**Reconceptualizing the Evaluation of Teaching in Higher Education**

**Abstract:**

Trends within higher education in the United States and Canada suggest that, although there are calls for recognition of teaching as a scholarly activity, teaching is not perceived as a significant aspect of scholarly work. Furthermore, policies, procedures, and criteria for the evaluation of teaching in higher education contribute to the marginalization of teaching within the reward structures of universities and colleges. Evaluation policies, procedures, and criteria tend to (1) emphasize technical, rather than substantive aspects of teaching, (2) focus on process rather than outcomes, (3) lack strategic concern for the use of evaluation data within the institution, and (4) are devoid of the very substance through which academics derive a sense of identity -- their discipline. Recommendations are offered for evaluating three aspects of teaching: planning, implementation, and results. Within each aspect, conceptual arguments and practical solutions are suggested for establishing criteria, deciding on sources of data, and determining the nature of data that must be gathered. The goal is to set in place evaluation policies, procedures, and criteria that will be perceived as rigorous and credible alongside more traditional forms of scholarship, while respecting the diversity of contexts and disciplinary identities within universities and colleges. Seven principles for evaluation of teaching are proposed.

NOTE: More information on this topic can be found in Pratt and Associates (1998), Five Perspectives on Teaching in Adult and Higher Education, Krieger Publishing Company, P.O. Box 9542, Melbourne, FL, USA 32902-9542. Figure 1 and parts of the text of this article appear in Chapter 11: Evaluating Teaching.

**Introduction**

We are in a curious and confusing time in higher education. On the one hand, institutions are awakening to demands that teaching be given more attention. One of the most hopeful contributions to this awakening and the re-valuing of teaching within the academy was sparked by the Carnegie Foundation for the Advancement of Teaching and the work of Boyer (1990) and Rice (1992). Their re-conceptualization of what is considered scholarship suggests four interdependent dimensions, one of which explicitly addresses teaching. Underlying this shift in thinking about scholarship is a belief that universities today must acknowledge, and reward, teaching as a legitimate and scholarly activity, alongside research and publication.

On the other hand, the tendency to subordinate teaching to other aspects of our work has never been greater. For example, in the United States, the number of faculty at research universities who believe it would be difficult to get tenure without publications has risen, over the last twenty years, from 44 percent, to 83 percent. This same trend holds within non-research institutions, doubling from 21 percent in 1970 to 44 percent in 1990 who believed it would be difficult to get

1

tenure based on the quality of one's teaching. (Boyer, 1990) Even the two and four-year institutions that identify themselves as "teaching institutions," are showing signs of a tendency among faculty to subordinate teaching to research. (Astin, 1991; Cheney, 1990, Daley, 1994) As these institutions are expanded into "university-colleges," and take on degree-granting status, their hiring policies attract larger numbers of faculty with their Ph.D. Consequently, with a great many applicants to choose from, the tie breaker is often a more advanced degree. What may not be considered in the decision to hire those with more advanced degrees is the enculturation they have gone through as part of getting that degree, and the effect it has had upon the individual's thinking about what kind of activity (and knowledge) is most valued.

Thus, while persuasive voices are calling for the recognition of teaching as a scholarly activity, faculty and administrators believe that teaching is not a valued activity within the reward structures of tenure and promotion, even at avowedly "teaching institutions." Additionally, the hiring practices associated with the expansion of two-year colleges into four-year university-colleges has increased the number of faculty who have been enculturated into norms of valuing research over teaching.

To make matters worse, prevailing conceptions of teaching, and current policies and practices of evaluating teaching in higher education, do little to counter the trends and forces mentioned above. While voices cry out for greater recognition of teaching in higher education, the very ways in which we think about teaching and evaluate its practice undermine the best intentions and most persuasive voices. How does this happen, and what must be done to change it? The intent of this article is to clarify the problem and reconceptualize the process of evaluation, so as to recognize teaching as a scholarly activity in higher education.

**Conceptions of Teaching**

Much of the research on effective teaching in higher education today focuses on conceptions of teaching. (e.g., Dall'Alba, 1991; Dunkin & Precians, 1992; Fox, 1983; Pratt, 1992a; Samuelowicz & Bain, 1992; Trigwell, Prosser, & Taylor, 1994) By 1995, at least thirteen empirical studies of conceptions of teaching in higher education had been published. (Kember, unpublished manuscript) Studies were conducted in several different countries, including Australia, Britain, Canada, The People's Republic of China, Hong Kong, Singapore, and the United States. Within these studies a number of conceptions of teaching emerged. Some authors found as many as nine different conceptions of teaching while other researchers found as few as four. In my own work (Pratt, 1992a; Pratt & Associates, in press), each of five qualitatively different conceptions of teaching were composed of a dynamic and interdependent trilogy of Actions, Intentions, and Beliefs.

The most visible part of someone's conception of teaching is their actions. Yet, conceptions of teaching are not defined by a particular set of actions, such as, lecturing, leading discussions, or demonstrating effective forms of practice. Actions are only a superficial representation of one's conception of teaching. Beneath the actions are a composite of intentions and beliefs which, though hidden from view, are crucial to understanding someone's conception of teaching. It is the collective interaction of actions, intentions, and beliefs that defines and distinguishes each

conception. Actions, without their accompanying intentions and underlying beliefs, are difficult to interpret and brimming with opportunity for misunderstanding.

Furthermore, conceptions of teaching are not confined to particular disciplines, contexts, or cultures. For example, conceptions of teaching that emphasize the imparting of information and structured knowledge have been found in the basic sciences, humanities, and social sciences. Similarly, faculty members holding conceptions that are concerned with facilitating understanding or conceptual change are also found in widely different professions, disciplines, and contexts. In other words, conceptions of teaching are associated more with the beliefs and intentions of the individual than with any particular disciplines or contexts.

Yet, despite the overwhelming evidence that ways of thinking about teaching and learning are rooted in deeper structures of intention and belief, for the most part, teachers within higher education assume that actions are the key to improving teaching and are reliable means of differentiating between effective and ineffective teaching. (Pratt, 1992a)

Many faculty development offices and programs operate on a similar assumption -- that good teaching can be defined in terms of a set of generic skills or actions. (e.g., TIPS -- Teaching Improvement Program for the Health Sciences) Effective teachers are expected to know how to develop goals or objectives, give lectures, ask questions, provide feedback, conduct discussions, provide examples, use audio visual materials, set reasonable exams and assignments, and so forth. As well, they are expected to treat students fairly, be accessible to students, and be enthusiastic about their course content. The assumption is that if teachers have sufficient knowledge of their content, are responsible in their use of time, and are willing to treat students fairly, all they need is a predetermined set of techniques and skills and they will be on the road to effective teaching. Thus, although qualitatively different conceptions of teaching have been documented in higher education, there seems to be a persistent belief that actions are the primary means of differentiating between effective and ineffective teaching.

Perhaps because of this, or in response to such limited understandings of teaching, current practices and policies governing the evaluation of teaching are often more concerned with a teacher's actions than with the intentions and beliefs that give direction and justification to teaching, resulting in technical, rather than substantive, approaches to evaluating teaching.

## Technical Approaches to Evaluation

### Focusing on 'duties'

In these times of educational accountability, one approach to evaluating teaching that has gained considerable attention is what Scriven (1995) calls the 'duties-based' approach. This approach cuts across conceptions of teaching and suggests that the best we can do, in the midst of philosophical and cultural diversity, is ask whether or not teachers are fulfilling their duties. Duties, as prescribed by those who take this view, are concerned with fair treatment of students, preparation and management of instruction, and evidence that students learn what is in the

objectives when they do the set activities. In response to a question about how to evaluate teaching while respecting and tolerating diversity, Scriven said:

In the recent teacher evaluation literature, there is a point of view which can handle the problem you raise, namely the 'style-free' approach. On this view, the only thing you can require of a teacher is the performance of duties, e.g., fairness in the treatment of students, effort in the preparation and management of instruction, and successful learning by students that do the prescribed activities. So the style [perspective] of teaching, e.g., didactic vs. inquiry, is totally irrelevant to teacher evaluation, and hence you can honor divergence while requiring success. The success is measured in the usual ways, with heavy but not exclusive use of student ratings at the pos-sec level.

(Michael Scriven <Scriven@AOL.COM>, 2 May 1995)

He goes on to explain that he is the author of this 'duties-based' approach and that the K-12 version has been through more than 40 revision cycles, using feedback from several thousand teachers and administrators. There is, he said, a less well worked out version for college-level teaching, though no source was given.

This approach seems to take the position that we cannot, in any morally and intellectually honest way, acknowledge all possible forms of 'good' teaching; there are too many possible variations on the 'good' of teaching. Thus, the only recourse for evaluators of teaching is to identify those who are negligent in the performance of their duty.

Some may say this is one way to be 'fair' to all conceptions of teaching; yet, I doubt that we could get teachers of philosophy, music, engineering, and medicine to agree on what are the most essential duties for a teacher. In fact, this approach ignores the very essence of different conceptions of teaching -- their beliefs about knowledge, learning, and the appropriate roles and responsibilities of a teacher -- and imposes a set of instrumental values that might be more amenable to some conceptions of teaching than others.

In addition, 'duty' is a socially constructed notion. One's sense of duty is bound up with one's social, historical, and cultural heritage. In China, for example, one's sense of duty as teacher often includes the establishment of a life-long relationship with students and a commitment to developing their moral character as well as their professional competence. (Pratt, 1991, 1992b; Wong, 1996) Many teachers in such societies would be insulted to have the notion of a teacher's 'duty' circumscribed in such limited ways as spelled out by Scriven. Furthermore, while Scriven's notion of duty may seem expedient and even appropriate for some institutions in North America, it perpetuates the belief that we cannot reliably differentiate between poor, adequate, and exemplary teaching. Logically, it follows that teaching, therefore, cannot be given serious consideration within the present reward structure of tenure and promotion.

**Focusing on 'technique'**

Another popular technical approach to evaluating teaching within higher education is the assessment of the surface aspects that cut across disciplines, contexts, and

philosophical perspectives of teaching. Faculty development workshops conducted through centralized offices focus, primarily, on the technical aspects of teaching -- planning, setting objectives, giving lectures, leading discussions, asking questions, communicating under difficult circumstances, and providing feedback to learners. These are important and, though not sufficient, are a necessary part of what makes an effective teacher.

As common as these technical aspects are to teaching across disciplines, contexts, and even cultures, there are many problems with taking this approach to evaluating teaching. First, we must agree on what technical aspects are universal and necessary to be a good teacher. 'Transmission' conceptions of teaching (e.g., Boldt, in press) suggests these should include: review and check of previous work, presenting new content/skills, guiding student practice, checking for student understanding, providing feedback and corrections, allowing for independent student practice, and giving weekly and monthly reviews. 'Apprenticeship' conceptions (e.g., Johnson & Pratt, in press) emphasize four different skills for teachers: modeling, scaffolding, fading, and coaching. 'Developmental' conceptions (e.g., Arseneau & Rodenburg, in press) emphasize the skills of diagnosing students' current ways of understanding and the building of conceptual and linguistic bridges from prior knowledge to desired ways of thinking and acting. While these are not all of the same 'technical' character implied earlier, they are, nonetheless, assumed to be universal aspects of teaching, applicable across disciplines, learners, and contexts.

This assumes that the role and responsibility of a teacher to represent and transform a particular body of knowledge to a particular group of learners is the same, despite the subject or the group of learners. It also ignores the obvious differences between disciplines and professional fields of study, and dismisses the differences between novices and advanced learners, laboratories and lecture halls, and teaching one student vs. one hundred students. Even the most generic of skills must bend to the conditions of who, what, and where the teaching is being done. Thus, we might have some difficulty reaching agreement on what should constitute the technical aspects against which all teachers should be judged.

In addition, much of the attempt to make evaluation more broadly applicable for personnel decisions results in diminished rigor. As evaluation procedures and criteria accommodate a wider range of disciplines and contexts they reduce the specificity of what it means to effectively teach "this group of students, this subject, in this context." When one looks only at the technical aspects of teaching, it matters little whether one is teaching English literature, medicine, or engineering. For example, I watched a friend in China teach for two hours in a language I didn't understand. At the end of the lesson, she asked for feedback on her teaching. Though I understood neither the language or the content, I was able to say something about the technical aspects of her teaching. I could see how often she asked questions, to whom she directed them, and what patterns of response occurred. I could see how much the 'discussion' centered around the teacher and how much of it spread horizontally to involve the students in the far corners and back of the room. I could comment on her use of the chalkboard and overhead projector, e.g., that she seemed to 'talk to the board' more than to the students. In general, I could say something (apparently) useful to her about her actions, without having understood a word of what was said.

5

I knew nothing about how she represented her discipline, i.e., her intentions and beliefs related to the content she was teaching. For example, I knew nothing about what she wanted people to learn, and why that was important; what crucial issues, key arguments, debates, and authors were to be considered; whose works were most central to those debates, whose works were omitted, and why. In short, I knew nothing of the substance of her teaching; only the surface techniques were visible to me. While those were important, they were not sufficient. Even as I talked with her after the session, I was aware of the superficiality of my comments. They contained only the most trivial aspects of what it meant to teach her content, to those students, in that particular context.

Consequently, the very aspect of teaching that is necessary for membership in one of Boyer and Rice's dimensions of scholarly activity, and from which most university faculty derive a sense of identity -- their disciplinary knowledge -- is left out of the dominant conceptions of teaching and the means by which it is evaluated. When the focus of evaluation moves away from the disciplinary home of academics, it decreases the likelihood that teaching will be recognized as a scholarly activity within the reward structures of our universities. Therefore, if evaluations are to be rigorous and credible they must acknowledge the essential and substantive aspects of one's discipline and/or profession, rather than just the most common attributes of teaching and learning.

Clearly, one cannot be a good or effective teacher without concern for duty and the technical aspects of teaching. These are necessary and important aspects of all conceptions of teaching. However, within such technical approaches to evaluation the focus is primarily on the actions of the teacher, with little exploration or concern for the deeper, underlying intentions and beliefs that give meaning and justification to those actions. Unless we understand what a person is trying to accomplish (intentions) and why they think that is justified (beliefs) we are very likely to misunderstand their actions. It is also unlikely we will have any meaningful way to evaluate or improve those actions. To do this, evaluations must go beyond duty and technique to assess the substantive, deeper aspects of teaching.

**Substantive Approaches to Evaluation**

**Academic Identity: acknowledging one's content**

As mentioned above, an essential ingredient left out of both 'technical' approaches is the very essence of most higher educators' identity -- their content. There can be no teaching without content; something (and someone) must be taught. Whether they teach in research universities, four year colleges, or two-year community colleges, most faculty think of themselves as a member of a profession, discipline, or trade, rather than as a teacher. (Becher, 1989) More often than not, they introduce themselves in terms of those associations, as historians, chemists, nurses, librarians, carpenters, and so forth. Their content is a pivotal aspect of their identity as an academic.

As a result, if evaluations are to be credible, the substantive aspects of academic identity and decision-making -- what is included/excluded from a course, what is emphasized/minimized, and what is assessed as evidence of learning -- must also be considered. Without an evaluation of the substantive aspects of teaching,

evaluators are likely to elevate duty and technique to a distorted sense of importance and omit that aspect of teaching upon which people build a life time of identity. In order to do this evaluators must consider evidence related to three aspects of teaching: planning, implementation, and results.

**Planning: assessment of intentions and beliefs**

**What to evaluate**

Within planning, there are at least five issues to evaluate. They each address the content and substance of teaching from a slightly different angle. Collectively, they reveal underlying intentions and beliefs about what is to be accomplished, why that is important, and how it is justified. They also provide a foundation upon which the next two aspects of teaching -- implementation and results -- can be evaluated.

- Mastery of content
  Selection of content
  Appropriateness of goals/objectives
  Appropriateness of course materials and requirements
  Articulation with other programmatic elements

The usual documents of teaching -- syllabus, assignments, examinations, projects, etc. -- are useful sources of information on planning. Within these documents one can see evidence of intentions and emphasis. However, the beliefs that provide justification and rationale for those decisions may be less apparent.

**Where to look**

One procedure for revealing the hidden beliefs within planning documents, is what Shulman and Hutchings call the 'reflective memo.' (1995) When someone's teaching is to be evaluated, s/he is asked to select a course and the most important assignment(s) for that course. With the course syllabus and assignment(s) in mind, the teacher writes a 'reflective memo' that discusses the goals for the course and how the assignment(s) will contribute to those goals. Prior to any observations of teaching, the instructor meets with the evaluators to discuss all three parts: syllabus, assignment(s), and reflective memo. The documents are used to open the discussion; the reflective memo is to take the discussion deeper.

If a reflective memo is to be useful it must probe the reasoning and justification that lies beneath the surface of the syllabus and assignment to reveal core intentions and beliefs. However, reflection on one's beliefs is particularly difficult. While beliefs are the most abstract aspect of commitment in teaching, they may also be the most significant. They represent underlying values held with varying degrees of clarity, confidence, and centrality. Some are vague and implicit; others are clear and readily explained. Some are held tentatively; others are considered incontestable. Some are marginal to the way a person thinks; others are central and even dominant.

However, the measure of centrality of a belief is not necessarily a matter of logic or rationality but, more often, the extent to which the belief itself is not in question. When a belief is held 'without question' it acts as 'arbiter' in determining whether

intentions, actions, or even other beliefs are reasonable and acceptable. Most teachers are able to accommodate a variety of changes in circumstances -- including changes in what they teach, whom they teach, and under what conditions -- as long as those changes do not challenge their 'core' beliefs, i.e., those most central to their values. The following questions are a sample of what can be used to help people reflect deeply upon their course materials.

- What are people expected to learn in this course? Why is that important?
- What crucial issues, key arguments, debates, and authors are to be considered?
- Whose work (authors, critics, practitioners, artists, etc.) is most central to the way you have developed this course and its content? Why?
- Whose work (authors) have you intentionally omitted? Why?
- What is the most important assignment in this course?
- Why is it important and how does it reflect your goals or intentions?
- What standards do you use in evaluating peoples' work on this assignment?
- How have your standards for this course changed over time? Why?
- What have you learned about improving this assignment, your course, or your teaching as a result of learners' responses to this assignment?
- How does your course link with or lay a foundation for other courses?

Neither the course documents or the reflective memo alone is as revealing as the combination of them. The teaching plan and assignment allow colleagues to review actual samples of teaching materials; the reflective memo supplies underlying thinking, and provides a context and framework for interpreting and judging substantive aspects of planning.

**Who to involve**

For some, this close scrutiny of teaching may run against assumptions and traditions of academic freedom in higher education. Yet, it seems ironic that academics are willing (and expected) to go public with other kinds of scholarly work, but close the door on their teaching. Teaching is, even more than other forms of scholarship, a public act within a community of students and colleagues that effects all members of that community. Those who claim it should be exempt from the kind of scrutiny given to research or publication unwittingly promote its continued marginalization within academic settings. If teaching is to count, it must be accountable to the same standards of evaluation as other forms of scholarship, and this must involve colleague and peer review.

Because judgments about teaching often have regard for the history, purposes, policies, and articulation with other program requirements within a department or institution, colleagues are usually asked to assess the quality of planning documents. (Centra, 1996) This is not an unusual occurrence. However, to increase their rigor and credibility, judgments must be perceived to be impartial and of high standards. One way to accomplish this is to have the same planning materials, and reflective memo, sent out for peer review by a respected academic, at arms-length, involved in similar work/teaching at another institution, much as is done with the review of academic writing.

The entire exercise is an example of how the evaluation of teaching can incorporate reflection on key commitments and belief structures. Incidentally, it also allows the evaluators to make explicit their own beliefs and commitments related to teaching and learning, providing a conversation among colleagues, peer evaluators, and the teacher about what constitutes good teaching.

**Implementation: assessment of the fit between actions, intentions, and beliefs**

**What to evaluate**

Within the implementation of that planning, there should be evidence of the technical skills and duties mentioned earlier, e.g., fair treatment of students, clarity of explanations, distribution of discussion, and so forth. However, the list of technical skills and duties should be derived in consultation with the person being evaluated, and it should complement what the person is trying to accomplish. In other words, evaluators should exit the examination of planning materials and reflective memo with a sense of what is to be accomplished, how it is to be done, and why that is important. The focus in evaluating implementation should be on how it is done. Thus, there should be evidence of agreement between the teacher's actions and the intentions and beliefs expressed in the planning documents and reflective memo.

**Where to look**

The usual way of gathering this evidence is through colleagues' observations. Regrettably, collegial observations have a poor record, and lack utility and credibility in the evaluation of teaching. As French-Lazovik reported in a review of literature on classroom observations in higher education:

The general finding is that it does not provide a sound method of evaluating the teacher's in-class activities. A few classroom visits by one colleague cannot be expected to produce a reliable judgment... Even when the number of colleagues is increased to three, and each makes at least two visits, the reliability of resulting evaluations is so low as to make them useless...(1975, p. 75)

This is not a surprising finding, given the usual way of conducting classroom observations in higher education. An observer (usually a colleague) enters the room armed with his or her personal and private 'yardstick' of good teaching against which this person's work is to be assessed. During the visit the observer may scribble notes or sit passively but, under the usual 'rules' of engagement, neither the notes or the thoughts of the observer are presented to the teacher as a way of explaining what was seen, what was considered important (or not), and what value was placed on the process and content observed. As a result, the usual process of observing a colleague is seldom rigorous or even compassionate; more over, the product is rarely helpful to the teacher or reliable to the evaluator.

**How to look**

The American Association for Higher Education (AAHE) Peer Review of Teaching Project offers several guidelines for improving observations by colleagues, many of

which would allow for increased rigor and credibility of those observations. They are presented below, with some additions and modifications, as part of a strategy for reducing the anxiety and increasing the reliability of implementation data.

Avoid 'parachuting' into a class, making your observations, and then exiting to make judgments without consulting with the teacher.

Make observations part of a consultation process which includes a pre- and post-visit meetings. The pre-visit meeting should establish the purpose of the observation, the goals for the class, and how those goals fit within the overall intentions for the course. Post-visit meeting should be a de-briefing to discuss what the observer saw. The teacher's own assessment of the session should be included in the discussion.

Link observations to reviews of teaching materials, the reflective memo, and discussions of these materials. Colleagues' observations and meetings with students should assess the internal consistency of a teacher's intentions, beliefs, and actions. At the end of it all, planning, implementation, and results should be complementary.

Conduct several observations, distributed across the entire term of a course, rather than a single snapshot look at one session.

Use a team approach, in which colleagues pair up or work in small groups to visit one another's classes over the term of a course.

As observers, be open to learning about different approaches to teaching while judging the quality of another's teaching.

Let students know what is happening and why; tell them the purpose of the evaluation and the role of observations in the process. (adapted from AAHE, 1995)

Notice that this method of observation is not the same as the 'glimpse' one would get through one or two observations of someone's teaching; nor is it entirely dependent upon a colleague's opinion of teaching. By being clear in purpose and procedure, to learners and teacher alike, this method offers a more open and reliable means of gathering data on implementation. In addition to increasing the likelihood that observations will focus on substantive aspects of teaching, it can also reduce the anxiety that usually accompanies peer observations. Most importantly, this approach is meant to complement the review of course materials and the teacher's reflections on intentions and beliefs.

**Who (not) to involve**

At this point, readers may be wondering why I have not recommended the involvement of students in evaluating the implementation of teaching. Certainly, this would be required if we were primarily concerned with the improvement of teaching. Student feedback is invaluable for helping instructors improve their technical skills and conduct in relation to Scriven's duties. However, because this article is concerned with summative, rather than formative, evaluation I have stressed the involvement of colleagues more than students. As you will see in the

next section, student or learner information is more convincing when it pertains to learning than when it addresses the process of teaching.

**Results: student assessment of learning, course, and instructor**

**Who to involve**

Overall, there is no more widely used source of data for evaluating teaching than student opinion. Learners are perceived by many to be the most reliable source of data about the relationship between teaching and learning on the grounds that they are witness to the teaching across time and the best judge of its effects on their achievements. (Cashin, 1988; Kahn, 1993) Student opinion is considered a necessary, and sometimes sufficient, source of evidence on which to judge the quality of teaching, even across diverse groups of learners, disciplines, and cultures. (e.g., Marsh, 1986; Marsh, Touron, and Wheeler, 1985; Watkins, 1992; Watkins, Marsh, and Young, 1987; Watkins and Thomas, 1991)

Yet, most student evaluations focus on the process rather than outcomes of teaching. For example, they often ask to what extent the teacher clarified goals and objectives, was organized and prepared, used time wisely, emphasized key concepts, provided timely feedback on assignments, was enthusiastic about the subject, and treated students with respect. While this may be useful information for the improvement of teaching, it is not very useful for assessing the effects of teaching. For example, it does not address what was taught and what was learned, the value of that learning, or the effects of teaching upon student learning. If student evaluations are to be credible, the data must reflect issues clearly related to the effects of teaching; and students must be a logical source of that data.

**What (and how) to evaluate**

With these conditions and cautions in mind, the literature on summative ratings of teaching suggests that students can provide reliable information on four topics:

An estimate of progress on course goals

Information on additional learning (beyond course goals)

An overall assessment of the value of a course

An overall rating of the instructor's effectiveness

Each will be discussed and illustrated through examples of questions and rating forms for gathering student opinion related to that topic.

**Topic 1. Estimating progress on course goals**

In virtually all teaching there should be some sense of purpose and direction, usually stated in the form of goals, objectives, aims, outcomes, or intentions. They may or may not be negotiated with students at the outset, but they should be an indication that teaching is planned and deliberate, and that students have been informed of the direction and intentions that will be guiding the course they are

11

taking. Students can then assess their progress on course goals, what was helpful to that progress, and what could have been done to help them make even better progress. A sample set of questions and response scale for soliciting student estimates of their progress on course goals is given below. (Figure 2) The blank lines are intended to be goal statements.

PROGRESS ON GOALS -- Using the scales below, rate your progress on each of the goals listed. These should be the same goals you were given at the beginning of the course. Then, comment on what was helpful and what else would have facilitated your progress related to these goals.

Rate your progress on each of the course goals listed below: (circle a number for each goal)

Goal (a)
_____
_____

Rate your progress: 0) none 1) little 2) some 3) average 4) good 5) excellent 6) extraordinary

Goal (b)
_____

Rate your progress: 0) none 1) little 2) some 3) average 4) good 5) excellent 6) extraordinary

What was particularly HELPFUL to your progress on these goals::

What else could have been done to FACILITATE your progress on these goals:

The sample form allows for only two goals; the number will vary, but probably shouldn't exceed ten. Goals that are rated should be the same as those given to students at the beginning of a course.

Prior to receiving the results of the evaluation, the instructor should select a few key goals and weight those same goals as an indication of emphasis within his or her teaching. Student learning (progress) can then be compared with an instructor's intentions, as indicated in the weighing of course goals.

**Topic 2. Assessing additional learning**

An assessment of progress on the stated goals is vital to the evaluation of teaching. However, good teaching very often reaches beyond the stated goals for its impact. Indeed, for some students evidence of progress on course goals, in the form of tests and assignments, as well as personal testimony, may be secondary to more important outcomes of teaching, e.g., an evolving sense of identity as a member of a professional 'community,' or having developed more complex ways of understanding one's discipline, or feelings of increased self-efficacy. (Pratt & Associates; in press) Figure 3 provides sample questions for gathering information on additional learning.

ADDITIONAL LEARNING -- If you can, identify any additional learning from this course or instructor that was particularly important to you. If you cannot think of any, leave this section blank and go on to the next part.

Within this course, was there something else you learned (in addition to the goals above) that was particularly important to you?

Why is that important to you?

For some students the unanticipated or 'incidental' learning may be more significant than what they learned related to course goals. Indeed, it may be an important indicator of the difference between adequate or even good teaching and exemplary teaching. Yet, this kind of data is usually omitted from student evaluations. Thus, for a broader, more encompassing estimate of the results of teaching, learners must be asked about outcomes that were not anticipated.

**Topic 3. Assessing the value of a course**

Another problem that has contributed to a lack of confidence in student data on teaching has been the contamination of evaluations of teaching with student opinion of a course or subject. It is possible to value a course or subject, but not the teacher; it is also possible to appreciate an instructor but not the subject or course. For example, learners may have taken a particularly critical course that opened new vistas of thinking; yet, their teacher may have been only marginally effective and not at all central to their awakening. Conversely, a teacher may have been critical for an individual, or even a group of students, but the subject or course may have been only incidental to that impact. Figure 4 shows how the value of a course can be assessed separately from judgments about teaching.

VALUE OF THE COURSE -- On the scale below, rate the value of the course. Then, comment on what was particularly valuable and what could be done to make the course more valuable.

Provide an overall rating of the VALUE of the COURSE: (circle one)

0) none 1) very 2) some 3) average 4) more than 5) far exceeded 6) among the best

little most courses most courses I've ever had

What was particularly VALUABLE:

What could be done to make the course MORE VALUABLE:

Again, if student evaluations of teaching are to be taken seriously, because they provide useful and reliable data on the quality of teaching, it is important to separate the perceived value of a course from the effectiveness of a teacher. It would be difficult to reconcile data that said a teacher was uncommonly effective, yet the course was of very little value; or if students said a teacher was only marginally effective, yet the value of the course far exceeded most courses. One

would hope there is agreement between these questions; but if there isn't, it is a clear indication for the need to inquire as to why there isn't.

## Topic 4. Rating the instructor's effectiveness

After students have assessed their learning and given an assessment of the value of their course, they should be asked to rate the effectiveness of the instructor. The order is important to this process. By asking about the effectiveness of the teacher last, students have had a chance to 'vent' any negative or positive feelings they might have that could confound the rating of the instructor (e.g., a required course or a favorite subject). Furthermore, this sequence allows them to consider their learning separate from the teacher's effectiveness, allowing for the influence of personal motivation, peer support, and other external factors known to effect learning. Figure 5 provides an example of how students can rate a teacher's effectiveness, while also giving information as to what was particularly effective and what might be done for this instructor to be more effective. Remember, all of this is in pursuit of a summative evaluation of teaching effectiveness.

EFFECTIVENESS OF THE INSTRUCTOR -- Using the scale below, provide an overall rating of your instructor. Then, comment on the ways in which s/he was effective and/or ways in which s/he could be more effective. As you can see, the highest ratings are reserved for those instructors perceived to be among the most effective in your program or in your experience as an adult learner.

Provide an overall rating of the instructor's EFFECTIVENESS: (circle one)

0) poor 1) marginally 2) sometimes 3) usually 4) always 5) one of my 6) exceptionally

effective effective effective effective most effective effective

What did the instructor do that was PARTICULARLY EFFECTIVE:

What could the instructor do to be MORE EFFECTIVE:

In addition to differentiating between the value of a course and the effectiveness of an instructor, a rigorous and credible evaluation should be able to distinguish between levels of effectiveness ranging from poor to exemplary teaching. The rating scales in Figures 4 and 5 are intended to do this, by allowing for course and instructor to be evaluated in four categories: poor (0), marginal (1-2), adequate (3-4), and exemplary (5-6). This is to counter the tendency with most evaluation forms to rate the majority of teachers at the higher end of the scale, thus loosing the ability to identify those who are 'uncommonly' effective. If learners' evaluations are to be useful and credible, they should differentiate between poor, marginal, adequate, and truly exceptional teaching. Thus, as is mentioned in the instructions to learners in Figure 5, the highest ratings are reserved for teachers who are perceived to be among the most effective -- in a program, an institution, or the learner's adult experience.

Through this kind of evidence we are asking students to tell evaluators to what extent their course is valuable, their instructor effective, and whether their learning

is significant. In terms of an evaluation's credibility, these are questions that must be addressed; they are also questions that only students can answer.

## Being Strategic

Two final notes about policies and procedures for evaluating teaching: first the need to consider the costs and benefits associated with this more elaborate approach to the evaluation of teaching; and second, the need to involve teachers in their own evaluation. Each of these is a strategic part of evaluating teaching.

## Involving key individuals

Clearly, the time and resources needed to assess and document the quality of teaching in the ways described here go well beyond the usual learner evaluations and colleague observations. In many institutions there will not be support for such an expenditure of diminishing resources on the evaluation of teaching. Yet, this is a catch twenty-two. Until such time as the evaluation of teaching is rigorously and credibly documented, it will not be considered a legitimate scholarly activity; but, until it is valued alongside the more traditional scholarly activities, resources will not be spent in documenting its substantive nature. Therefore, we must be more strategic in the development and implementation of evaluation procedures.

One way to begin that process is by developing evaluation criteria and procedures that help the person writing the report that goes forward within the institution. Too often, the work of evaluation has focused only on helping the faculty member document his or her teaching, without giving enough thought to the one who must write the recommendation that goes forward on behalf of that person. For example, in most of higher education the key person responsible for making a case for re-appointment, tenure, or promotion is the department or division head. He or she is responsible for taking the evidence (and possibly a personnel committee's recommendation) to the next level in the institution. In effect, this individual becomes the chief advocate for the person who is under review. Evaluation criteria, procedures, and evidence should facilitate this individual's work. Indeed, without this strategic concern, evaluations are at risk of being ignored or dismissed, thus perpetuating the catch twenty-two of teaching not being valued because of the ways in which it is conceptualized and documented, yet not being rigorously documented because it is not highly valued.

Self-evaluation: the teaching portfolio

A final source of information about teaching effectiveness that should be included is the teacher's self-evaluation. If teachers are to participate in all that has been said thus far, they should also have the opportunity to more broadly document their efforts and accomplishments in teaching. One approach to this is called the teaching portfolio and holds great promise. It had its start in Canada, under the sponsorship of the Canadian Association of University Teachers (CAUT) and was intended to expand the body of evidence administrators could consider when making personnel decisions. According to the original guide called The Teaching Dossier, it is a "summary of a professor's major teaching accomplishments and strengths. It is to a professor's teaching what lists of publications, grants, and academic honors are to research" (Shore, et al. 1986, p. 1), a kind of "extended

teaching resume...a brief but comprehensive account of teaching activity over a defined period of time." (AAHE, 1991, p.3). A great deal has been written on the subject (e.g., Foster, Harrap, & Page, 1983; King, 1990; Seldin, 1991; Stark & McKeachie, 1991; Vavrus & Collins, 1991) and does not need to be repeated here. However, if teaching portfolios are to be taken seriously, they must be perceived as rigorous pieces of evidence about teaching. How can that be accomplished?

One way in which teaching portfolios can be more rigorous is to incorporate reflection as part of the substance upon which judgments are made. Rather than judging the volume of work submitted, or the sheen of its lamination, evaluators might ask teachers to provide evidence of growth and change, success and failures, plans and aspirations, with reflective comments that take the evaluator deep into the substance and reasoning of the teacher's evolving thinking and approaches. In other words, much of what has been said thus far, applies as well to self-evaluations; they must rigorously document substantive aspects of teaching if they are to be credible.

## Summary

While some authors argue convincingly that teaching should be recognized as scholarly work, many institutions of higher education evaluate teaching in ways that trivialize teaching as a set of technical skills, and/or rely heavily on sources of data that are suspect in the eyes of administrators and others involved in tenure and promotion decisions. The challenge of this article has been to suggest guidelines and principles that increase the rigor and credibility of evaluations while honoring the diversity of content, context, and perspectives on teaching. It would be relatively easy to offer evaluation guidelines that are equitable but which do not concern themselves with rigor. Technical approaches based on duties and/or technical aspects of teaching do this. Yet, good teaching is rigorous; to be less than that in its evaluation is to perpetuate its marginal position in the reward structures of higher education and be blind to the substantive aspects that characterize truly effective teaching.

To be rigorous in the evaluation of teaching requires a fundamental change in approach -- one that shifts the focus of evaluation from surface features to deeper structures, and one that asks 'why' as well as 'how.' Without this crucial shift in approach, teaching will continue to be seen as a relatively mechanistic activity, devoid of its most essential ingredient -- the substance of our professional identity.

The benefits, therefore, from equitable, credible, and appropriately respectful evaluation processes are legion. It simply makes sense, from the learners', the teachers', and the institutions' points of view to embrace evaluation methods that attend to substantive rather than technical aspects of teaching. With that in mind, I close with a list of seven principles which, if used to guide evaluation, support both the reality of multiple, legitimate perspectives about excellence in teaching, and promote its rigorous and credible evaluation.

## Seven Principles for Evaluating Teaching

Principle 1: Evaluation should acknowledge and respect diversity in actions, intentions, and beliefs..

Principle 2: Evaluation should involve multiple and credible sources of data.

Principle 3: Evaluation should assess substantive, as well as technical, aspects of teaching.

Principle 4: Evaluation should consider planning, implementation, and results of teaching.

Principle 5: Evaluation should parallel other forms of judging scholarly work.

Principle 6: Evaluation should contribute to the improvement of teaching.

Principle 7: Evaluations should be done in consultation with key individuals responsible for taking data and recommendations forward within an institution.

## References

American Association for Higher Education (AAHE) (1991). The Teaching Portfolio: Capturing the Scholarship in Teaching. Washington DC: The American Association for Higher Education.

American Association for Higher Education (AAHE) (1995). From Idea to Prototype: The Peer Review of Teaching -- Choosing from a menu of strategies, Washington DC: The American Association for Higher Education.

Arseneau, R. and Rodenburg, D. (in press). 'The developmental perspective', in Pratt, D.D. and Associates, Five Perspectives on Teaching in Adult & Higher Education. Malabar, FL: Krieger Publishing.

Astin, A. (1991). The American College Teacher. Los Angeles: Higher Education Research Institute.

Becher, T. (1989). Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines. Milton Keynes: SRHE and Open University Press.

Boldt, A. (in press). 'The transmission perspective', in Pratt, D.D. and Associates, Five Perspectives on Teaching in Adult & Higher Education. Malabar, FL: Krieger Publishing.

Boyer, E. (1990). Scholarship Reconsidered. Princeton, N.J.: Carnegie Endowment for the Advancement of Teaching.

Cashin, W.E. (1988). Student Ratings of Teaching: A Summary of the Research. Idea Paper No. 20. Manhattan: Center for Faculty Evaluation & Development, Kansas State University.

Cashin, W.E. and Downey, R.G. (1995). 'Disciplinary differences in what is taught and in students' perceptions of what they learn and of how they are taught', in Hatva, N. and Marincovich, M.(eds.) Disciplinary Differences in Teaching and Learning: Implications for Practice, New Directions for Teaching and Learning, number 64, Winter, San Francisco: Jossey-Bass, pp. 81-92.

Cheney, L. (1990). Tyrannical Machines. Washington, D.C.: National Endowment for the Humanities.

Dall'Alba, G. (1991). 'Foreshadowing conceptions of teaching', Research and Development in Higher Education 13, 293-297.

Daley, W.T. (1994). 'Teaching and scholarship: adapting American higher education to hard times', Journal of Higher Education 65(1), 45-57.

Day, C., Calderhead, J. and Denicolo, P. (1993) (Eds.). Research on Teacher Thinking: Understanding Professional Development. Washington, D.C.: The Falmer Press.

Dunkin, M.J. & Precians, R.P. (1992). 'Award winning university teachers' concepts of teaching'. Higher Education 24, 483-502.

Foster, S.F., T. Harrap, and G.C. Page (1983). 'The teaching dossier', Higher Education in Europe 8(2), pp. 45-53.

Fox, D. (1983). 'Personal theories of teaching', Studies in Higher Education 8(2), 151-163.

French-Lazovik, Grace (1975). Evaluation of College Teaching: Guidelines for Summative and Formative Procedures. Washington, DC: Association of American Colleges.

Johnson, J. and Pratt, D.D. (in press). 'The apprenticeship perspective', in Pratt, D.D. and Associates, Five Perspectives on Teaching in Adult & Higher Education. Malabar, FL: Krieger Publishing.

Kahn, S. (1993). 'Better teaching through better evaluation: a guide for faculty and institutions'. To Improve the Academy, 12, 111-126.

Kember, D. (in press). 'A Reconceptualisation of the Research into University Academics' Conceptions of Teaching'. Learning and Instruction, Pergamon Press.

King, B. (1990). Linking Portfolios with Standardized Exercises: One Example from the Teacher Assessment Project. (Technical Report). Stanford, CA: Stanford University, School of Education, Teacher Assessment Project.

Marsh, H.W. (1986). 'Applicability paradigm: students' evaluations of teaching effectiveness in different countries', Journal of Educational Psychology, 78, 465-473.

Marsh, H.W., Touron, J. And Wheeler, B. (1985). 'Students' evaluations of university instructors: the applicability of American instruments in a Spanish setting', Teaching and Teacher Education, 1, 123-138.

Pratt, D.D. (1991). 'Conceptions of self within China and the United States', International Journal of Intercultural Relations, (USA) 15 (3), 285-310.

Pratt, D.D. (1992a). 'Conceptions of teaching', Adult Education Quarterly, 42(4), 203-220.

Pratt, D.D. (1992b). 'Chinese conceptions of learning and teaching: A westerner's attempt at understanding'. International Journal of Lifelong Education (UK) 11 (4), 301-319.

Pratt, D.D. & Associates (1998). Five Perspectives on Teaching in Adult & Higher Education. Malabar, FL: Krieger Publishing.

Rice, R.E. (1992). 'Toward a broader conception of scholarship: the American context', in Whiston, T.G. and Geiger, R.L. (eds.), Research and Higher Education in the United Kingdom and the United States, Lancaster, England: Society for Research into Higher Education, pp. 117-129.

Samuelowicz, K. & Bain, J.D. (1992). 'Conceptions of teaching held by academic teachers'. Higher Education, 24, 93-111.

Scriven, M. (1995). Personal communication, May 2.

Seldin, P. (1991). The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions. Boston, MA: Anker Publishing.

Shore, B.M., et al. (1986). The Teaching Dossier: A Guide to Its Preparation and Use. Revised edition, Montreal: Canadian Association of University Teachers.

Stark, J.S., and W. McKeachie (1991). National Center for Research to Improve Postsecondary Teaching and Learning: Final Report, Ann Arbor, Mich.: NCRIPTAL.

Shulman, L. and Hutchings, P. (1995). 'Exercise I - teaching as scholarship: reflections on a syllabus', in Hutchings, P. (ed.) From Idea to Prototype: The Peer Review of Teaching -- A Project Workbook, Washington DC: The American Association for Higher Education.

Trigwell, Kl, Prosser, M. & Taylor, P. (1994). 'Qualitative differences in approaches to teaching first year university science', Higher Education, 27, 75-84.

Vavrus, L., and A. Collins. (1991). 'Portfolio documentation and assessment center exercises: a marriage made for teacher assessment', Teacher Education Quarterly, 18(3), 13-39.

Watkins, D. (1992). 'Evaluating the effectiveness of tertiary teaching: a Filipino investigation', Educational Research Journal, 7, 60-67.

Watkins, D., Marsh, H.W. and Young, D. (1987). 'Evaluating tertiary teaching: a New Zealand perspective', Teaching and Teacher Education, 3, 41-53.

Watkins, D. And Thomas, B. (1991). 'Assessing teaching effectiveness: An Indian perspective', Assessment and Evaluation in Higher Education, 16, 185-198.

Wong, W. (1995). 'Use of teaching portfolio to capture your scholarship in teaching', Teaching-Learning Tips, Issue 10/95, Educational Technology Centre, The Hong Kong University of Science and Technology, Kowloon, Hong Kong.

Wong, M. (1996). A study of traditional Chinese apprenticeship: A look at five Chinese Masters. Masters Paper, The University of British Columbia, Vancouver, B.C., Canada.



Department of Educational Studies
Faculty of Education, UBC
2125 Main Mall, Vancouver, BC V6T 1Z4

http://www.edst.educ.ubc.ca/faculty/pratt/DPrecon.html