the amount of material included to ensure the efficiency of the evaluation process. Suggestions are made for what data can be included, how that data should be structured and presented to avoid bias and superficial treatment. Case examples are included, where possible. Chapter II ends with a discussion of what the construction of a teaching portfolio can accomplish in terms of guiding peer review of teaching, the improvement of teaching practice and the establishment of standards for evaluation.

Chapter III provides some guidelines, based on a synthesis of research findings, for collecting and employing evaluation data from students and peers. It will be useful for department heads and faculty groups seeking guidelines for data collection and their proper use.

Chapter IV includes an in-depth discussion of evaluating a candidate's teaching. As such it will be especially relevant to faculty serving on tenure committees. It begins with a consideration of general criteria relevant to evaluation, followed by an analysis of the evaluation process, including some suggestions for criteria of teaching excellence specific to each data source.

Chapter V broadens the discussion to encompass the relationship between tenure decisions and the improvement of teaching practice. It has been written to assist the collaborative development of teaching between faculty colleagues. It begins with a model of how an individual develops knowledge about teaching through experience and describes how to maximize the development of teaching practice through the evaluation process. Resources available to Cornell faculty members to further assist them in the documentation and evaluation of teaching through the Center for Teaching Excellence are provided.

## Chapter I—Conceptual Overview

During the last 40 years research on teaching effectiveness has shifted away from behavioral 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 40 years ago, which produced many experimental studies, proved less fruitful than was hoped because methodological constraints limited (and probably trivialized) the contributions to understanding how students reach comprehension and application of knowledge taught. The focus on *behaviors*, (rather than the richer and more powerful realm of teachers' and students' *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 relatively recently. The practice of teaching in higher education very strongly reflects the educational system that prepares and qualifies its teachers during their graduate education. Teachers in higher education are, for the most part, educated to be justifiable authorities on the subjects they teach, but only indirectly are they typically educated 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 teaching practices and experimentation, and who were more likely to adopt those stimulating approaches themselves. Recent research suggests students can gain more in terms of deeper learning when they are more actively engaged in the classroom. If we are to educate our students to effectively deal with the complex problems of today's world we must be concerned with the quality of the learning we can stimulate through our teaching approaches.

Over the last 30 years, there has been a significant increase in the concern for the training of graduate students in how to teach, more attention has been placed on the relative importance of undergraduate education, and strong criticism has been levied at the degree to which higher education is adequately educating students (Arum, 2011). All of these factors have contributed to creating an atmosphere where the central administration at Cornell University has supported a refocusing on the quality of teaching.

## 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.<sup>2</sup> 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. Faculty members need to be supported for their on-going efforts at developing their teaching quality. Investment in any efforts to continually develop teaching skill should not automatically imply substandard teaching quality.

However, there is a price to be paid for treating these two evaluation functions separately: the summative evaluation process may become too oriented toward 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.

One of the most problematic areas in the use of ratings is comparison of raw mean scores among faculty or against some other kind of implicit or explicit normative standard. Comparisons are often made explicitly (for example, 3.4 versus 3.5) or implicitly based on the reviewer's personal opinions about where to draw the boundaries (for example, 3.0 is bad, and 3.5 is okay). The difficulty with either approach is that the problem of error in ratings and its effects on ratings means is ignored. (Franklin, 2001)

Regardless of relative experience and skill, everyone has the potential to improve. In fact, a truly scholarly approach to teaching (Becker & Andrews, 2004; McKinney & Cross, 2007) would imply that one is never finished learning about it, just as one may never come to fully understand a phenomena under investigation in research.

---

<sup>2</sup>Scriven, M. (1981). Summative teacher evaluation. In J. Millman (ed.). *Handbook of Teacher Evaluation* (pp. 244). Beverly Hills: Sage Publications.

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 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 on-going improvement of teaching practice.

To achieve efficiency in the evaluation of teaching, specific procedures have been identified (Arreola, 2007; Blackmore, 2005; Chism, 2007) that 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. Peers should conduct summative evaluation of teaching. However faculty peers may not always feel they have the expertise or time necessary to properly conduct evaluations of teaching (Atwood, Taylor & Hutchings, 2000).

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 results to be used to continually evolve it, 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 the point of hire. These expectations should be communicated to faculty members when they are hired so they can begin documenting their efforts right away.

The practice of peer evaluation of teaching should not be limited to classroom observation but should encompass the entire range of teaching activities that take up most of an instructor's time outside of the classroom. These activities need documentation for review from data gathered, that can include student and alumni letters, course teaching materials and measures of student learning. Departmental review committees must establish and maintain guidelines and criteria

for the evaluation of teaching, establish procedures to be followed, and standards to be set (Arreola, 2007). Colleagues can be available to assist the tenure candidate in making his or her case for quality teaching through mentoring efforts during the candidate's early years. Where faculty colleagues do not have the time or expertise to assist in these areas, instructional development resources are available through the Center for Teaching Excellence.

### **The question of defining excellence in teaching**

A frequently raised question both in the literature on the evaluation of teaching (Chism, 2004; Desimone, 2009; Knapper & Cranton, 2001) 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 be necessarily different from those for consideration at the assistant 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, and intrinsic instructional worth to the department and college?

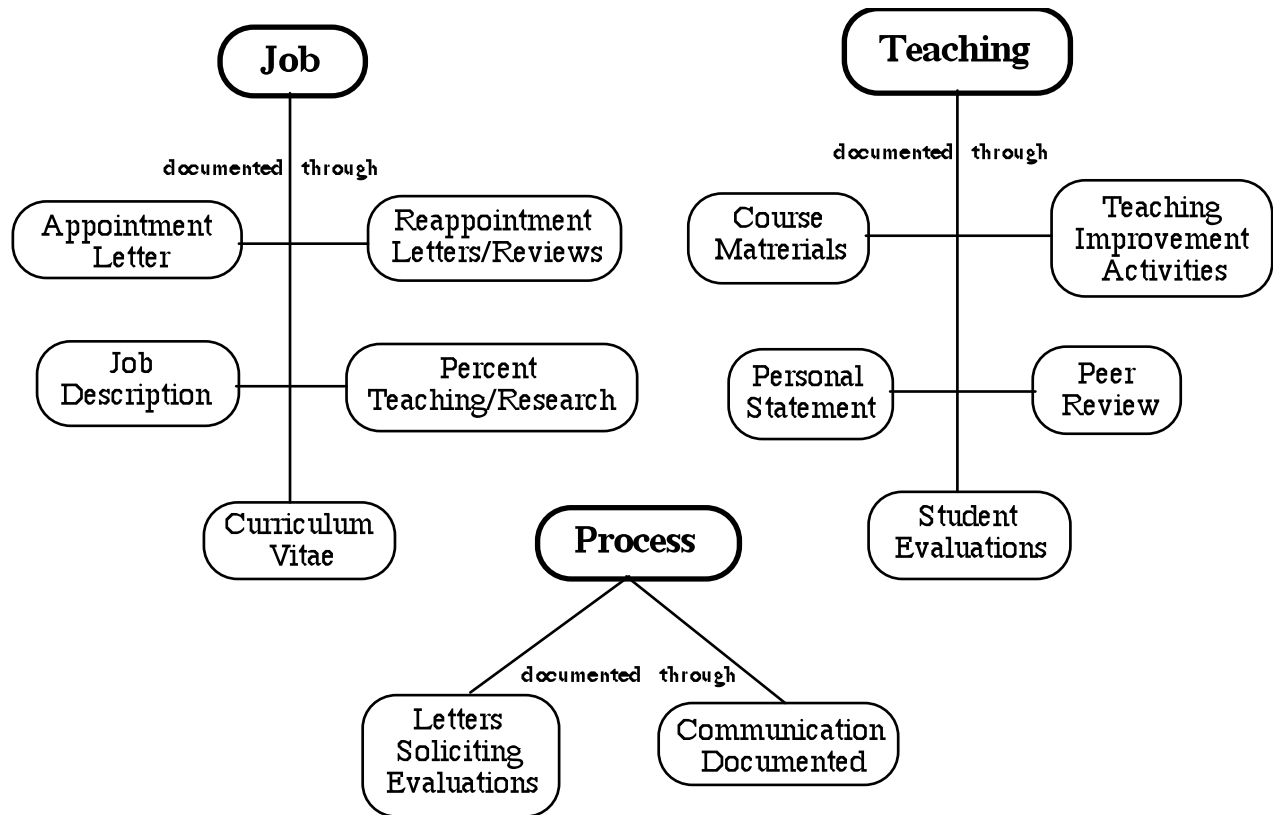
It will be far more difficult to agree upon and evaluate an absolute definition of excellence in teaching than a relative one. Assuming that adequate and appropriate standards of merit have been applied in hiring a candidate, his or her continual worth to the unit will be closely related to the capacity for improvement in performance. The fact that the candidate has been hired to teach at Cornell 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 continual effectiveness based upon a comprehensive range of criteria and data sources (Arreola, 2007). 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 at Cornell have guidelines for evaluating faculty for tenure (Lynn, 1991). Since the Lynn study, the university has created the Faculty Advisory Committee on Tenure Appointments (FACTA) that advises the provost on tenure decisions (<http://theuniversityfaculty.cornell.edu/governance/committees/facta/FACTAlegislatioRev.pdf>). College guidelines can be most helpful if they 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 departmental and college faculty. The roles of everyone involved in a tenure case should be known and spelled out including chairs, *ad hoc* committee members, students, colleagues, dean, chief academic officer (Arreola, 2007; Chism, 2004).

Tenure and promotion files should include the documents that establish the contracts between the chair, the candidate and the department. 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. 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. 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 that explicates his or her efforts at improving teaching effectiveness?

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

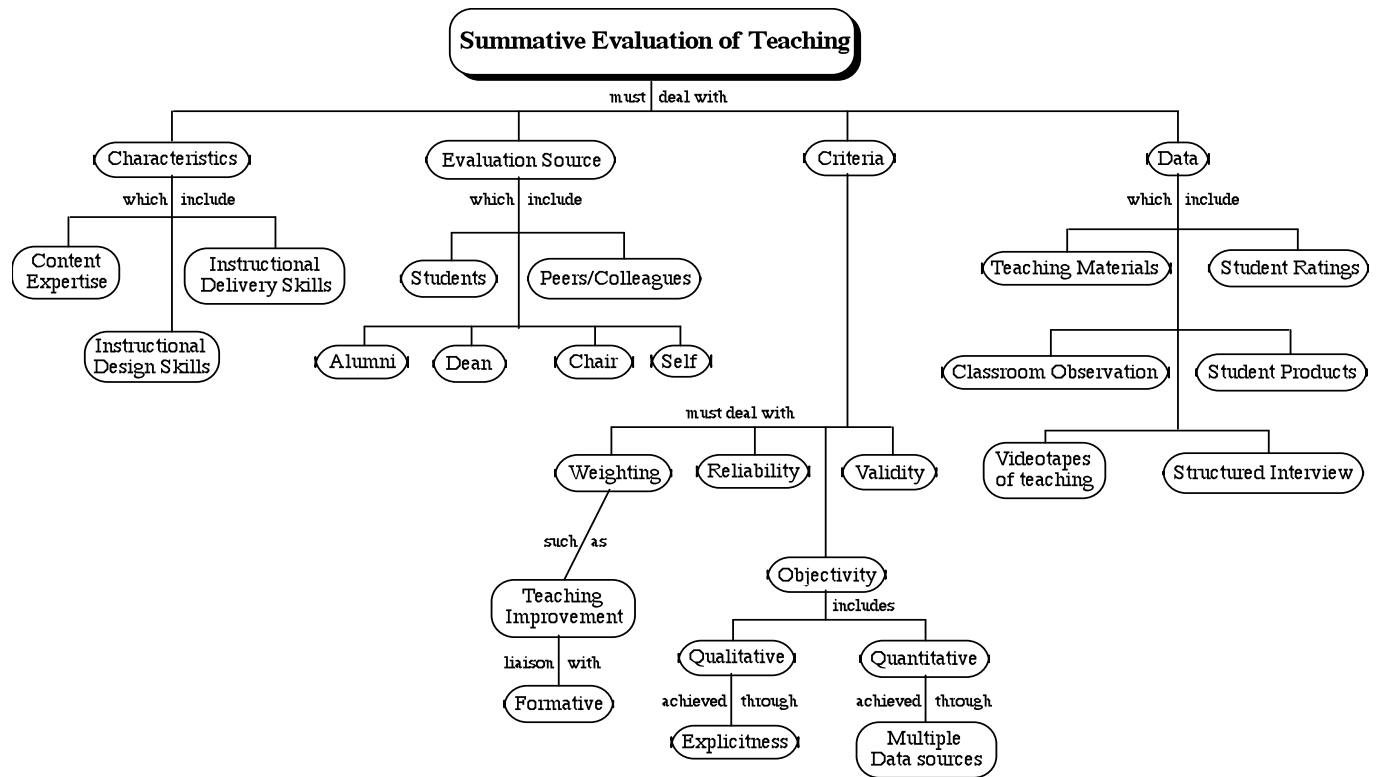


Figure 2

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. Peers in the discipline and faculty colleagues in the department must evaluate content expertise. Peers both within the candidate's department and discipline and outside of it can best evaluate instructional design skills. 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 effectively teaching the content 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 (Franklin, 2001). The weighting attributed to their contributions of evaluating instructional and course design skills is an important consideration.

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 (Berk, 2005). 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. 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 to a representative and manageable form 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 detailed view of what has transpired, but helps in interpreting abstract data, like numerical student evaluation scores (Franklin, 2001; Knapper, 2001).

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 work.)

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 reliable (Arreola, 2007). 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 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 workload assigned? These matters will be addressed further in Chapter IV.

If quantitative objectivity is an important criterion 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? Of primary consideration is the most appropriate material 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, appropriate and effective use of technology, and the choice of texts and course readings. 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 that must be decided on the department and college level. This chapter has brought up some of the major issues that should be considered in setting departmental and college policy. To help make those informed decisions 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 (Edgerton, Hutchings, Quinlan, 1991), in collaboration with some of the leading authors in the area of documentation of teaching—the Teaching Portfolio.

## Chapter II—The Teaching Portfolio: A Model for Documenting Teaching and Its Improvement

This discussion of the documentation of teaching for tenure and promotion is based on the assumption that a tenure file should provide thorough documentation of the process whereby the candidate was evaluated, in terms of both research and teaching. Lack of thorough documentation is a liability to the candidate, department and college, on legal, ethical, intellectual and efficiency grounds. It is recognized that there is much more to the tenure decision process than what is, or can be, put in a file. However, the file is the primary document used to make the candidate's argument for tenure and should therefore reflect a degree of thoroughness and detail sufficient to stand on its own as a source of evidence.

Inclusiveness and brevity are two competing factors that must be considered in documenting teaching and its development. Inclusiveness has to do with whether there is enough data available to all those who must make a decision and whether that data represent the full range of activities and responsibilities associated with the candidate's teaching. Offsetting inclusiveness is the issue of brevity: has the available data been reduced so it is in a manageable and digestible form without biasing or distorting the facts? What format the data is in is another important factor. Different data will require different format guidelines, yet all data are related in various ways and should not be presented in isolation from each other. As suggested in Chapter 1, a range of data sources on a candidate's teaching effectiveness improves the quantitative objectivity by which that candidate is evaluated. If a major criterion by which the candidate is evaluated is the improvement of teaching practice, the *candidate* is one of the best sources of data to document *improvement*.

This chapter will focus on a tool that can be used by the tenure candidate for documenting the improvement of teaching, the "Teaching Portfolio" (Edgerton, Hutchings, Quinlan, 1991; Seldin, 2006; Smith & Tillema, 2007). According to Seldin, the teaching portfolio

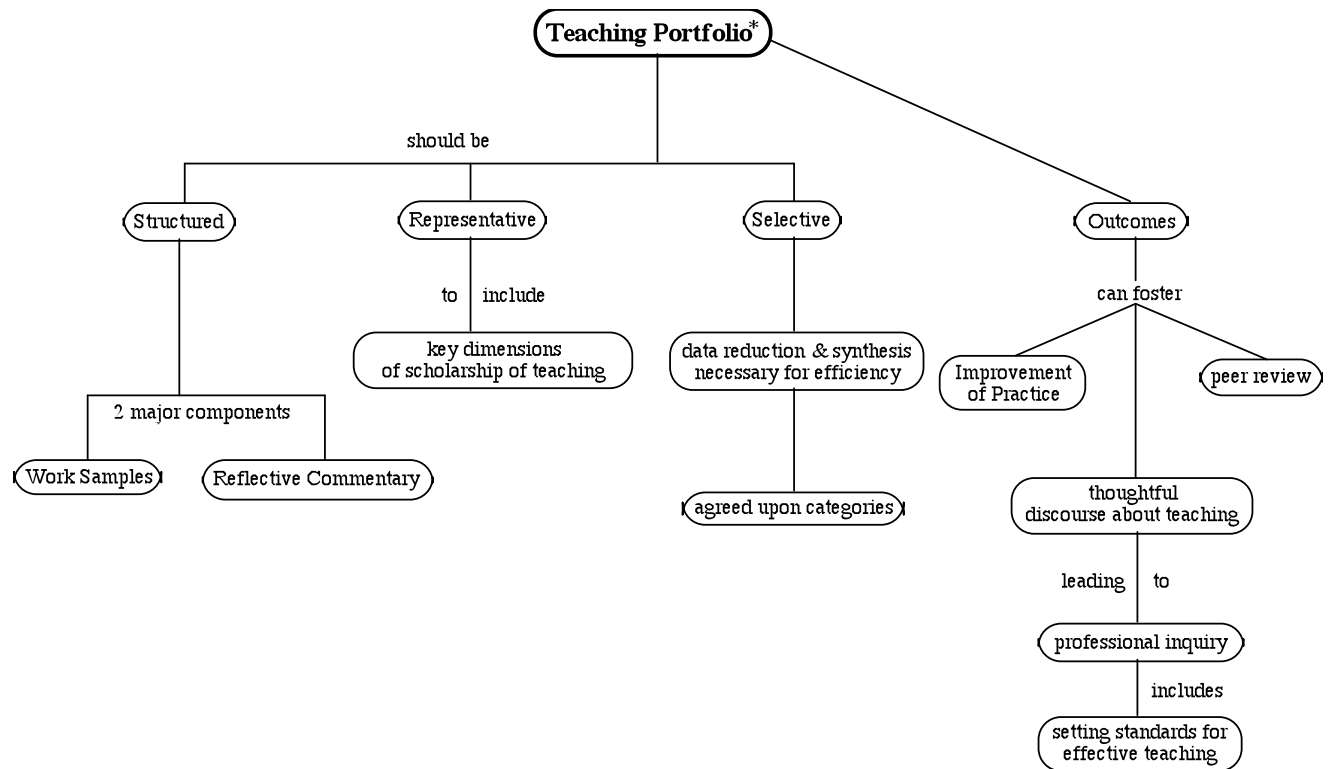
...would enable faculty members to display their teaching accomplishments for examination by others. And, in the process, it would contribute both to sound personnel decisions and to the professional development of individual faculty members....It is a factual description of a professor's major strengths and teaching achievements. It describes documents and materials which collectively suggest the scope and quality of a professor's teaching performance..<sup>3</sup>

---

<sup>3</sup>Seldin, P. (1991). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions*. Bolton, MA.: Anker Publishing Co., 3.

The Teaching Portfolio has been chosen as a model in this handbook because it connects summative and formative evaluation functions in a single process, it honors teaching as a scholarly activity, it is a practical and efficient way to document teaching and its development over time, and it has a growing history of use at a variety of higher education institutions. The construction of a teaching portfolio raises issues and questions that must be considered by the candidate and administrators engaged in the evaluation of teaching. Figure 3 illustrates what a Teaching Portfolio is and can do.





\*adapted from Edgerton, Hutchings & Quinlan, AAHE, 1991

Figure 3

The teaching portfolio should be representative enough that the key dimensions of teaching as a scholarly activity are evident.

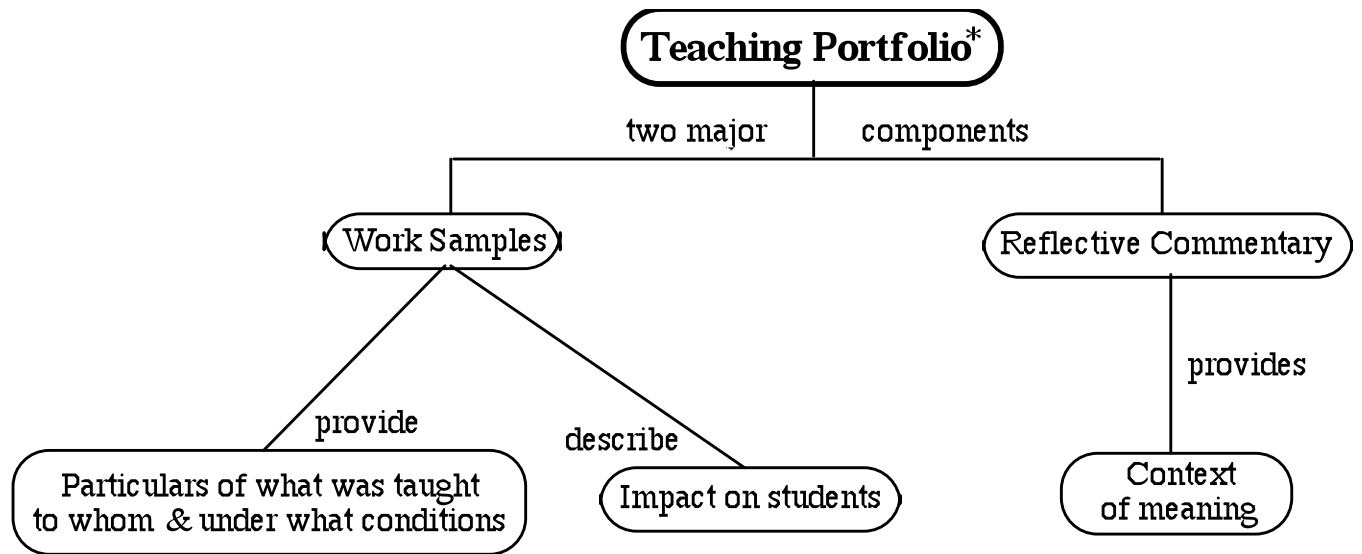
When defined as scholarship...teaching both educates and entices future scholars....As a *scholarly* enterprise, teaching begins with what the teacher knows. Those who teach must, above all, be well informed, and steeped in the knowledge of their fields....Teaching is also a dynamic endeavor involving all the analogies, metaphors, and images that build bridges between the teacher's understanding and the student's learning. Pedagogical procedures must be carefully planned, continuously examined, and relate directly to the subject taught....teaching, at its best, means not only transmitting knowledge, but *transforming* and *extending* it as well.<sup>4</sup>

When so conceived teaching encompasses both a logical dimension—in the selection and representation of the subject matter, and a psychological dimension—in the consideration of the process of structuring and translating that subject matter into a form which is accessible to students.

A teaching portfolio has both representative breadth while it is also selective. Criteria for inclusiveness must be established which limit the bulk and form of data to a manageable amount. The selection process should preserve the criteria of representativeness of the primary teaching responsibilities, yet should reduce and transform the available data into a manageable form that insures efficiency during the subsequent evaluation process. Selectivity is governed by structuring the portfolio into 2 major components (see Figure 4 below). The first of these are *work samples* or what some have called “artifacts of teaching”(<http://depts.washington.edu/gs630/Spring/STEMStrategies.pdf>) which consist of the details of what was taught and what its impact was on students. These are often structured into appendices. The major component of the portfolio is its *reflective commentary* or narrative account (Franklin, 2001) that extends the meaning of the work samples selected by providing a context in which to comprehend their design and choice from the teacher's own point of view.

---

<sup>4</sup>Boyer, E. (1990). *Scholarship reconsidered*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching, 23-24.



\*adapted from Edgerton, Hutchings & Quinlan, AAHE, 1991

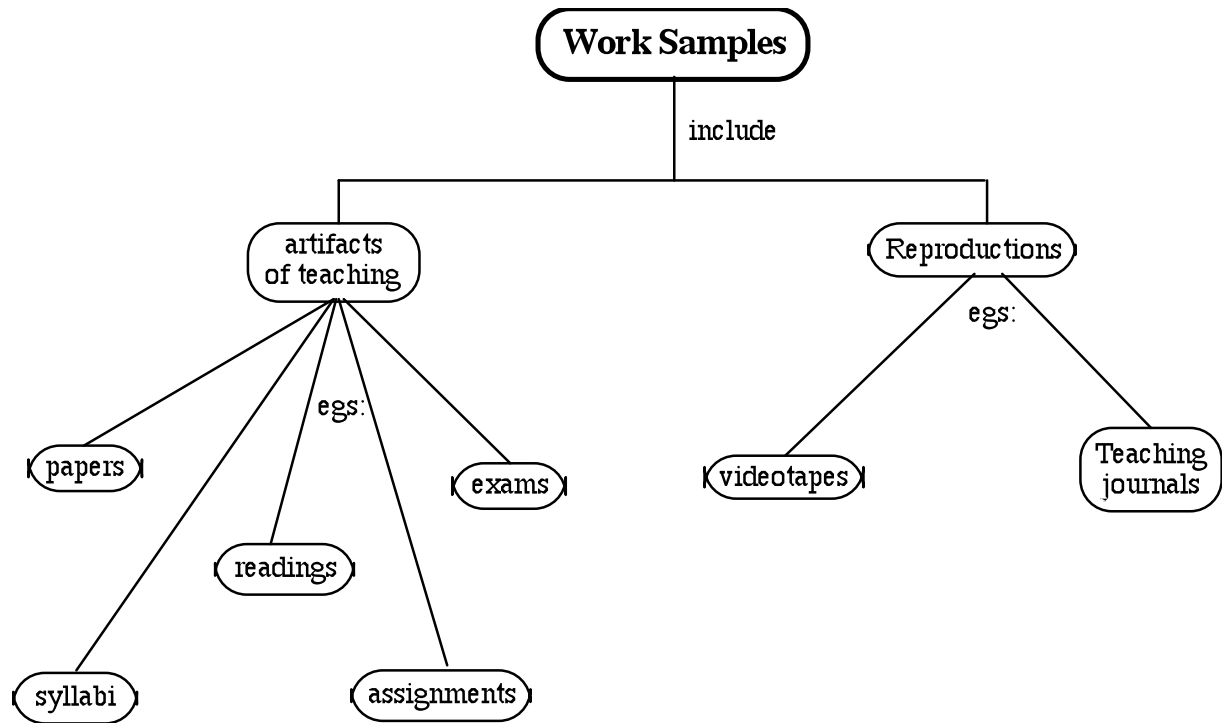
Figure 4

Work samples (see Figure 5 below) constitute direct evidence of teaching such as facts, objects, and reproductions of events from daily practice. Work samples should be selected which “...highlight what is unique about an individual’s approach to teaching.”<sup>5</sup> Just what samples are selected must be negotiated between the candidate and department. From the candidate’s point of view, selection will most probably be governed by an intimate knowledge of what was done, its effect, and how it changed over time. Department decisions about what constitutes adequate sampling will be governed by a consensus of the key scholarly dimensions of teaching in that field.

According to Edgerton, Hutchings and Quinlan (1991), the work samples are artifacts of teaching performance, while the reflective commentary that accompanies each artifact provides the teacher’s rationale for using that artifact, how it was developed and what its impact was on students’ learning. The reflective component of the portfolio is a kind of annotation to each sample of work. While teaching consists of both behavioral and cognitive aspects, there are also ethical aspects that somehow must be discussed. These can be documented in the reflective part of the portfolio and include discussions of what ethical principles and values guided the candidate’s approach to key decisions made about teaching and its improvement, like how diversity factors were dealt with in both course design and in interactions with students.

---

<sup>5</sup>Edgerton, R. (1991). The teaching portfolio as a display of best work. Paper presented at the National Conference of the American Association for Higher Education, Washington, D.C.



\*adapted from Edgerton, Hutchings & Quinlan, AAHE, 1991

Figure 5

The following is an example of how work samples can be connected to a reflective statement.

Artifact: Class syllabus [A copy is included in the dossier.]

Reflective annotation: Course goals, evaluation procedures (and dates), test(s), office hours, small group feedback sessions, reading assignments and a schedule of lectures are listed. The innovative section of the syllabus includes: (a) expectations that students will develop and use effective thinking skills (See Small Group Problem Sets, and Goals of Small Group Work on p.2) ; (b) rationale for small group work on problem sets (see Small Group Problem Sets, Goals of Small Group Work, and Calculation of Problem Set Grades on pp. 2-3); (c) an explanation of effective thinking and problem solving skills (See The Process of Effective Thinking and Problem Solving on pp. 3-4); and (d) purpose of the research project and benefits to students from participation (See The Motivated Strategies for Learning Questionnaires [MSLQ] and the “ordered tree” of concepts on p. 5).

1990/91: Examples of problem sets and answer keys.

Problem sets are designed to give students an opportunity to use their knowledge of economics to solve problems and reach conclusions. Students are specifically asked to follow explicit problem solving steps (e.g., Guided Design steps) which have been shown to be helpful for novice solvers. [Examples are included in the dossier.]

1988 to present: Tests with higher and lower order thinking skill requirements marked to evaluate student learning problems.

All tests questions are indicated as HO, for higher order thinking skills required, or LO, for lower order thinking skills required to reach a correct answer. This technique (a) continues the rationale of naming thinking skills when they are useful and (b) helps students learn how to study in the future. For instance, if students miss LO questions it may mean that they are not reading the text carefully enough. If students miss HO questions it may mean that they are memorizing and not learning how to use their knowledge. If students miss HO questions they are directed to certain areas of the Study Guide and problems which require students to practice using their knowledge. If students miss LO questions it is suggested they read the text assignments more carefully. [Copies are included in the dossier.]

To document teaching on this level will require departments and colleges to agree upon categories and key dimensions reflecting the scholarship of teaching. This, in turn, may require faculty and administrators to examine the roles they play in the summative evaluation of teaching. The candidate must play a very active role in monitoring his or her teaching, while colleagues must play a collaborative role. The collaborative construction of a teaching portfolio between

departmental colleagues connects the summative and formative evaluation functions together in a single process, since the decisions made by the department faculty in determining policy on what is selected for inclusion and how it is structured in a portfolio (summative functions) can foster the improvement of practice (formative function). These decisions will of necessity require reflective discussions about teaching between the candidate, his or her peers, chairperson and dean. The intention is that the activity of building a teaching portfolio during the first six years of teaching practice encourages peer mentoring and review, resulting in a profile of how the candidate's teaching has developed over that period of time. This can itself lead to a kind of professional inquiry since, after enough candidates have undergone the process, it is likely that a clearer set of standards for what constitutes effective teaching may emerge.

The following is an example of what a Teaching Portfolio might consist of, together with a faculty member's rationale for how it was constructed.

### **Portfolio Outline<sup>6</sup>**

Professor Margaret Ackman  
College of Pharmacy, Dalhousie University

#### **A. Statement of Teaching Responsibilities**

1. Courses Taught
2. Student Advising
  - a. individual students
  - b. student committees
3. Practica: Organized and Supervised

#### **B. Statement of My Teaching Philosophy and Goals**

#### **C. Efforts to Improve Teaching**

1. Formal Courses in Education
2. Conferences Attended
3. Workshops Attended
4. Participation in Peer Consultation

#### **D. Redevelopment of Existing Courses**

1. Addition of tutorials, role-playing, case studies, etc.
2. Incorporation of Writing Skills
3. Incorporation of Oral Presentation Skills
4. Appendix of Representative Course Syllabi and Assignments

---

<sup>6</sup>O'Neil, C., & Wright, A. (1992). Recording teaching accomplishment: A Dalhousie guide to the teaching dossier. ( 2nd ed.). Dalhousie University, Halifax, Nova Scotia: Office of Instructional Development and Technology, 66-67.

E. Information from Students

1. Summary of Student Ratings
2. Comments from Student Committees Regarding Advising

G. Service to Teaching

1. Evaluating Term Papers, Chair
2. Faculty Evaluation, Co-Chair
3. Curriculum Committee, Member
4. Lectures to Special Interest Groups of the Public

H. Information from Colleagues

I. Information from Other Sources

1. Guest Lectures to Other Faculties
2. Continuing Education Lectures for Peers
3. Lectures to Special Interest Groups of the Public

J. Future Teaching Goals

*Professor Ackman provided the following explanation of her plan to prepare her teaching dossier.*

I began my teaching dossier with a statement of my teaching responsibilities. This section would discuss not only the courses which I teach, but also the student advising which I do for both individual students and student committees. Our students are required to complete a variety of practicums during their undergraduate education. Therefore, my organization and supervision of these practicums would also be included. The second section would be my Statement of Teaching Philosophy and Goals. The rest of the dossier will complement and enhance this statement and provide a framework of how I intend to achieve my goals.

This is only my third year of teaching. As a result, I am very conscious of my lack of experience and formal education in teaching. I have spent a great deal of time and effort trying to improve my teaching and learn more about teaching and learning. I feel that this ongoing effort is very important and that my teaching has improved because of it. Therefore, this would be presented as one of the initial sections of my dossier—Efforts to Improve Teaching. I would follow this section with the section on Redevelopment of Existing Courses. For me, this is a very logical progression. When I began teaching, I was presented with an existing course which I did not feel contributed to the goals and objectives of the College. However, it was only after attending a number of workshops and conferences that I was able to begin redesigning the course.

The section, Information from Students, would follow. This would indicate improvement in overall ratings since my initial attempts at teaching. This would



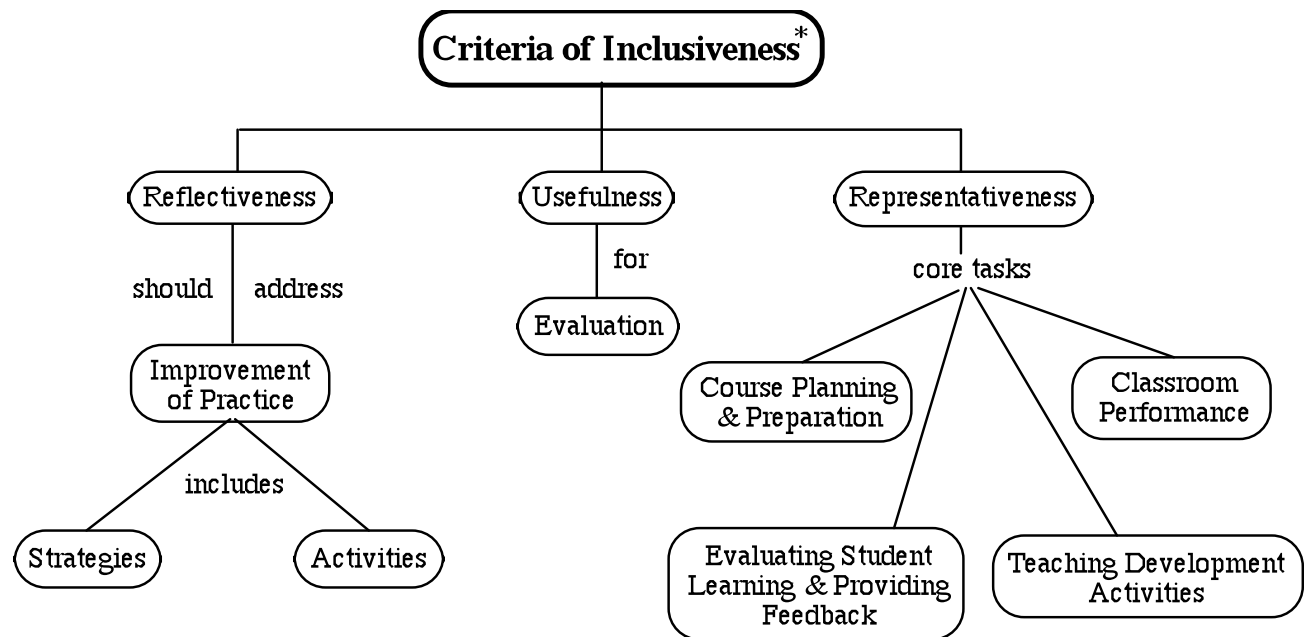
relate to Efforts to Improve Teaching and also to Redeveloping of Existing Courses. A section on Student Achievement would be next. This would indicate that my students do reasonably well on both regional and national levels in comparison to their peers.

Service to Teaching would be the next section. This also relates to the section on Efforts to Improve Teaching. As I have become better informed concerning teaching, I have become more involved in the administrative aspects of teaching. This is reflected by the committees which I currently chair or of which I am an active member. Information from Colleagues would fit best at this point in my dossier. My colleagues would be in a position to comment not only on my teaching abilities, but also on my Service to Teaching. Information From Other Sources is a miscellaneous category. It is important because it demonstrates my degree of involvement with the profession and my ability to teach at a variety of levels. However, for evaluative purpose, it should not be given the same weight as student evaluations. Therefore, it would be included near the end.

The final section would be my Future Teaching Goals. This section may actually be one of the most important in my dossier because I am relatively new to teaching. I considered placing this section after My Teaching Philosophy and Goals. However, since it would relate to all of the other material in the dossier, it is only reasonable to conclude with this information. In order to ensure that this section would be given the appropriate consideration, I would refer to it in My Teaching Philosophy and Goals and in other pertinent sections.

## **Criteria of Inclusiveness**

Three criteria for inclusiveness have been suggested for the teaching portfolio: reflectivity, usefulness and representativeness, as shown in Figure 6 below.



\*adapted from Edgerton, Hutchings & Quinlan, AAHE, 1991

Figure 6

The simplest and most obvious of these criteria is usefulness: are the data chosen for inclusion—and the way they have been structured and presented—useful for the purposes of evaluation? The framework presented in Figure 1, Chapter 1, suggests three areas of data useful in documenting the position and how effectively the candidate filled it: *the job*, how it was described and how the candidate's filling the job subsequently shaped it, documentation of *the process* whereby the job was filled by the candidate, and *teaching data*. These kinds of data are useful in providing an overall profile of the relative weights to be attributed to the candidate's research and teaching data.

The second criterion for inclusiveness in a portfolio is reflectivity: to what degree are artifacts, such as course materials and other abstract data such as student evaluation scores, accompanied by a reflective narrative by the candidate which ground them in a meaningful context and help their interpretation by evaluators? Reflectivity (Brookfield, 1995; Brookfield & Preskill 2008) helps those evaluating the portfolio to understand the activities and strategies undertaken by the candidate to improve teaching practice and the data presented to demonstrate that improvement. Reflections on the efforts of improving teaching can include statements about what experience has taught the candidate about teaching, what he or she has worked on changing, what experimental actions were taken to effect change, and what change was accomplished, either intentionally or unintentionally through gathered evidence. The following two examples contrast a more descriptive personal statement about teaching (Example 1) with a more reflective statement (Example 2).

### **EXAMPLE 1<sup>7</sup>**

Dr. Thomas H. MacRae  
Department of Biology, Dalhousie University

#### **Teaching Advanced Students**

Eleven honours students, eight of whom were NSERC Undergraduate Scholars, have studied in my laboratory (Appendix I) and I have served on several other honours committees. Mr. John Apple, who spent two years with me, is a co-author on two papers. Mr. Bill Brown is co-author on two abstracts and three papers now submitted. Ms. Ann Christopher will be co-author on one paper. One other student was co-author of an abstract and another, an NSERC Undergraduate Scholar from McMaster University who worked in my laboratory, is now doing her Ph.D. under my direction. I have supervised several undergraduate students in research and/or writing-based special topics courses. I normally support only

---

<sup>7</sup>O'Neil, & Wright, 1992, pp. 41.

one undergraduate application per year for an NSERC Summer Scholarship, although in 1988 I supported three students all of whom received awards. I am very selective of the undergraduates I accept for honours, usually taking only those who can generate sufficient money to allow them to work during the summer. This approach is necessary as the time required to master technical aspects of the projects in my laboratory is usually great. The student thus has the time to undertake a meaningful project, is subject to closer supervision during the initial phase of his/her work since I do not lecture in the summer, and develops a truer appreciation of research as the distraction of course work is lessened.

As my first two years at Dalhousie were on a term appointment, I was not able to supervise graduate students until my third year. Since 1982, two MSc students and one Ph.D student have graduated from my laboratory, while two other graduate students are in progress (Appendix II). Two of the students have held NSERC Graduate Scholarships.

### **EXAMPLE 2<sup>8</sup>**

Dr. Graham J. Fishburne  
Department of Elementary Education, University of Alberta

#### **Commitment to the Improvement of University Teaching**

At the outset of this report, I attempted to articulate my own philosophy of teaching. I believe that all University Professors are working toward the common goal of educating students in the most effective manner possible. We all possess skills and knowledge that are needed to achieve this common goal. By sharing our skills and knowledge, we will learn from each other and improve the effectiveness of the instruction that we offer to our students.

Since entering into higher education teaching I have spent a great deal of my time in shared and cooperative ventures aimed at improving the quality of University teaching. These activities have been above and beyond my own areas of research and the academic discipline for which I was hired....It would have been easy for me to remain within the domain of the areas of research enquiry that I have established and within the academic discipline for which I was hired; however, because of my commitment to University teaching, I have endeavored throughout my career to share and assist in any way possible my knowledge of teaching and learning with colleagues who are also involved in the world of instruction:

“As a teaching colleague, Dr. Fishburne...(was not only open to new approaches to instruction but also) contributed many creative and innovative ideas...He is an extraordinarily competent, effective teacher. He has been actively involved in departmental affairs, and takes seriously

---

<sup>8</sup>O'Neil, & Wright, 1992, pp. 51-52.

his obligation to be of service to the University, the profession, and the public.” Professor of Mathematics Education, 1985.

Since joining the Faculty of Education, I have been involved in the development of integrated courses. This has led to team teaching experiences where several professors team teach and share teaching experiences. Over the years, I have been involved as a team leader in such courses. I have found this experience to be extremely valuable since we must plan and share our various teaching methods as we strive for the most effective learning environment. We engage in a great deal of “reflective practice” and learn much from each other. The teams regularly change which has allowed me the opportunity to play a leading role in peer teaching with many of my colleagues. Due to the shared learning commitment the Faculty enjoys with the Alberta Teachers’ Association, we regularly have seconded staff from various school boards involved in our undergraduate program. My leadership role has been that of “mentor” to these people and to new members of staff. Over the years I have team taught with approximately 30-35 University teachers. Feedback on these mentoring experiences has been very positive as the following peer comments attest:

“Dr. Fishburne’s teaching is exemplary. Over the years we taught on the same team, I had the unique opportunity to observe and evaluate his teaching as well as his contributions in collaborative program development. His teaching is ‘dynamic’ He is enthusiastic about his work and demanding in his expectations of students and colleagues. I consider him to be a model of professionalism-one to whom students and colleagues look for strength in the teaching role and for guidance in their own development. His general support and encouragement is always a positive influence on both students and colleagues.” Professor of Mathematics Education, 1987.

Over the past few years I have been consistently involved with the University Committee for the Improvement of Teaching and Learning (CITL). I have made regular presentations to University colleagues on issues aimed at supporting campus who frequently call for advice, research findings, resource material, and general information on effective teaching.

I am also a member of CITL’s Peer Consultation Program. This involves working on CITL’s Peer Consultation Team with University Professors from Faculties other than my own who have requested some form of help or assistance with their University teaching.

I am constantly involved in Peer Teaching Evaluation. I not only work with colleagues in my own Faculty but I am requested by staff in other Faculties to conduct peer teaching evaluations. I not only offer a perspective on the colleague’s teaching but frequently work with the colleague to help improve their teaching. I have performed this voluntary activity for the past 10 years and the feedback from colleagues has suggested that it has been a worthwhile endeavour

for improving University teaching. A typical response to my work in this area appears below:

“I have often recommended to professors needing teaching evaluations that they seek out Dr. Fishburne because he is not only quick to render expert judgement but he continues to help professors improve their teaching skills. He shares information and material willingly with colleagues and provides help whenever he can...” Chair and Professor, Department of Elementary Education.

Some examples of questions a candidate can provide reflective responses to include:

#### Discipline and Classroom Approach

Within your discipline, which area do you regard as your strongest? Your weakest?

What is your greatest asset as a classroom teacher? Your greatest shortcoming?

Which teaching approach works best for your discipline? Why?

Do you change methods to meet new classroom situations? Can you give a recent example?

What is your primary goal with respect to your students?

How would you describe the atmosphere in your classroom? Are you satisfied with it?

#### Knowledge of Subject Matter

In what ways have you tried to stay current in the subjects you teach?

How would you judge your knowledge in the subjects you teach?

Do you think your colleagues agree with that judgment?

What have you done or could you do to broaden and deepen your knowledge of the subject matter?

#### Questions About Teaching

What is the one thing that you would most like to change about your teaching?

What have you done about changing it?

What would you most like your students to remember about you as a teacher ten years from now?

Overall, how effective do you think you are as a teacher? Would your colleagues agree? Your students?

Which courses do you teach most effectively?

In what way has your teaching changed in the last five years? Ten years? Are these changes for the better? Why or why not?

The Stanford Teacher Assessment Project (King, 1990) suggests four core tasks of teaching: a) planning and preparation, b) performance, c) evaluating student learning and providing them feedback on their learning, and d) teaching development activities. Together, these constitute the third criterion of inclusiveness of the portfolio: does the information provided adequately represent the range of activities of a candidate's teaching?

According to Edgerton, Hutchings and Quinlan (1991) course planning and preparation can include such artifacts as course syllabi, a series of assignments given to students, readings and teaching plans. For the reviewer to gain a sense of development in practice, the selection of these planning and preparation artifacts should be governed by how they have changed over time. Thus, all artifacts may not necessarily be exemplary cases but serve to show how the candidate has developed in various areas. The reflective commentary accompanying these artifacts will help the reviewers interpret their relative meaning and value.

Classroom performance can be represented by peer observation reports, videotapes of selected classes taught by the candidate, and student evaluation results. If videotapes are chosen for inclusion, the reflective commentary accompanying them must provide a rationale for why and how those teaching episodes were selected for inclusion. Peer observation of classroom performance is a practice that must be carried out with great care if it is to be useful and just. Peer observers must be qualified in terms of the subject matter being taught and in terms of what to observe and report on. In most cases peer observers must be trained in carrying out their responsibilities (Chism, 2007). The candidate can help interpreting observational reports through a reflective commentary with reactions to what was written by observers. Peer observation will be discussed in more detail in Chapter IV.

Documenting evaluation of student learning and providing students with feedback on their progress might include selecting a student paper as evidence of a high standard with reflective comments accompanying the paper addressing why that standard was appropriate, together with an explication of what was done to help that student achieve such a high standard. Another work sample in this area might be a copy of a student's exam which reflects misconceptions students often bring to the course, accompanied by comments on what strategies were developed and used to deal with those misconceptions and a rationale for evaluating the effectiveness of those strategies.

Through the trials and errors of teaching much is learned about effective practice. Most academic fields have some journal devoted to publishing faculty studies on their teaching experiences. Keeping up with a professional field's knowledge base of teaching, or contributing to that body of knowledge is a legitimate part of a faculty member's teaching responsibilities and can be documented in a portfolio (McKinney & Cross, 2007). At the very least, the candidate could include a paper heard at a professional meeting or read in the professional press, accompanied by a reflective essay on how it influenced changes in course design or teaching practice.

General guidelines for constructing Teaching Portfolios suggested by Edgerton, et al., include:

- minimize size and length through the following question, “What will this entry add to the description of knowledge, skills and perspective of the candidate?”
- orient the portfolio away from raw data and move toward judgments of data’s meaning
- maintain an attitude of flexibility and experimentation while lessons about function and process are being learned.
- think about portfolios and the entire evaluation process as connected to the improvement of practice and toward “developing a more professional discourse about teaching”
- work involved in developing a portfolio must be considered an integral part of faculty’s responsibility, both for the candidate developing it and for colleagues evaluating it<sup>9</sup>

A teaching portfolio can clearly reflect a candidate’s teaching and teaching development through the prudent selection of included work samples. Selection of content should do justice to the range of teaching responsibilities and activities engaged in by the candidate without resulting in a cumbersome, redundant and lengthy document for others to wade through. The key in preserving representativeness while restricting bulk is in the establishment of a clearly articulated set of criteria and categories that should be developed by department faculty. Once these have been established and the data categories selected, there are fundamental principles that have been developed through research and experimentation that can govern the collection and presentation of that content. We turn now to a detailed discussion of data collection and its representation.

---

<sup>9</sup>adapted from Edgerton, R., Hutchings, P., & Quinlan, K. (1991). *The teaching portfolio: Capturing the scholarship in teaching*. Washington, D.C.: American Association for Higher Education.



## Chapter III—Supporting Data: Collection & Presentation

To increase the efficiency of the review process for all involved, it is imperative that the supporting data presented be collected in a way that reduces bias, is representative, and yields valid assessment. It is also important to understand that the manner in which the data is presented can itself bias the evaluative outcome. Fortunately this is an area that has stimulated a very prolific amount of research. This chapter will provide a synthesis of some of that work, focusing on two major sources of evaluation data: students and peers.

### Use of Student Evaluation Data

It has been argued that students are not valid sources of evaluation information, that their numerical and written responses on questionnaires used to make tenure and promotion decisions are based on superficial criteria, like appearance and popularity. This assumption has not been empirically supported (Arreola, 2007).

Based on the findings of the meta-analysis, we can safely say that student ratings of instruction are a valid index of instructional effectiveness. Students do a pretty good job of distinguishing among teachers on the basis of how much they have learned. Thus, the present study lends support to the use of ratings as one component in the evaluation of teaching effectiveness. Both administrators and faculty should feel secure that to some extent ratings reflect an instructor's impact on students.<sup>10</sup>

Key in the use of student data is the notion that as a data source, it is only one component which is available for a committee to make an informed judgment. When incorporated into a thorough analysis, student evaluation data is useful not only because it represents the learner's perspective, but it can stand to round off a picture of a candidate's teaching quality when presented in relationship to peer data and data supplied from the candidate's own perspective.

Student data can be solicited and presented in both quantitative and qualitative forms. Quantitative data, in the form of numerical student evaluation questionnaire scores, was the more prevalent form that appeared in the tenure files that were analyzed during the preparation of the report *Evaluation and Recognition of Teaching* (Lynn, 1992). Qualitative student evaluation data, in the form of written letters of evaluation, appeared in less

---

<sup>10</sup>Cohen, P. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, (51), 3, 305.

abundance, although when such data was provided it took up a lot of space in the file, as in one case where a tenure file contained 88 student evaluation letters.

Student questionnaire items must be validated to insure what was being measured by the question is, in fact, what the question purports to measure. Quantitative data such as student evaluation scores can be very efficient in that many teaching factors from many individual perspectives can be presented in relatively little space. Student evaluations of teaching are the most commonly used data source in the evaluation of teaching both at Cornell and elsewhere in higher education. However, there are problems with this practice:

- They constitute a limited data source
- The conditions under which they are filled out can introduce bias
- There are limits to what they can measure
- No set of numerical values can be sufficient as the sole indicator of teaching effectiveness.

Interpreting results:

- Look for patterns over time (standards for normative comparisons)
- Focus on the highest and lowest rated items. Do they represent a pattern over time?
- Keep in mind that a specific sample is not necessarily representative
- Don't over interpret small variances in median scores
- Don't use university-wide, or even college-wide norms for normative comparisons; determine discipline-based norms
- Focus on overall items: "Overall, how would you rate the quality of this person's teaching."
- Look for patterns in student comments; don't over generalize from one student's comment
- Consider that a higher percentage of student comments will come from very satisfied or very dissatisfied students
- Ask whether the written comments help explain the numeric scores
- Take context and experience level into account
- Seek confidential consultation advice to interpret scores from a staff member in the Center for Teaching Excellence
- Expand your sources of student feedback by using regular mid-term evaluations with an on-line tool such as Qualtrics (<https://sri.cornell.edu/SRI/services.qualtrics.cfm>) to generate frequent discussions with your students regarding the quality of their experience in the course and to guide your responses while they are still engaged in the course
- Consider to what degree is there evidence that the instructor has taken student teaching evaluations to heart and used them to improve practice

### Instrumentation:

Research in the area of student evaluation of instruction has resulted in the publication of over 3,000 studies. Much has been learned about proper questionnaire design. One finding is that the purpose of the evaluation should determine the format and kinds of questions included in the evaluation instrument. In general, summative evaluation questionnaires designed for tenure and promotion decisions that are traditionally given at the end of a student's enrollment in a course contain fewer items than formative questionnaires that may be carried out earlier in the student's course experience. Summative instruments focus on global items ("Overall, how would you rate the quality of the instructor's teaching?") and use evaluative scales (Excellent, Good, Fair, Poor... or Strongly Agree...Strongly Disagree) whereas formative questionnaires focus on behaviors (Frequently, Somewhat Frequently, Rarely, Never...) or other aspects that are flexible or adaptable ("To what degree are the clicker questions helpful?")

The use of "core" items can allow an individual's scores to be compared to those determined from a group aggregation, such as within a department or field, or across a college's faculty. Core items are more generic aspects of teaching that are not influenced as much by course design or size. Core items enable the development of normative scores so an individual can be validly compared to his or her peers. Examples of such core items that have been validated through controlled quantitative methods include:

The instructor is well prepared for class.  
The instructor has a thorough knowledge of the subject.  
The instructor communicated his/her subject well.  
The instructor stimulated interest in the course subject.  
The instructor demonstrates a favorable attitude towards students.  
The instructor shows enthusiasm while teaching.

### Administration of Questionnaires:

Research on questionnaire validity suggests that if the following guidelines are followed for administering end-of-semester evaluation questionnaires, reliability and validity of results will be improved.

- response format should be clear and consistent
- students should remain anonymous
- students should be given adequate time to complete the questionnaire
- students should not be allowed to discuss their ratings while they are being administered
- questionnaires should be administered during the last 2 weeks of semester (but not the last day and not during or after an exam)

- someone other than the one being evaluated should administer the questionnaire, or at the very least, the one being evaluated should leave the room
- a student should collect the questionnaires and mail them to an independent office for scoring
- 80% minimum attendance of the student population in a course is necessary on the day an evaluation is administered
- don't use a numeric questionnaire in courses with fewer than 10 students (use open-ended, written response items instead.)<sup>11</sup>

For an in-depth discussion of the common problems with student evaluation of teaching data, we recommend the following article (Franklin, 2001):

<http://onlinelibrary.wiley.com/doi/10.1002/tl.v2001:88/issuetoc>

### Reporting scores:

How summative evaluation scores are reported in a tenure file or in the tenure/promotion process can bias that process, either positively or negatively. Some general principles for proper questionnaire score reporting include:

- report frequency distribution for each item
- don't carry mean scores beyond one decimal place
- multiple sets of scores should be provided for each type of course (survey, lab, seminar) and collected *over a period of time*
- narrative (qualitative) data from the candidate, colleagues or chair about the contextual circumstances of the quantitative student rating scores is an aid in their interpretation.
- normative data sets should be established yearly for course type (elective, required, lecture, lab, etc.) on both a department level and college level if comparison with a tenure candidate's own scores is departmental policy.
- appropriate normative data should be provided wherever possible

In addition, the practice of comparing individual scores to a department's aggregate scores requires that confidence intervals be determined for the overall department faculty to better interpret where an individual's score falls (Franklin, 2001).

Figure 7 below is an example of a simple format for reporting student evaluations scores for a single course. Figure 8 is an example of a visually clear way of reporting a candidate's relative standing in relation to departmental normative data.

---

<sup>11</sup>Sell, G. R. & Chism, N. (1988). *Assessing teaching effectiveness for promotion and tenure: A compendium of reference materials*, Center for Teaching Excellence, Columbus, Ohio: Ohio State University, 5-6.

Instructor: Professor X		Course: 100, sec 22	Spring '92	# choosing each response					
Item	Mean	Standard Deviation	Count	Strongly disagree	disagree	neutral opinion	agree	strongly agree	
1. well prepared for class	4.6	0.8	27	0	1	1	7	18	
2. knowledge of the subject	4.1	0.5	27	0	0	2	19	6	
3. communication effectiveness	4.6	0.6	27	0	0	1	10	16	
4. stimulated interest in course subject	4.5	0.6	27	0	0	1	11	15	
5. one of the best Cornell Professors	4.5	0.7	27	0	0	3	8	16	
6. clearly interprets abstract ideas and theories	4.4	0.6	27	0	0	1	13	13	
7. favorable attitude toward students	4.5	0.5	27	0	0	0	14	13	
8. willing to experiment & flexible	4.4	0.6	27	0	0	2	12	13	
9. encourages students to think for themselves	4.3	0.6	27	0	0	2	15	10	
10. You found the course intellectually challenging and stimulating?	3.8	0.8	27	0	2	6	14	5	
11. Instructor's explanations were clear.	4.4	0.6	27	0	0	1	15	11	
12. Instructor presented the point of view other than his/her own when appropriate	4.1	0.9	27	0	1	6	10	10	
13. Feedback on examinations/graded material was valuable.	4.4	0.9	27	1	0	2	7	17	
14. Methods of evaluating student were fair and appropriate.	4.1	0.7	27	0	0	4	15	8	
15. Instructor gave lectures that facilitated taking notes.	3.8	0.8	27	0	2	5	16	4	
16. Instructor contrasted the implications of various theories.	2.8	0.9	27	1	10	11	4	1	
17. Students were encouraged to participate in class discussion.	4	0.8	27	0	1	6	13	7	
18. You have learned and understood the subject materials in this course.	3.7	0.6	27	0	0	9	16	2	
19. overall rating	4.6	0.6	27	0	0	1	9	17	
				poor	fair	average	good	excellent	

Figure 7

### Professor X Fall 1992 Departmental Placement

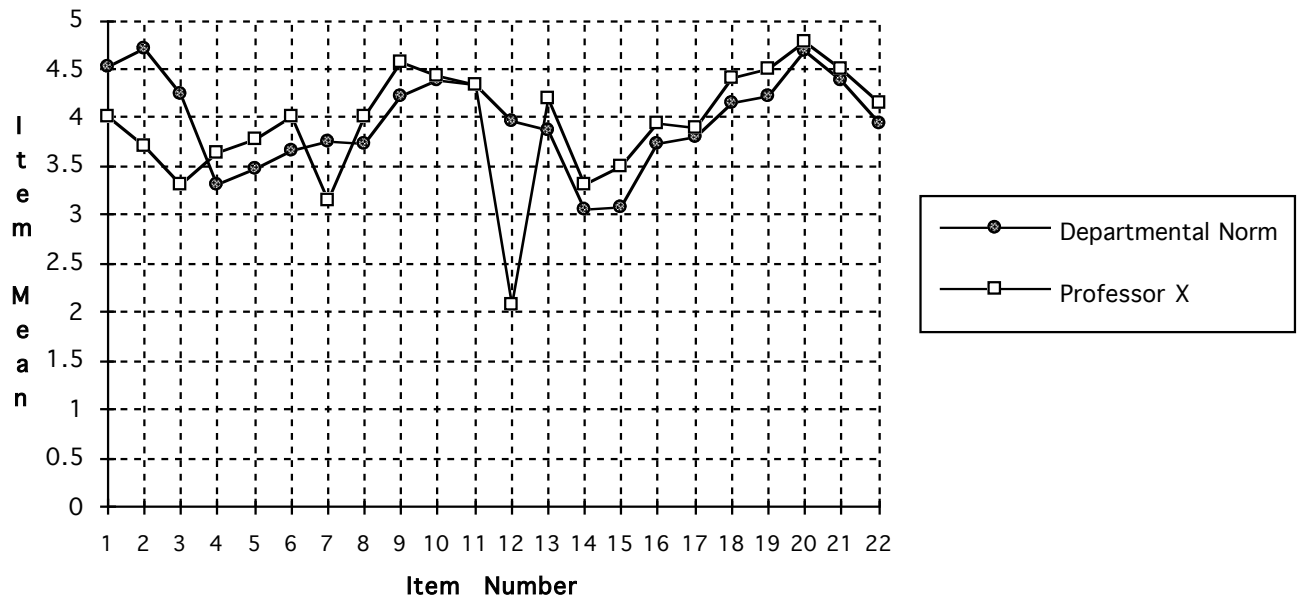


Figure 8

Qualitative data such as student written survey comments is harder to generalize and aggregate due to the fact that it is in a more open-ended form. Its potential bulkiness can be reduced through an interpretive narrative by the instructor that synthesizes the major themes and can help in the interpretation of the data (Franklin, 2001). The work of synthesizing can be spread out over time, on a year-by-year basis, as part of an annual review process.

During the preparation of the report *Evaluation and Recognition of Teaching* (Lynn, 1992), the deans of the colleges were interviewed. One dean raised the issue of anonymity of student evaluation data. Quantitative questionnaire scores aggregated by class section preserve anonymity. However, letters in their original form do not preserve the anonymity of the student. While students, either undergraduate or graduate, are still working with the candidate they are in what one dean called the candidate's "power web." This may prevent students from being as candid in their written remarks if they know the candidate may identify them at some point during the tenure decision process. Ways of preserving student anonymity include if letters by students are returned to someone other than the candidate—the department head or *Ad Hoc* chair, for example—and then input on a computer or supplied through an on-line response system that keeps the respondent anonymous. If students are informed that these precautions are being taken when they are asked to write their letters or respond to open-ended questions, the validity of their responses will be enhanced.

An example of the department chair synthesizing the relevant comments from undergraduate student reviewers who were asked to write letters of recommendation is included below.

...undergraduates uniformly describe him as an unusually effective, conscientious, enthusiastic teacher who enables students to do their best work, master difficult subject-matter, and gain confidence in their own intellectual abilities.

This [student quote from a review letter] clear and convincing testimony describes the experience of all the students who wrote to us from the courses he taught in spring 1988 and in fall 1989. Since the most disturbing aspect of some of the student responses two years ago was the suggestion that he could be authoritarian and coercive in his teaching, we are reassured by all these letters which suggest precisely the opposite.

It seems clear that like many young assistant professors [candidate] was too demanding in his first dealings with graduate students, imposing admirable but often excessive standards of professionalism both in the classroom and as a special committee member, and expecting his students to share his commitment to his own projects. As the letter from [student]

suggests, however, he has since become more realistic and flexible. And all the letters attest that he is always extremely conscientious and helpful.

One should conclude, I think that [candidate] is an intellectually stimulating and enabling graduate teacher, with an expertise and commitment that many of our students find particularly valuable, one who has had trouble finding the appropriate mode in which to exercise authority, but who has now learned to do so.<sup>12</sup>

The *usefulness* and *reliability* of student letters of evaluation, whether undergraduate or graduate, can be improved if specific criteria are communicated when letters are solicited to help focus the students. If the students are all requested to respond to the same questions, reliability will be enhanced and it will be easier to summarize all the letters. The following is an example of the kinds of questions about teaching that can be used to aid students in writing evaluation letters and thereby provide consistency in their responses:

1. Factual Knowledge: how well did the candidate help you acquire and integrate new terms, information and methods? Please give explicit examples where possible.
2. Concepts and Principles: how well did the candidate organize the material covered into a comprehensive whole? Were important concepts and principles from theory interrelated? Please give explicit examples where possible.
3. Application: Do you feel that the candidate's teaching and course structure enabled you to apply what you learned in the course to concrete problems? Were you able to generalize beyond the text? Please give explicit examples where possible.
4. Motivation: Did you feel the candidate was sufficiently motivated about the subject matter to excite your own interest in it? Describe how the candidate communicated a sense of enthusiasm about teaching.
5. Self Understanding: To what degree did the candidate help you become more aware of yourself as a learner? Describe specific experiences where the candidate contributed to your feeling empowered in your ability to learn.
6. Improvement of Instruction: Did the candidate seek out information from you and experiment with ways of improving his or her teaching? To what degree was the candidate open to feedback on improving the

---

<sup>12</sup>Cornell University Select Committee. (1992). *Evaluation and Recognition of Teaching Appendices*, Ithaca, N.Y.: Cornell University, 21.



course? How confident are you in the candidate's ability to continually develop as a teacher? Please be as specific as possible.

To avoid biasing faculty opinion of a candidate's teaching effectiveness, student letters, in any form, summarized or not, should not be available to the voting faculty until all file data on both teaching and research has been assembled into the tenure file. This is true for all data: everyone voting on the candidate should have the same base of data to make an informed and unbiased decision.

## **Peer Data**

Evaluation of the candidate's teaching by peers is a practice that has become more prevalent in tenure and promotion decisions over the last 30 years (Berk, 2005) and has taken on an increasingly significant role in the tenure decision process. Effective peer review depends to a large degree on the explicitness of the criteria by which candidates are to be judged.

Colleagues and peers are necessary contributors to evaluating a tenure candidate's teaching. They are best qualified to evaluate the candidate's breadth and depth of subject matter knowledge, course design skills and assessment strategies for determining students' learning the material. The information necessary for colleagues and peers to evaluate these kinds of skills must be thorough without being redundant. The candidate can help in peer evaluation by supplying the kind of information described in Chapter 2. However, colleagues from within the institution, both within the candidate's own department and outside it, and peers outside the candidate's institution, representing the discipline, will be required to provide their own data.

## **Classroom Observations**

Research has shown that peer review of classroom teaching can be important for maintaining the quality of teaching and learning in a department (Arreola, 2007; Kaplan, Meizlish, 2008; Seldin, 2006). It has the following potential benefits:

- It draws upon the disciplinary expertise of colleagues
- It contributes to a collegial academic culture
- It can benefit both parties involved
- Faculty participation is an expression of commitment to the quality of teaching and student learning

Experience with practice in this area suggests the following conditions for effective peer review:

- It must be carried out within an atmosphere of trust and respect
- The support, guidance and training of participants is necessary
- Faculty participation in peer review of teaching needs to be recognized and acknowledged within the departmental culture

In general, peer review of classroom teaching can include the following steps:

- a pre-observation meeting
- a class observation where the observer makes notes on what was observed
- a post-observation debriefing
- a written report by the observer

#### A suggested model for classroom observations

##### Pre-Observation Meeting

Best practices (Robertson, 2006) suggest that the observer meet with the instructor before the class observation to discuss the instructor's class and teaching. It may also be helpful to review teaching materials that were developed for the class. Materials may include:

- The course syllabus
- Any teaching materials the colleague has prepared for that class that might be relevant such as handouts, pre-class quizzes, homework assignments due that day, teaching notes, PowerPoint slides, or an overall teaching plan

During the pre-observation meeting the process can benefit from discussing the following:

- Which learning outcomes does this class address?
- How will class time be used? What activities can the observer expect to see?
- How does this class fit in with the overall course?
- What preparations for this class have the students been required to perform?
- Are there specific points on which the instructor would like to receive feedback?

##### Class Observation

Best practices (Robertson, 2006) in peer review propose that a core set of criteria be used in the observation process. Research in this area (Arreola, 2007) suggests that departments discuss and establish criteria that work for their field. These criteria may vary among fields. In the pre-observation meeting the instructor and observer might discuss two to three criteria on which the observer will focus during the class. The following criteria have been identified as factors that contribute to better student learning and can be a useful starting point when discussing criteria to evaluate teaching.

#### Sample Criteria:

- *Clarification of class purpose:* Does the instructor convey the focus for the class?
- *Organization of class structure:* Are the class materials and activities well organized?
- *Reinforcement of major concepts:* Does the instructor emphasize the major concepts being covered? Do the activities and materials utilized in class reinforce the major concepts?
- *Pacing and scope:* Is the material presented at a suitable rate? Is the amount of material covered reasonable?
- *Classroom atmosphere:* Has the instructor established a safe and respectful classroom atmosphere conducive to student learning? Has the instructor created an inclusive class environment?
- *Consideration of diversity:* Does the instructor acknowledge or interact with a broad range of students? Is the instructor respectful of diverse opinions and perspectives? Does the instructor employ a diverse set of activities or methods to accommodate a range of student learning styles?
- *Class management:* Does the instructor effectively manage the class?
- *Balance between abstract and concrete:* If applicable, is there an appropriate balance between abstract and concrete concepts?
- *Classroom assessment:* If applicable, in what ways does the instructor check for comprehension and solicit feedback?

#### Post Observation Meeting and Reporting

It is recommended that departments discuss and develop a post-observation process that reflects departmental teaching expectations and priorities for peer review.

A written report to document the teaching evaluation and the observations that were made can include a summary or copies of the following:

- peer observation criteria
- pre-observation-meeting notes and plan
- class observation notes that emphasize selected criteria
- post-observation meeting notes including identified goals for improvement by the one observed

#### **Template Letter for Peer Review of Teaching Evaluation**

*The template below is designed to help guide in constructing a letter for the peer review evaluation of teaching. The template is intended to be used in concert with the Peer Review of Teaching Forms available on the CTE web site:*

*<http://www.cte.cornell.edu/faculty/peerreview.html>*

#### **Example of Introductory Paragraph**

“This letter will summarize our recent discussions and my observation of your teaching for COURSE NAME AND NUMBER. Your SUBJECT OF CLASS was attended by NUMBER OF STUDENTS on DATE from TIME in LOCATION. I was requested by YOU or YOUR SUPERVISOR to evaluate your teaching skills FOR PROMOTION/REAPPOINTMENT, FOR SELF-DEVELOPMENT, or IN RESPONSE TO STUDENT EVALUATIONS. This

letter describes the three phases of the peer review of teaching process: pre-observation meeting, classroom observation, and post-observation feedback session.”

### **Body of Letter**

#### **Paragraph 2: Pre-Observation Meeting Summary**

Refer to your notes from the pre-observation meeting on the Peer Review Observer Form [link]; identify aspects of the class or teaching methods to be specifically observed, and indicate any relevant prior contextual issues discussed during the meeting such as previous student evaluation data.

#### **Paragraph 3: Classroom Observation Summary**

Based upon observation notes generated using the Guidelines for Peer Review of Teaching [link] and the focal points developed during the pre-observation meeting, generate a descriptive analysis of the teaching methods observed.

#### **Paragraph 4: Post-Observation Feedback Summary**

Using the notes generated from the pre-observation meeting and the class observation, summarize with the instructor positive traits and opportunities for improvement. This is where goals for instructional development can be collaboratively generated, if necessary.

### **Final Conclusion Paragraph**

Generate an overall impression of teaching skills with references to the core criteria outlined in the Guidelines for Peer Review of Teaching [link]. Begin with an inventory of effective teaching methods recognized during the peer review process. Offer constructive suggestions for improvement, if needed; refer to campus-wide resources (e.g., teaching mentorship within the department, College of Agriculture and Life Sciences Teaching Experience, Cornell’s Center for Teaching Excellence programs) where necessary. If recommendations for improvement of teaching are made, follow-up re-evaluation with a time line should be suggested.

*This template was modified from one developed in collaboration with faculty in the Department of Clinical Sciences at the Cornell College of Veterinary Medicine.*

### **Insuring Fairness in the Overall Process**

Several factors are critical in insuring a valid and fair overall peer review process. What questions are asked and answered by the reviewers is central. Some sort of replicable protocol is necessary to insure fairness and accountability for the process. This is true for whatever data is being reviewed, whether course materials, classroom observations or student evaluations. Developing a set of questions to focus the reviewer can make the task less arbitrary and subjective. The entire review process by peers should be governed by a set of procedures established within the department. Examples of such procedures include:

...peer ratings should be used in conjunction with student ratings. . . .  
dimensions [of teaching] should be decided upon in advance. . . [the]  
procedure should guarantee the anonymity and independence of the rater.

. . . at least three colleagues be chosen to rate an instructor's teaching... these raters...may come from...an elected committee of the college faculty whose function is to evaluate teaching....raters do not meet together and preferably do not know who else is involved in the evaluation process. Rather, each judge independently rates the instructor on the preselected dimensions and submits the ratings to the dean [or department head], who then compiles a pooled rating for each dimension.<sup>13</sup>

Developing these procedures and the questions used to review the candidate can be a useful accomplishment of a departmental standing committee on teaching.

### Qualification of Peer Reviewers

How peer reviewers are selected is another critical factor in establishing validity in peer review. No one should be placed in a position to review or observe a colleague for tenure or promotion decisions who does not feel qualified to carry out that task. A very consistent finding of peer observation studies is that observers should have some kind of training that prepares them for that responsibility (Bell, 2001; Chism, 2007). Peers are typically capable of evaluating subject matter knowledge, what must be taught by the candidate, whether the appropriate methodology is being employed for teaching specific content areas, the degree to which the candidate has applied adequate and appropriate evaluation techniques for course objectives, and the degree to which professional behavior has been exhibited according to current ethical standards.

---

<sup>13</sup>Cohen, P. & McKeachie, W. (1981).The role of colleagues in the evaluation of college teaching. *Improving College and University Teaching*(28), 4, 150.

## Chapter IV—Criteria for Evaluating Data on Teaching

This discussion of evaluation criteria is meant to assist the department or college in establishing its own system for the evaluation of teaching. Tenure decisions require a general rule for defining excellence that can accommodate the variety of disciplinary-based instructional traditions, while honoring the individual's freedom to express and develop personal style in teaching. An example of such a general rule might be, "To what degree does the data supplied support the reasoned opinion of those most competent to judge that the candidate has—and will continue to—demonstrate the capacity to maintain agreed upon instructional standards?"

A distinction has been made between promotion and tenure criteria: *promotion* criteria focus on merit of the candidate's professional and scholastic contributions and promise, whereas *tenure* criteria focus on the long-time worth of the candidate's professional and scholastic contributions and promise. "Worth requires merit, but merit is not a sufficient condition for worth."<sup>14</sup>

"Merit is free of the specifics of the pool; it is criterion referenced and deals with the candidate's ranking on those criteria. Worth is utility to the hiring party."<sup>15</sup> Based on these definitions, if tenure candidates sufficiently prove their capacity and commitment to continually improving teaching practice, their long-time worth to the unit and institution will be greater than if they have merely made a case that their performance has measured up to a universally established absolute standard. Normative standards are necessary for determining merit, but merit is a concept which may be more relevant at the point of hiring and when the candidate is being considered for promotion.

In general, criteria for evaluating teaching will be more useful in the tenure and promotion process if they:

- can discriminate between teachers in terms of specific competencies
- can reliably and consistently measure a specific competency both for the same individual over time and between individuals
- maintain a neutral orientation relative to individual style and viewpoint

---

<sup>14</sup> J. Aubrecht, Delaware State College: personal communication.

<sup>15</sup> Leinhardt, G. (1991). Evaluating the New Handbook of Teacher Evaluation. *Educational Researcher*(20), 6, 24.

- yield information about instructional situations where the teacher functions best<sup>16</sup>

This chapter provides suggested criteria with which to evaluate teaching, based on the data categories described in Chapter III. It begins with a general discussion of effective teaching and goes on to include criteria for use by students, criteria relevant to evaluating teaching materials, and criteria for use by peers, including classroom observations, all of which have been developed through controlled inquiry in work carried out within the last 35 years.

### **Effective Teachers—A Description**

During the past 50 years the debate over effective teaching has moved from a discussion of technical, classroom skills, or *process skills* as they have been called, to a focus on skills necessary to make the subject matter understandable to the student, what one author calls "the pedagogy of substance." Thinking dichotomously about teaching as either a technical process skill divorced from the subject matter, or solely a matter of translating abstract and technical information into understandable terms limits the conception of what teaching is. Looking at teaching as a scholarly activity that is connected to research suggests a dialogue between the tasks of understanding a body of knowledge and explaining it. Effective teaching must be concerned with both of these areas of expertise: I am no more effective if I have a body of knowledge to profess but am unable to communicate it than I am if I can hold students rapt in wonder but do not know what I am talking about (Naftulin, Ware & Donnelly, 1973).

This dichotomy can be avoided by a more integrative model of teaching: effective teachers are able to understand enough about their students' learning the content they teach that they can translate their own understanding of the subject matter into a form that is accessible to their students.

...one of the things we see when we look at teaching analytically is this combination of an emphasis on understanding the subject matter, understanding how it is represented in the heads of students and then being able to generate representations of your own as a teacher that will be a bridge between the subject matter and the students.<sup>17</sup>

---

<sup>16</sup>McNeil, J.D. and Popham, W.J. (1973). The assessment of teacher competence. In R.M.W. Travers (ed.). *Second handbook of research on teaching*. Skokie, IL.: Rand McNally, 218-244.

<sup>17</sup>Shulman, L. (1989). Toward a pedagogy of substance. *AAHE Bulletin*, June, 11.

Work on teacher effectiveness has yielded the following observations that support an integrative model which is both *process* and *content*-based:

...Teachers promote learning by communicating to their students what is expected and why.

Effective teachers not only know the subject matter they intend their students to learn but also know the misconceptions their students bring to the classroom that will interfere with their learning of that subject matter

Effective teachers are clear about what they intend to accomplish through their instruction, and they keep these goals in mind both in designing the instruction and in communicating its purposes to the students. They make certain that their students understand and are satisfied by the reasons given for why they should learn what they are asked to learn.

Effective instruction provides students with metacognitive strategies to use in regulating and enhancing their learning. It also provides them with structured opportunities to exercise and practice independent learning strategies.

Effective teachers create learning situations in which their students are expected not just to learn facts and solve given problems but to organize information in new ways and formulate problems for themselves. Such learning situations include creative writing opportunities in language arts, problem-formulation activities in mathematics, and independent projects in science, social studies and literature.

...Effective teachers continuously monitor their students' understanding of presentations and responses to assignments. They routinely provide timely and detailed feedback, but not necessarily in the same ways for all students.

...Effective teachers realize that what is learned is more likely to be remembered and used in the future if it serves students' purposes beyond meeting school requirements.

...effective teachers...take time for reflection and self-evaluation, monitor their instruction to make sure that worthwhile content is being taught to all students, and accept responsibility for guiding student learning and behavior....the same research...has made it clear that few teachers follow all of these practices all of the time.



...teachers must cope with a full agenda that typically precludes time for serious reflection...<sup>18</sup>

The last two points raised in the quotes above deserve emphasizing. First, teachers are human and not machines. Strict adherence to a set of principles does not in itself establish effectiveness. I may, for any number of acceptable reasons, occasionally exhibit inconsistency in teaching practice. The more important issues are: to what *degree* is my practice governed by some explicit pedagogical framework, and how *frequently* am I unable to follow my own guiding principles of teaching which my experience has shown to produce desirable results. More is said about guiding principles and teachers' pedagogical frameworks in Chapter V.

Second, the extent to which I can be effective will be governed, to a certain degree, by the environment and conditions under which I must work. I only have so much time and energy, and I have a life beyond my work, which has its own demands. These are facts we take for granted, but because we take them for granted, we may be in danger of forgetting them during the rigor of a tenure decision. A case where a newly hired faculty member is assigned to teach five courses represents a much more stressful situation than a case with a lighter teaching load. Workload is an important factor to be considered when evaluating a candidate on the following departmentally-based criteria:

- Has the candidate assumed the responsibilities related to the department's or university's teaching mission?
- Does the candidate recognize the problems that hinder good teaching in his or her institution and does he/she take a responsible part in trying to solve them?
- If all members of the faculty were like this individual, what would the college be like?
- To what extent is the candidate striving for excellence in teaching?<sup>19</sup>

If teaching is to be adequately rewarded as a valued activity and contribution to the department or unit, the degree to which a candidate has accomplished the following should be recognized.

whether there is sufficient data on teaching quality  
whether alternative teaching methods have been explored

---

<sup>18</sup>Porter, A. & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership*, 78-83.

<sup>19</sup>French-Lazovik, G. (1981). Peer review: Documentary evidence in the evaluation of teaching. In J. Millman, (ed.). *Handbook of Teacher Evaluation*. Beverly Hills: Sage Publications, 77-78.

whether changes have been made in the candidate's courses over time  
whether the candidate sought aid in effectively incorporating new teaching methods  
whether the candidate developed special teaching materials  
whether the candidate participated in teaching improvement opportunities

## **Evaluating Course and Teaching Materials**

Teaching and course materials are evidence that a department can use to evaluate a tenure candidate's course design skills as well as those skills necessary to effectively evaluate student learning.

*Suggested criteria for evaluating teaching and course materials*<sup>20</sup>

### Course organization

1. The course objectives are congruent with the department curricula.
2. The course objectives are clearly stated.
3. The syllabus adequately outlines the sequence of topics to be covered.
4. Is the syllabus current and relevant to the course outline?
5. Are the outline and topic sequence logical?
6. The intellectual level of the course is appropriate for the enrolled students.
7. Time given to the various major course topics is appropriate.
8. The course is an adequate prerequisite for other courses.
9. Written course requirements, including attendance policies, are included in the course syllabus.

### Course content

1. The required or recommended reading list is up to date and includes works of recognized authorities.
2. A variety of assignments is available to meet individual needs.
3. Laboratory work, if a part of the course, is integrated into the course.
4. The assignments are intellectually challenging to the students.
5. Is it up to date?
6. Is the instructor's treatment fair and engaging?
7. Are conflicting views presented?
8. Are the breadth and depth of coverage appropriate for the course?
9. Has the instructor mastered the subject matter?

### Evaluating Student Learning

1. The standards used for grading are communicated to the students in the course syllabus.
2. The written assignments and projects are chosen to reflect course goals.
3. The examination content is representative of the course content and objectives.
4. The tests used in the course have been well designed and selected.
5. The examination questions are clearly written.
6. The examinations and assignments are graded fairly.

---

<sup>20</sup> Miller, R. (1987). *Evaluating faculty for promotion and tenure*. Jossey-Bass: San Francisco.

7. The grade distribution is appropriate to the level of the course and the type of student enrolled.
8. The examinations and papers are returned to the students in a timely fashion.
9. Students are given ample time to complete the assignments and take-home examinations.
10. The amount of homework and assignments is appropriate to the course level and to the number of credit hours for the course.
11. Is the examination suitable to content and course objectives?
12. Are tests graded and returned promptly?
13. Are the grading standards understood by students?
14. Is the grade distribution pattern appropriate for the course level?
15. How do students perform in more advanced courses?
16. Do students apply the principles learned in the course in their papers and projects?
17. What is the general quality of major homework assignments?

#### Course Objectives

1. Have the objectives been clearly communicated to the students?
2. Are they consistent with the department's relevant program objectives?
3. If the course is a building block for a more advanced course, are the students being properly prepared?

#### Instructional Methodology

1. Are the instructor's teaching approaches (lectures, discussion, films, fieldwork, outside speakers) suitable to the course objectives?
2. Is the pacing appropriate?
3. Are students learning to use the library resources and necessary research skills for the course?
4. Would other technology strengthen the course?

#### Homework Assignments

1. Do homework assignments supplement lectures and class discussions?
2. Do assignments reflect appropriate course goals?
3. Is the reading list relevant to course and program goals?
4. Is it appropriate to the course level?

Once a set of evaluation criteria have been established within a department, thought must be given to weighting the various evaluation sources. A departmental standing committee on teaching can be responsible for determining the relative weight attributable to each data source in order to insure consistency between tenure cases, and to explicitly communicate the department's expectations regarding teaching. An example of how this might be done is included in Table 3.

Table 3 - Weighting of Teaching Evaluation Data<sup>21</sup>

---

<sup>21</sup> adapted from Cohen, P. & McKeachie, W. (1981) The role of colleagues in the evaluation of college teaching. *Improving College and University Teaching*(28), 4, 152.

<u>Sources of information</u>	<u>Percentage of Total Evaluation</u>
Student ratings of in-class activities	30
Peer Rating of Course design features	
organization	5
goals	5
instructional materials	5
evaluation devices	5
Peer rating of teaching qualities	
intellectual breadth	5
commitment to teaching	5
improvement of teaching practice	20
Peer rating of student achievement	10
Self-rating of overall teaching effectiveness & improvement	10

A major premise of this handbook has been that the demonstrated development of teaching practice should be a major criterion by which a tenure candidate is evaluated. Just what that means, how it can be accomplished, documented and evaluated is the subject of the last chapter.

## Chapter V—Improving Instructional Practice

Improvement of teaching is the result of an integration of refinements in how teachers think about teaching, what their beliefs are in terms how they can influence student learning, what they have come to know and understand, based on experience and experimentation in action, and the degree to which they see teaching as an important part of their scholarly work which is connected to research. As such, the development of teaching practice will be manifested on several levels and the criteria used to measure that development must encompass cognitive, attitudinal, value and behavioral indices. Relying on a single indicator or data source such as student evaluations is not sufficient. Higher student evaluation scores may not always indicate improvement. Improvement grows out of reflection on feedback about one's teaching (Brookfield, 1995; McKeachie & Svinicki, 2011). This reflection helps teachers develop their professional knowledge, which will in turn affect their future instructional planning. The key concepts here are *feedback* and *reflection*. It is not sufficient to merely solicit feedback on instructional quality from students and peers and put it in a tenure file, reflection on the meaning of that feedback and responding to it is necessary if development in teaching is to take place.

In most cases the degree to which the individual can adequately reflect on his or her teaching will be determined by the working environment created within his or her department. Overburdening new faculty with too many courses, or courses that are very labor intensive with insufficient instructional resources in the form of teaching assistants or money for materials to teach, or focusing too heavily on the establishment of a research agenda will create a working environment where reflection on teaching may seem a luxury too costly to indulge in for the newly hired faculty member. There are cases where the individual can adequately reflect on his or her teaching practice in spite of these added burdens, but many promising teachers may, without being that aware of it, assimilate an adaptive attitude of low priority toward teaching in order to survive. When such a thing happens, everybody loses: the candidate, their students and the department that has made the investment in hiring the faculty member to start with.

Reflection on practice can take place in the planning stage of teaching as well as after a teaching episode—when the teacher may be identifying course objectives, developing assignments, choosing readings or revising a teaching plan. What may make a significant difference in the degree to which an individual improves over time is a different kind of reflection. In his book *The Reflective Practitioner*, Donald Schön uses the terms “reflection-on-action” and “reflection-in-action” as key to the practice of a new “epistemology of

practice.”<sup>22</sup> His basic premise is that in order to make significant and effective changes in our professional practice we need to honor a different way of producing knowledge about what we do. This involves taking the time necessary to reflect on our practice so we become more aware of why we do things the way we do, what doesn’t work and why and what may be more productive approaches. An example of this relevant to teaching is thinking metacognitively, which means identifying and examining our assumptions and using that process to guide the development of practice. Paying more attention to student comprehension, reflecting on ways to better determine student comprehension of course material is critical in adjusting teaching practice. Another reflective approach to developing one’s teaching involves using misconceptions to guide instruction. If I have spent enough time observing and reflecting on how my students learn or don’t learn what I teach, I may gain some insight into how to be more effective as a teacher by explaining things in more effective ways.

Some people may do this as a matter of course, but given the very heavy content orientation that graduate education tends to reinforce in preparing future faculty, the chances are that most reflection on action is limited to trial and error and focuses on course content. Reflection on the *process* of learning the subject matter and what is psychologically necessary to learn it adequately is very important in improving one’s teaching. An example of this is in the area of setting objectives. Many novice teachers begin by planning their classes in terms of how to fill the time allotted. They begin by thinking egocentrically: how will *I* plan to use this class time to cover what *I* think the students should know. In contrast, a more learning-centered teacher may make adjustments to practice based upon reflecting on students’ performance and response. A more fruitful way to plan a class might be to start with the students’ frame of reference and try to build a conceptual bridge to my own frame of reference as the instructor. From a planning point of view this would involve a very different set of questions for setting class objectives, focused around: what do I want my students to be able to say, do, think, or feel when they leave at the end of class that they couldn’t say, do, think, or feel when they walked in to it? This is the difference of developing teaching based on inputs (content) versus outcomes (students’ exiting knowledge, skill and attitudes.)

Regardless of their graduate training or experience, all teachers bring to their practice what might be called personal theories of teaching and learning (Hofer, 2001; Richards, 2002). Such theories may not be very conscious but the authors studying teacher thinking suggest

---

<sup>22</sup>Schon, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books 276-78.

that teachers' personal theories are what guide them in making planning, course design and classroom behavioral decisions. Personal theories of teaching and learning grow out of our experiences as students and teachers and begin developing while we are children. We develop predispositions toward certain learning "styles" as some authors describe it (Pashler, et al., 2009), just as teachers may gravitate around certain teaching styles. The degree to which we are conscious of our personal theories, how they guide us, how we refine them and how flexible we are from central stylistic tendencies can be thought of as indices of our instructional development. One can read books that can influence the way we think about teaching but we learn how to do it and improve through trial and error practice, or doing as one author describes it as "conducting on-the-spot empirical experiments." The problem from the point of view of efficiently developing our teaching is that, for the most part, we make no records of these experiments so our learning curve is dependent on our memory.

If I am to trace the improvement of my teaching practice, a useful focal point is an analysis of the degree to which my practice is consistent with my thinking about teaching (my personal theory.) Developing my personal theory is an epistemological issue as I discover and refine my knowledge about teaching and learning. The basis of this epistemology lies in the interrelationship between what I say I do when I describe and explain my personal theory of teaching and learning, on the one hand, and what, in fact, I do in practice, on the other. That part of my theory that is evident in what I *say* I do and which I have reflected on enough to be able to articulate it is my "espoused theory" (Argyris & Schön, 1974). In contrast to my espoused theory is my "theory in use" (Argyris & Schön, 1974) that is evident in my *behavior*. The developmental question that is useful for me to continually reflect upon is "To what degree is my espoused theory consistent with my theory in use?" Or, to put it more simply, "To what degree am I able to do in practice, what I think and believe I should be doing?" Discrepancies between the two can lead me to further refine my thinking or my practice, as the case may be.

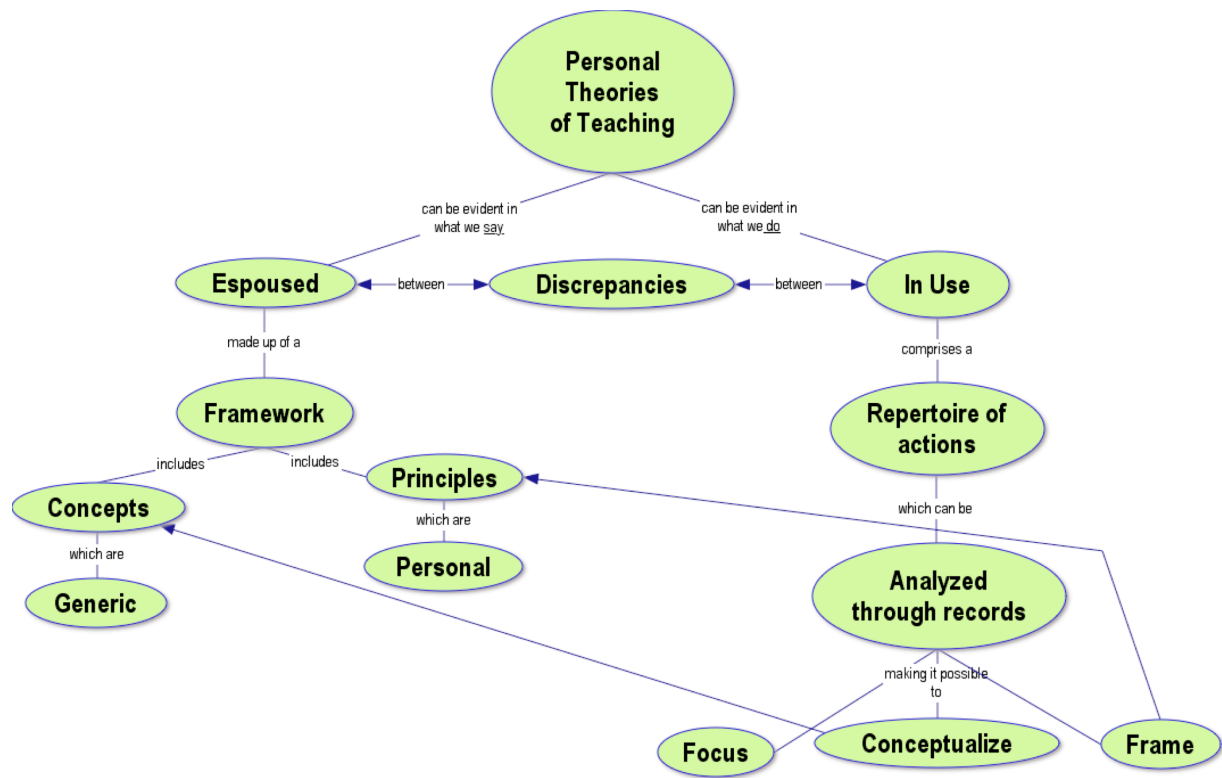
This can be illustrated with a very prevalent, but simple example. A frequently used teaching technique we all are familiar with is the use of the phrase, "Are there any questions?" This question is usually asked after we have just finished explaining some important or particularly difficult point. Another recurrent use of this question is at the end of a class or seminar to initiate discussion. In many cases, particularly with novice teachers, but even with veterans, particularly when teaching something for the first time, the use of this question proves ineffective. What may be forthcoming, rather than a flood of inquiry, are confused looks and vacant stares. I may have to experience this frustrating effect many times before I even begin

to acknowledge that something in my teaching is in need of change, much less pinpoint the specific problem. Acknowledgement may slowly dawn on me from simple repetitiveness, or someone—a student or colleague—may risk bringing it to my attention.

If a colleague were to point this problem out to me and ask me, "What are you expecting to accomplish when you ask 'Are there any questions?'" I would be articulating my espoused theory. I might say something like, "I want to give the students a chance to clear up any confusion they may have about what I have just explained." If my colleague has observed me enough to notice a consistent pattern in my use of this question she might reply, "If that's what you are trying to do, you're going about it the wrong way. You consistently look down at your watch immediately after you ask that question."

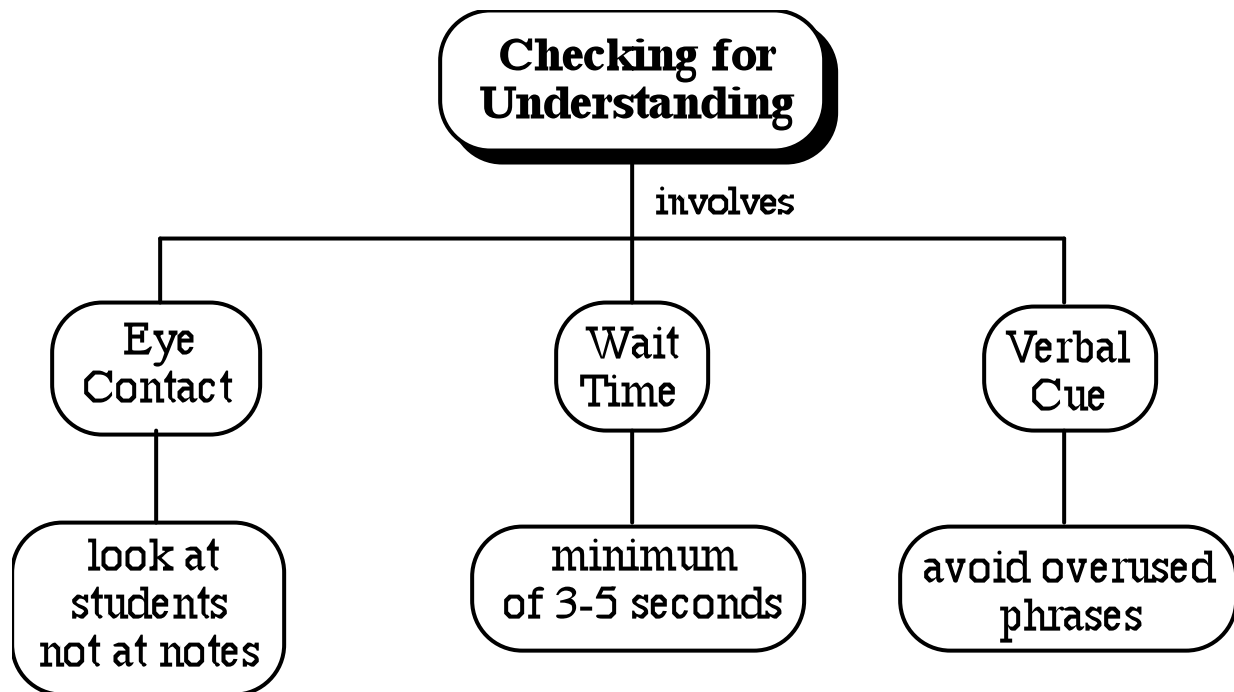
When this discrepancy is analyzed from a personal theory point of view to encourage the development of my teaching practice, I must begin to conceptualize this sequence of events so that I can begin to think strategically about it. In Schön's terms, I am "reflecting on action." I might begin to develop a framework of teaching concepts to help me clarify my thinking about the problem. The first step involves naming the situation, like "checking for understanding." *Checking for understanding* now has become a concept of teaching that I can begin to think consciously about. I can begin to use it in my planning for classes, choosing more strategically where it might be appropriate and necessary. My colleague's observation about my looking down at my watch might help me further conceptualize this situation. If I ask myself why am I looking at my watch at that point, I might begin to recognize that I am more anxious about the time than in waiting sufficiently for students to respond. Upon hearing this, my colleague might point out that from the students' perspective, I don't look like I want to be interrupted. This situation can be conceptualized as a "mixed cue" where I send out a *verbal* cue to my students which says, "Give me some feedback? What don't you understand?" while simultaneously sending out a *behavioral* cue which directly conflicts with my verbal cue. The net result is silence, filled with frustrated and confused looks from my students. This mixed cue example points out where my espoused theory (checking for understanding) is discrepant from my theory in use (time management) The diagram below helps put this in epistemological perspective.





This diagram illustrates how our espoused theory is made up of a framework of concepts and principles that we use to think about teaching and which guide, in some ways, our theory in use. Reflecting on our actions taken helps us to integrate the two, so that we are more consistent, successful in achieving our intentions, and more knowledgeable about teaching practice in the sense that we can explain it to others. If we can articulate the knowledge that is evident in our actions the implication is that our actions become more governed and justified by a coherent framework that is part of a personal theory.

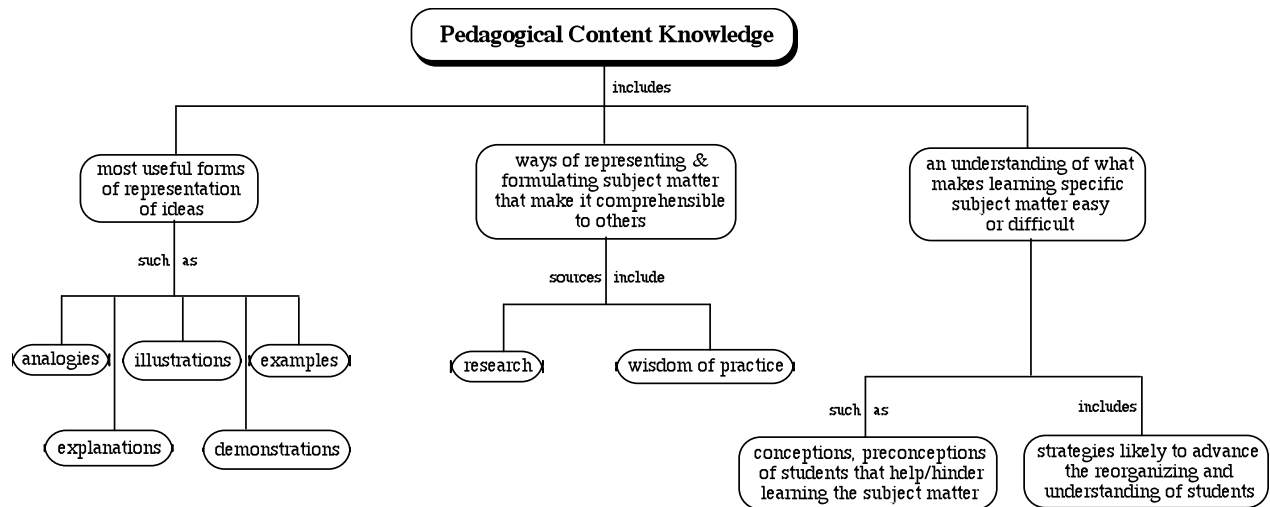
Research comparing novice and experienced teachers (Meskill, et al., 2002) has indicated that as teachers learn from experience they develop a repertoire of concepts, which guide their practice and help them deal with the variety of situations they come to face. These *concepts* of teaching are useful as they are developed into *principles* of teaching. A principle of teaching is a guide to action that includes at least one concept of teaching. Using the checking for understanding example, we can see how concepts of teaching are generic in the sense that they do not, in and of themselves, guide us. We all must deal with the issue of checking for understanding but we all do it in our own ways, based on our preferred teaching style and what our experience has taught us. A principle of teaching is our personal behavioral guide to how we check for understanding (the concept): I will pause after explaining an important point, make eye contact and ask, "Who would like me to go over that again?" leaving at least 3 seconds for students to react. This principle is now part of my espoused theory that guides my practice (theory in use) and can be explicitly represented as part of a framework:



Whereas a concept of teaching is more generic, such as "warm up," "wait time," or "summarizing" a principle of teaching is an individualized way of dealing with that generic regularity. An example of a *principle* for warming up in a foreign language class might be to engage students in casual conversation in the target language as they enter the classroom in order to get them thinking in that language by the time formal instruction begins. An example of a *principle* to ensure enough wait time after asking a question in order for students to respond effectively is to count silently to yourself from one to five, or seven. An example of a *principle* to deal with effective summarizing might be to budget time into your teaching plan at the end of the period and to review briefly the major concepts that have been covered, but in a fresh way. Of course there are many other principles to deal with each of these situations, which implies that effective teachers have extensive repertoires of principles so that if one principle proves ineffective in a situation, they have others available to choose from.

Colleagues can help each other improve practice in a way that respects individual differences and personal style in teaching when they observe each other and discuss their teaching. As they do so they are articulating their espoused theories and presenting an observational record of theories in use. If I am invited to reflect on my teaching actions by a colleague who has observed me and who asks me questions like, "What were you doing there? Was it effective? What alternative strategy might have been more productive?" I can begin to get used to reflecting on my teaching, which in turn may help me learn how to "reflect in action," as Schön (1983) called it. I am able to reflect in action once I have articulated my personal theory to such a degree that my thinking and doing are consistent with each other. When I function on this level I am very efficient. My attention is not cluttered by the necessity to think through every response and pay attention to every detail available to me. I can now use part of my attention to monitor my performance by asking myself questions like, "What am I doing now? Is it getting me anywhere? What other principle do I know, or can I invent, which may be more productive?" In this way, both collaboratively with my colleagues, and individually through reflection, I continually develop and refine my knowledge about teaching and learning which constitutes a personal theory I use to guide and improve my practice.

The example of checking for understanding is a process-related teaching concept. It has more to do with how I handle the process of explaining something. Teaching involves other knowledge relating to the content that is taught. In his work on the knowledge base of teaching, Lee Shulman (1986) used the term "pedagogical content knowledge" which can be illustrated as follows:



adapted from Shulman, L. "Those Who Understand: Knowledge Growth in Teaching,"  
*Educational Researcher*, Feb. 1986, pgs. 4-14

It is in the area of pedagogical content knowledge where colleagues can be extremely valuable in helping each other improve their teaching. There is a lot of creativity involved in devising analogies and choosing effective examples which prove to "achieve shared meaning" with students, as D.B. Gowin (1981)<sup>23</sup> has defined teaching. The relationships that develop between faculty members are central to the evaluation of teaching within the tenure process. It has been the premise of this handbook that the most efficient use of faculty time with regard to the evaluation of teaching is to integrate the summative and formative forms of evaluation. Documenting teaching for tenure files and the evaluation of the data gathered will be much more efficient if the necessary work has been engaged in collaboratively between faculty colleagues during the six years leading up to the tenure decision process. Direct observation of teaching performance is the final outcome of a lot of prior thinking and work that colleagues can collaborate on. The Teaching Portfolio described in Chapter II is designed to foster both thoughtful discourse about teaching and professional inquiry into how to contribute more effectively to student learning.

### **Using the Teaching Portfolio to Improve the Quality of Discourse on Teaching**

The analysis of tenure files conducted in preparation for the report *Evaluation and Recognition of Teaching* (Lynn, 1992) revealed a significant imbalance between the level of documentation regarding teaching compared to that dealing with research. This was evident in the following quote from a department head to the tenure candidate:

We discussed your teaching activities and my indications are that your teaching...is going well and that it is well received by your students. We hope that the renovations of the laboratories...will provide opportunities for further improvements in our...teaching program.<sup>24</sup>

In spite of the fact that this letter is out of a larger context in which the tenure case was judged, it fails to document for anyone outside the department (the dean or provost) just what "indications" were used as evidence in making the judgment that the candidate's teaching was "going well" or in what way was it "well received by [the candidate's] students." Supervision by senior colleagues and department heads is critical in establishing an atmosphere where teaching is valued and constructively evaluated. The documentation of a candidate's teaching

---

<sup>23</sup>Gowin, D.B. (1981). *Educating*. Ithaca, N.Y.: Cornell University Press, 62, 2.

<sup>24</sup>Cornell University Select Committee. (1992). *Evaluation and Recognition of Teaching Appendices*, Ithaca, N.Y.: Cornell University, 20.

practice and level of thinking about teaching can be reviewed through the teaching portfolio. The level of discourse about teaching that is exchanged between the candidate and supervisor establishes a standard by which teaching is evaluated.

### **Cornell Resources to Assist in the Development and Evaluation of Teaching**

Cornell's Center for Teaching Excellence has developed resources to assist faculty members develop and document their teaching effectiveness as well as conducting the peer review of teaching. These resources can be accessed on our web site here:

<http://www.cte.cornell.edu/faculty/peerreview.html>

In addition, faculty working on developing course design can benefit from participating in the CTE Course Design Institute. Registration information, its schedule, and agenda are posted on the CTE web site.

CTE staff members are also available to discuss the evaluation and documentation of teaching at department faculty meetings by scheduling meetings through e-mail at

[cornellcte@cornell.edu](mailto:cornellcte@cornell.edu). Some departments have found it helpful to provide training in conducting peer review of teaching to their faculty members by engaging CTE staff to facilitate that training. Newly hired faculty members can document their efforts at developing their teaching and courses by participating in the several certificate programs available.

Information about these opportunities is available on the CTE web site here:

<http://www.cte.cornell.edu/faculty/teachcert.html>

<http://www.cte.cornell.edu/faculty/jrfaculty.html>

It has been the goal of this handbook to provide some useful ideas and information to assist colleges, departments and faculty members in the complex and critical task of evaluating teaching within the tenure and promotion process. Further assistance and resources are available from the Center for Teaching Excellence:

420 Computing and Communications Center  
Garden Avenue Extension  
Cornell University  
Ithaca, NY 14853-6601  
607-255-3990  
[cornellcte@cornell.edu](mailto:cornellcte@cornell.edu)

## References

- Argyris, C. & Schon, D. (1974) *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.
- Arum, R. (2011). *Academically adrift: Limited learning on college campuses*. University of Chicago Press.
- Becker, W. E. & Andrews, M. L. (2004). *The scholarship of teaching and learning in higher education: Contributions of research universities*. Indiana University Press.
- Boyer, E. (1990). *Scholarship reconsidered*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching, 23-24.
- Brookfield, S. (1995). *Becoming a critically reflective teacher*. San Francisco, CA: Jossey-Bass.
- Brookfield, S & Preskill, S. (2008). *Learning as a way of leading: Lessons from the struggle for social justice*. Jossey-Bass.
- Centra, J.A. (1979). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Centra, J.A. (1975). Colleagues as raters of classroom instruction. *Journal of Higher Education*. 46, 327-337.
- Chism, N.V. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Bolton MA: Anker Publications.
- Cohen, P. (1981) "Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies." *Review of Educational Research*, 51, 3, 281-309.
- Cohen, P. & McKeachie, W. (1981). The role of colleagues in the evaluation of college teaching. *Improving College and University Teaching* (28), 4, 147-154.
- Cornell University Select Committee. (1992). *Evaluation and Recognition of Teaching Appendices*, Ithaca, N.Y.: Cornell University.
- Edgerton, R. (1991). The teaching portfolio as a display of best work. Paper presented at the National Conference of the American Association for Higher Education, Washington, D.C.
- Edgerton, R., Hutchings, P., & Quinlan, K. (1991). *The teaching portfolio: Capturing the scholarship in teaching*. Washington, D.C.: American Association for Higher Education.
- French-Lazovik, G. (1981). Peer review: Documentary evidence in the evaluation of teaching. In J. Millman, (ed.). *Handbook of Teacher Evaluation*. Beverly Hills: Sage Publications, 77-78.
- Gowin, D.B. (1981). *Educating*. Ithaca, N.Y.: Cornell University Press.



- Hoffman, R. (1989, Summer). Research/teaching—A complex dance. In *Cornell Undergraduate Education*, ( 2), 2, Office of Instructional Support, Cornell University.
- King, B. (1990). *Linking portfolios with standardized exercises: One example from the teacher assessment project*. (Technical Report). Stanford, Calif.: Stanford University, School of Education, Teacher Assessment Project.
- Leinhardt, G. (1991). Evaluating the New Handbook of Teacher Evaluation. *Educational Researcher*(20), 6.
- McNeil, J.D. and Popham, W.J. (1973). The assessment of teacher competence. In R.M.W. Travers (ed.). *Second handbook of research on teaching*. Skokie, IL.: Rand McNally, 218-244.
- Miller, R. (1987). *Evaluating faculty for promotion and tenure*. Jossey-Bass: San Francisco.
- Millis, B.J. (1991). Putting the teaching portfolio in context. *To Improve the Academy*, 10, 215-232.
- Naftulin, D.H., Ware, Jr., J.E, &Donnelly, F.A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48(July), 630-635.
- O'Neil, C., & Wright, A. (1992). *Redording teaching accomplishment: A Dalhousie guide to the teaching dossier*. ( 2nd ed.). Dalhousie University, Halifax, Nova Scotia: Office of Instructional Development and Technology, 66-67.
- Porter, A. & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership*, 74-85.
- Seldin, P. (1984). *Changing practices in faculty evaluation*. Jossey-Bass.
- Seldin, P. (1989). The teaching portfolio. Paper presented at the annual meeting of the Professional and Organizational Development Network in Higher Education, Jekyll Island, GA.
- Seldin, P. (1991). *The teaching portfolio—A practical guide to improved performance and promotion/tenure decisions*. Anker Publishing Co., Bolton, MA.
- Sell, G. R. & Chism, N. (1988). *Assessing teaching effectiveness for promotion and tenure: A compendium of reference materials*. Center for Teaching Excellence, Ohio State University.
- Schon, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books, 276-78.
- Scriven, M. (1981). Summative teacher evaluation. In J. Millman (ed.). *Handbook of Teacher Evaluation*. Beverly Hills: Sage Publications.

Shulman, L. (1989). Toward a pedagogy of substance. *AAHE Bulletin*, June, 11.

Sullivan, A. and Donald, J. (Eds.). (1985). The role of two types of research on the evaluation and improvement of university teaching. *Using Research to Improve Teaching: New Directions for Teaching and Learning*, 23, Jossey-Bass, San Francisco.

Whitman, N. & Weiss, E. (1982). Faculty evaluation: The use of explicit criteria for promotion, retention and tenure. *AAHE-ERIC/Higher Education Research Report*, No. 2.

## **Bibliography**

### Peer Review of Teaching

Atwood, C. H., Taylor, J. W., & Hutchings, P. A. (2000). Why are chemists and other scientists afraid of the peer review of teaching? *Journal of Chemical Education*, 77(2), 239-243.

Bell, M. (2001). Supported reflective practice: A program of peer observation and feedback for academic teaching development. *International Journal for Academic Development*, 6(1), 29-39.

Bernstein, D. J. (2008). Peer review and evaluation of the intellectual work on teaching. *Change*, 40(2), 48-51.

Bernstein, D. J., Burnett, A. N., Goodburn, A., & Savory, P. (2006). *Making teaching and learning visible: course portfolios and the peer review of teaching*. Massachusetts: Anker.

Blackmore, J.A. (2005). A critical evaluation of peer review via teaching observation within higher education. *International Journal of Educational Management*, 19(3), 218-232.

Cestone, C.M., Levine, R.E., & Lane, D.R. (2008). Peer assessment and evaluation in team-based learning. *New Directions for Teaching and Learning*, 116 (Winter), 69-78.

Chism, N.V. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Bolton MA: Anker Publications.

Courneya, C., Pratt, D., & Collins, J. (2008). Through what perspective do we judge the teaching of peers? *Teaching and Teacher Education*, 24, 69-79.

Gosling, D. (2002). Models of peer observation of teaching.  
[http://www.heacademy.ac.uk/resources/detail/id200\\_Models\\_of\\_Peer\\_Observation\\_of\\_Teaching](http://www.heacademy.ac.uk/resources/detail/id200_Models_of_Peer_Observation_of_Teaching) (

Keig, L. (2000). Formative peer review of teaching: Attitudes of faculty at liberal arts colleges toward colleague assessment. *Journal of Personnel Evaluation in Education*, 14(1), 67-87.

Kell, C., & Annetts, S. (2009). Peer review of teaching: Embedded practice or policy-holding complacency? *Innovations in Education and Teaching International*, 46(1), 61-70.

Lomas, L. & Kinchin, I. (2006). Developing a peer observation program with university teachers. *International Journal of Teaching & Learning in Higher Education*, 18(3), 204-214.

McMahon, T., Barrett, T., & O'Neill, G. (2007). Using observation of teaching to improve quality: Finding your way through the muddle of competing conceptions, confusion of practice, and mutually exclusive intentions. *Teaching in Higher Education*, 12(4), 499-511.

McManus, D.A. (2001). The two paradigms of education and the peer review of teaching. *Journal of Geoscience Education*, 49(5), 423-434.

Orsmond, P. (2004). Evaluating our peers: Is peer observation a meaningful process? *Studies in Higher Education*, 29(4), 489-503.

Pagani, F. (2002). Peer review: A tool for global co-operation and change. *Observer*, 235, December.

Peel, D. (2005). Peer observation as a transformatory tool? *Teaching in Higher Education*, 10(4), 489-504.

Roberson, W. (Ed.). (2006). Peer Observation and Assessment of Teaching. The Center for Effective Teaching and Learning & Instructional Support Services, University of Texas at El Paso, :  
<http://academics.utep.edu/LinkClick.aspx?link=UTEP+Peer+Observation+Booklet.pdf&tabid=58396&mid=129801>

### The Evaluation of Teaching

Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems* (3rd ed.). Bolton, MA: Anker.

Chism, N. V. N. (2004). Characteristics of effective teaching in higher education: Between definitional despair and certainty. *Journal of Excellence in College Teaching*, 15(3), 5-36.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199.

Hatzipanagos, S., & Lygo-Baker, S. (2006). Teaching observations: A meeting of minds? *International Journal of Teaching and Learning in Higher Education*, 17(2), 97-105.

Knapper, C & Cranton, P. (2001) Fresh approaches to the evaluation of teaching. *New Directions for Teaching and Learning*, (2001),88.

Knapper, C. (2001).Broadening our approach to teaching evaluation.*New Directions for Teaching and Learning*, (2001), 88, 3-9.

McKinney, K., & Cross, P. (2007). *Enhancing learning through the scholarship of teaching and learning: The challenges and joys of juggling*. JB Anker Series. Wiley and Sons.

Seldin, P., & Associates. (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*. Bolton, MA: Anker.

Smith, K., & Tillema, H.H. (2007). Use of criteria in assessing teaching portfolio: Judgmental practices in summative evaluation. *Scandinavian Journal of Educational Research*, 51(1), 103-117.

Stake, R. E., & Cisneros-Cohernour, E. J. (2000). Situational evaluation of teaching on campus. , (2000), 83. 51-72.

Student Course Evaluations

Franklin, J. (2001).Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. *New Directions for Teaching and Learning*,(2001), 87 85-100.