Classroom performance can be represented by peer observation reports, videotapes of selected classes taught by the candidate, or 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 (Sell & Chism, 1988). The validity of peer observational reports will be increased if they also serve a formative evaluation function (Batista, 1976). Observational reports must also be accompanied by a reflective commentary by the candidate that includes reactions to what was written by observers. Peer observation will be discussed in more detail in chapter 4.

Documenting evaluation of student learning and providing feedback on their progress might include selecting an exemplary 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 that 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.

The publish or perish tradition has distracted faculty from writing about their teaching experiences. Yet through the experience of teaching—the trials and errors—much is learned about effective teaching practice. Most academic fields have a journal devoted to publishing faculty reports on their teaching experiences. Some are more empirically rigorous than others, yet 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. 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 by Edgerton, *et al*., include:

- Keep size "lean and clean" by applying the following principle to all data considered for inclusion, "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 of portfolios and the entire evaluation process as connected to the improvement of practice and toward "developing a more professional discourse about teaching."

- Consider the work involved in developing a portfolio to be an integral part of faculty responsibility, both for the candidate developing it and for colleagues evaluating it.[8]

Through the prudent selection of work samples, a teaching portfolio should be able to clearly reflect a candidate's teaching and teaching developmen. It should do so in a way that does justice to the range of responsibilities and activities engaged in by the candidate without resulting in a cumbersome, redundant and lengthy document. The key to preserving representativeness while restricting bulk is in the establishment of a clearly articulated set of criteria and categories, which should be negotiated between the candidate and department at the point of initial employment. Once these have been established and the data categories selected, there are fundamental principles that have been developed through research and experimentation which should govern the collection and form of presentation of that data. We turn now to a detailed discussion of data collection and its representation.

---

[8] Adapted from R. Edgerton, P. Hutchings, and K. Quinlan (1991). *The Teaching Portfolio—Capturing the Scholarship in Teaching* (Washington, D.C.: American Association for Higher Education).

Chapter III—Supporting Data: Collection and Presentation

To increase the efficiency of the review process for all involved, it is imperative that the supporting data 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 are presented can itself bias the evaluative outcome. Fortunately this is an area that has stimulated abundant 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. "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."[1] Key in the use of student data is the notion that as a data source it is only one component available for a committee to make an informed judgment. When incorporated into a thorough analysis, student evaluation data are useful not only because they represent the learner's perspective, but they 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, were the more prevalent form in the tenure files analyzed during the preparation of the report, *Evaluation and Recognition of Teaching*. Qualitative student evaluation data, in the form of written letters of evaluation, appeared in less abundance, although when such data were provided they took up a lot of space in the file—in one case a tenure file contained 88 student evaluation letters.

The potential to statistically manipulate quantitative data is a mixed blessing. Assuming the data were gathered properly—using questionnaire items that were validated to ensure that what was being measured by the question was, in fact, what the question purported to measure—quantitative data can be very efficient in that many teaching factors from many individual perspectives can be presented in relatively little space. In the analysis of tenure files conducted for the report, many of the guidelines listed below were not followed, however, which resulted in a students' picture of the candidate's teaching that was not only limited but possibly biased.

---

[1] Peter Cohen (1981). " Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies." In *Review of Educational Research*, 51, no. 3: 305.

Instrumentation

Research in the area of student evaluation of instruction has resulted in the publication of more than 2,500 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 contain fewer items than formative questionnaires. 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) rather than frequency scales (Frequently, Somewhat Frequently, Rarely, Never).

The use of "core" items allow an individual's scores to be compared to scores determined from a group aggregation, such as interdepartmental, 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 following:

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 is one of the best Cornell teachers I have known.
The instructor clearly interprets abstract ideas and theories.
The instructor demonstrates a favorable attitude toward students.
The instructor is willing to experiment and be flexible.
The instructor encourages students to think for themselves.

Administration of Questionnaires

Research on questionnaire validity suggests that if the following guidelines are followed for administering student 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)[2]

[2] G.R. Sell and N. Chism (l988). *Assessing Teaching Effectiveness for Promotion and Tenure: A Compendium of Reference Materials*, Center for Teaching Excellence (Columbus, Ohio: Ohio State University), 5-6.

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 interpreta tion.
- normative data sets should be established yearly for course type (elective, required, lecture, lab, etc.) on both a department level and college level for comparison with a tenure candidate's own scores.
- appropriate normative data should be provided wherever possible

Figure 7 below is an example of a simple format for reporting student evaluations scores for a single course.

Instructor: Professor X	**Course: 100, sec 22**	Spring '92		# choosing each response				
		Standard		Strongly		neutral		strongly
Item	**Mean**	**Deviation**	**Count**	disagree	disagree	opinion	agree	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

Figure 8 is an example of a visually clear way of reporting a candidate's relative standing in relation to departmental normative data.

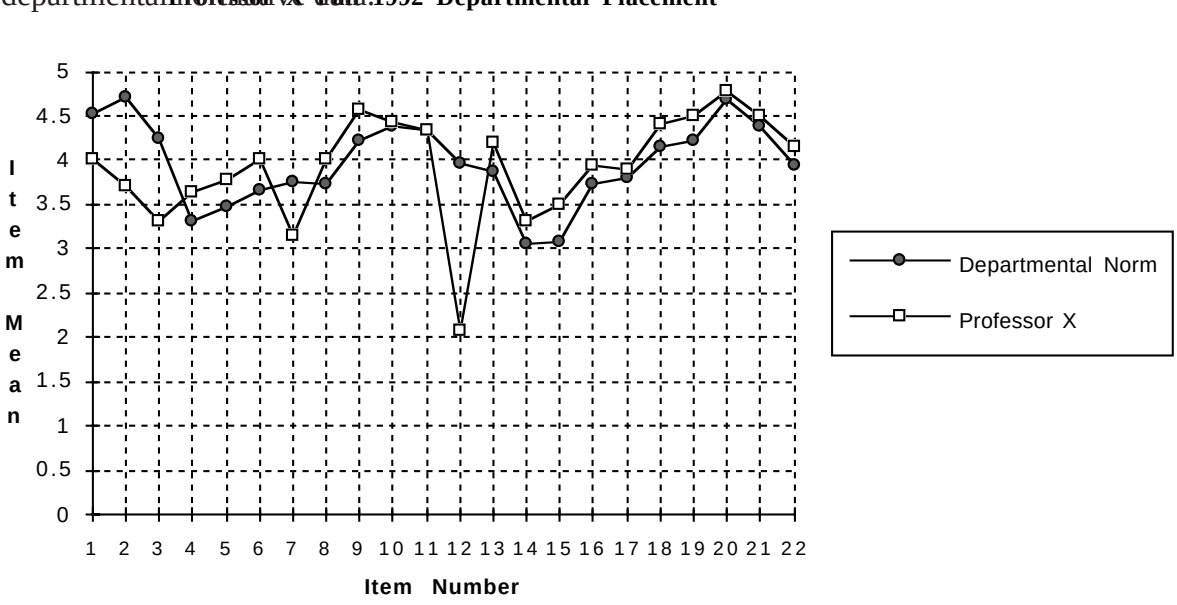

Figure 8

Qualitative data is less generalizable and harder to aggregate because it is in a more open-ended form. Its potential bulkiness can be reduced through a synthesis by an objective individual familiar with the case, such as a department head. For others who must review this kind of synthesized data and who are less familiar with the candidate's situation, such as deans and administrators outside of the department, a supplementary reflective statement from the candidate synthesizing the student letters can be useful in concert with the department head's report. If these reports are well written and address major developmental issues in the candidate's teaching practice, the time necessary to write them is well justified, especially if their creation leads to improved practice. 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*, the deans of the colleges were interviewed. One dean raised the issue of anonymity of student evaluation data. Quantitative questionnaire scores, combined with letters of recommendation, provide a good balance of general and specific information. Letters in their original form, however, 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 they may be identified at some point by the candidate during the tenure decision process. If letters by students are returned to someone other than the candidate—the department head or *ad hoc* chair, for example—and if they are then keyboarded on a computer and summarized by an independent person (a member of a departmental standing committee on teaching) and students are informed that these precautions are being taken when they are asked to write their letters, the validity of their responses will be enhanced.

An example of a department chair's synthesis of 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.[3]

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:

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.

---

[3] A Report of the Select Committee, Jan. 14, 1992. *Evaluation and Recognition of Teaching,* Appendices, (Ithaca, N.Y.: Cornell University), 21.

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 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 data base 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 20 years (Seldin, 1984). During that time peer review has taken on an increasingly significant role in the tenure 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 evaluation of 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 who represent the discipline, will be required to provide their own data.

To be most effective, the peer evaluation process should be neutral, open, relatively unthreatening and structured, all of which can be enhanced through the use of standardized rating and observation procedures and criteria. Standardization is a precaution stimulated by the evidence that colleagues' ratings may not be statistically reliable. In one study (Centra, 1979), the average correlation between colleagues was .26 per item. Another study[4] revealed the potential for positive bias of peer evaluation: fifty-four teachers were evaluated on the basis of two classroom visits by each of three different colleagues, which showed that 94 percent of all ratings were in the top two categories of a 5-point scale.

Several factors are critical in ensuring a valid and fair peer review process. What questions are asked and answered by the reviewers is central. Some kind of replicable protocol is necessary to ensure fairness and accountability for the process. This is true for whatever data is being reviewed, whether course

[4] J. A. Centra, (1975). "Colleagues as Raters of Classroom Instruction," *Journal of Higher Education* 46: 327-337.

materials, classroom performance or student learning. 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 the following:

> 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.[5]

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 is not qualified to carry out that task. A consistent finding of peer observation studies is that observers should have some kind of training that prepares them for this responsibility. Peers typically are 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. The following questions may be useful to undertake a specific peer review task:

- Do you believe you can properly judge the teaching-learning process in the classroom visited?
- Would you recommend this instructor to students advised by you? Why or why not?
- What specific changes are needed to strengthen teaching performance?
- How would you rate this instructor against others teaching similar courses in the department?

---

5 Peter Cohen and Wilbert McKeachie. "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. The dictionary uses the terms "standard, rule or test" in defining criterion. 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 improve instructional practice?"

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."[1]

"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."[2] Based on these definitions, if tenure candidates sufficiently prove their capacity and commitment to continually improving research and 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 that 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
- yield information about instructional situations where the teacher functions best[3]

This chapter provides suggested criteria to use to evaluate teaching, based on the data categories described in chapter 3. It begins with a general discussion of effective teaching and goes on to include criteria for use by students, criteria relevant to evaluation of 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 25 years.

---

1 J. Aubrecht, Delaware State College: personal communication.
2 G. Leinhardt. (1991)." Evaluating the New Handbook of Teacher Evaluation" *Educational Researcher*, 20, 6, p. 24.
3 J. D. McNeil and W. J. Popham (1973). " The Assessment of Teacher Competence." In R.M.W. Travers, ed. *Second Handbook of Research on Teaching* (Skokie, Ill.: Rand McNally), 218-244.

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 who do not know what I am talking about.

The dichotomy can be avoided by a more integrative model of teaching: effective teachers are able to understand enough about their students' ways of thinking that they can translate their own understanding of the subject matter into a form that connects with 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.[4]

Recent work on teacher effectiveness has yielded the following observations which support an integrative model that is both process- and content-based (italics in original):

> *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

[4] L. Shulman (1989). "Toward a Pedagogy of Substance," *AAHE Bulletin*, June, 11.

> 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 . . .[5]

The last two points raised in the quotes above deserve to be emphasized. 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. 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. Work load 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 or 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?[6]

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:

---

5 Andrew Porter and Jere Brophy (l988). "Synthesis of Research on Good Teaching: Insights from the Work of the Institute for Research on Teaching."*Educational Leadership*, 78-83.

6 G. French-Lazovik (l981). "Peer Review: Documentary Evidence in the Evaluation of Teaching." In *Handbook of Teacher Evaluation*, J. Millman, ed.(Beverly Hills: Sage Publications), 77-78.

whether there is sufficient data on teaching quality
whether alternative teaching methods have been explored
whether changes have been made in the candidate's courses over time
whether the candidate sought aid in trying new teaching ideas
whether the candidate developed special teaching materials
whether the candidate participated in teaching improvement opportunities

A study carried out at Berkeley (Hildebrand, Wilson & Dienst, 1971) was designed to discriminate between best and worst teachers. One set of scales were factor-analyzed out of student survey data that are relevant to evaluating teaching by students. They are:

Scale 1: *Analytic/Synthetic Approach*, relates to scholarship, with emphasis on breadth, analytic ability, and conceptual understanding.
Scale 2: *Organization/Clarity*, relates to skill at presentation, but is subject-related, not student-related, and not concerned merely with rhetorical skill.
Scale 3: *Instructor-Group Interaction*, relates to rapport with the class as a whole, sensitivity to class response, and skill at securing active class participation.
Scale 4: *Instructor-Individual Student Interaction*, relates to mutual respect and rapport between the instructor and the individual student.
Scale 5: *Dynamism/Enthusiasm* relates to the flair and infectious enthusiasm that comes with confidence, excitement for the subject, and the pleasure in teaching[7]

A second set of scales were factor-analyzed out of survey data from faculty colleagues. These surveys were also designed to discriminate between the best and worst teachers. These scales are relevant for use by colleagues in evaluating a candidate's teaching. They are:

Scale 1: *Research Activity and Recognition*
Scale 2: *Intellectual Breadth*
Scale 3: *Participation in the Academic Community*
Scale 4: *Relations with Students*
Scale 5: *Concern for Teaching*[8]

The criteria identified with each scale for use by students and colleagues and which were the most discriminating between the best and worst teachers are included in the tables below. (Factor analysis coefficients that were used to associate the item with the particular scale are included.)

Student Evaluation Criteria

Table 1[9]

Components of Effective Teachers as Perceived by Students*
[scale 2 is not included here because it is colleague, and not student related]

Scale 1. Analytic/Synthetic Approach		Factor coefficient
1.	Discuss points of view other than their own	.70
2.	Contrast implications of various theories	.66

[7] Milton Hildebrand, Robert C. Wilson, Evelyn R. Dienst (l971) *Evaluating University Teaching*, Center for Research and Development in Higher Education, University of California, Berkeley, 18.

[8] *Ibid*., p. 20.

[9] *Ibid*., pp. 18-19.

3.	Discuss recent developments in the field	.64
4.	Present origins of ideas and concepts	.60
5.	Give references for more interesting and involved points	.53
6.	Present facts and concepts and related fields	.53
7.	Emphasize conceptual understanding	.46
8.	Explain clearly	.78
9.	Are well prepared	.63
10.	Give lectures that are easy to take notes in	.62
11.	Are careful and precise in answering questions	.61
12.	Summarize major points	.51
13.	State objectives for each class session	.50
14.	Identify what they consider important	.47

Scale 3. Instructor-Group Interaction

15.	Encourage class discussion	.70
16.	Invite students to share their knowledge and experiences	.65
17.	Clarify thinking by identifying reasons for questions	.64
18.	Invite criticism of their own ideas	.62
19.	Know if the class is understanding them or not	.58
20.	Know when students are bored or confused	.57
21.	Have interest and concern in the quality of their teaching	.48
22.	Have students apply concepts to demonstrate understanding	.43

Scale 4. Instructor-Individual Student Interaction

23	Have a genuine interest in students	.74
24.	Are friendly toward students	.71
25.	Relate to students as individuals	.69
26.	Recognize and greet students out of class	.69
27.	Are accessible to students out of class	.65
28.	Are valued for advice not directly related to the course	.64
29.	Respect students as persons	.60

Scale 5. Dynamism/Enthusiasm

30.	Are dynamic and energetic persons	.80
31.	Have an interesting style of presentation	.76
32.	Seem to enjoy teaching	.74
33.	Are enthusiastic about their subject	.65
34.	Seem to have self-confidence	.64
35.	Vary the speed and tone of their voice	.63
36.	Have a sense of humor	.53

*Based on 1968 Survey, N = 1015

These items can be used either to develope end-of-semester summative evaluation questionnaires or to evaluate other student data on teaching, such as letters. If a numeric evaluation schema is adopted,

caution should be exercised in establishing normative data. "The usual overall evaluation of teaching will provide for evaluation on a five-point scale and will permit a classification of teachers as poor, adequate, good, excellent, or outstanding. In practice, the bottom end of the scale is rarely used and the actual range varies between a little under 3.0 to a little over 4.5 That is, anything under 3.0 is poor, and anything over 4.5 is outstanding; the other classifications are arranged in between these two extremes. . . . The concept of improvement implies progressing up the scale."[10]

Evaluation Criteria for Use by Peers

Making global assessments of an instructor's overall teaching effectiveness is a practice that is unsatisfactory to the candidate, to those who must make the evaluation, and to the department. A more useful and practical practice is for colleagues to focus on certain qualities associated with good teaching that they are in a good position to judge. The items listed below were those most discriminative between best and worst teachers as perceived by their colleagues (Hildebrand, et al., 1971). They are included here to provide a general profile of effective teaching from which a department may develop its own profile. Because these items discriminated between the best and worst teachers at the $p < .001$ level they have a high level of validity. The authors of the study suggest they be used as a supplement (and not as a substitution) for student ratings.

Table 2[11]
Components of the Activities of Effective Teachers
as Perceived by Colleagues*

	Scale 1. Research Activity and Recognition	Factor coefficient
1.	Do work that receives serious attention from others	.69
2.	Correspond with others about their research	.69
3.	Do original and creative work	.64
4.	Express interest in the research of colleagues	.55
5.	Give many papers at conferences	.55
6.	Keep current with developments in their field	.49
7.	Have done work to which I refer in teaching	.48
8.	Have talked with me about their research	.38
	Scale 2. Intellectual Breadth	
9.	Seem well read beyond the subject they teach	.66
10.	Are sought by others for advice on research	.60
11.	Can suggest reading in any area of their general field	.59
12.	Know about developments in fields other than their own	.51
13.	Are sought by colleagues for advice on academic matters	.43

[10] A. Sullivan (l985). "The Role of Two types of Research on the Evaluation and Improvement of University Teaching." In Arthur Sullivan and J. Donald, eds., *Using Research to Improve Teaching: New Directions for Teaching and Learning* (no. 23) (San Francisco: Jossey-Bass), 76.
[11] Hildebrand, et al., pp. 21-22.

Scale 3. Participation in the Academic Community

14.	Encourage students to talk with them on matters of concern	.60
15.	Are involved in campus activities that affect students	.58
16.	Attend many lectures and other events on campus	.47
17.	Have a congenial relationship with colleagues	.39

Scale 4. Relations with Students

18.	Meet with students informally out of class	.58
19.	Are conscientious about keeping appointments with students	.57
20.	Meet with students out of regular office hours	.57
21.	Encourage students to talk with them on matters of concern	.55
22.	Recognize and greet students out of class	.37

Scale 5. Concern for Teaching

23.	Seek advice from others about the courses they teach	.70
24.	Discuss teaching in general with colleagues	.60
25.	Do not seek close friendships with colleagues (Negative)	-.47
26.	Are people with whom I have discussed my teaching	.45
27.	Are interested in and informed about the work of colleagues	.44
28.	Express interest and concern about quality of their teaching	.40

*Based on 1967 survey, N = 119

Once a candidate has been evaluated on these or other criteria, certain precautions are necessary to ensure fairness: "include a review of central tendencies and variations in the rating results; an analysis of the effects of ecological factors, including different types of courses, students, and time frames on ratings in the unit; and the establishment of agreed-upon standards and steps to be taken in the application of the standards."[12]

Classroom Observation by Peers

What happens in the classroom can have a substantial impact on student relationships with the course material. It is therefore important to add to students' and the candidate's own perspective a third view of classroom performance by peers through planned observations. Studies seeking to determine whether peers can reliably and validly evaluate classroom performance through observations have been discouraging, however. "It is not clear . . . whether the validity and reliability of classroom observation procedures warrant their being considered as a legitimate approach for summative evaluation."[13]

Reliability and validity of classroom observations can be enhanced if guidelines are established that address the following issues: how many visits and when are they carried out; who does the observing;

---

[12]A. Sullivan (l985), p. 16.

[13] Peter Cohen and Wilbert McKechie, p. 148.

how are observers selected and how many people are involved in the observations; what is observed, and, consequently, what is the character of the observational report; and to whom do observers report?

The following guidelines can enhance the quality of classroom observation by peers:

1. use with caution, training of observers is suggested to minimize bias
2. use several observations by several people over time
3. select observers with no biases (use multiple observers)
4. observations should be done with prior notification of candidate
5. observational criteria should be oriented towards currency/accuracy of material & ethical conduct (content & professionally oriented) rather than stylistically/ rapport-oriented
6. records of colleague observational data should be summarized with explicit descriptions of the context of the observation[14]

Staff in the Office of Instructional Support have developed a protocol for classroom observation and performance review, which they have taught successfully to many individuals. This protocol is based on a cognitive development paradigm that fosters the improvement of practice, rather than on a remedial approach that limits teaching to a set of technical skills. More will be said about this process in chapter 5. The following questions are included to assist the department in developing a comprehensive and consistent classroom observation protocol.

Classroom Observation Items

Structure and Goals
The instructor was fully prepared for class.
The instructor provides an overview of what is planned for the class period.
The instructor emphasizes the conceptual basis of the material.
The instructor's lectures are well organized.
The instructor provides periodic summaries of what has been covered or discussed.
The instructor uses class time efficiently.
The instructor ties things together at the end of class.
The instructor chooses appropriate activities for learning the material.

Teaching Behaviors
The instructor asks questions that encourage students to think about the subject.
The instructor is animated.
The instructor clearly explains instructions for completing required tasks.
The instructor leaves enough wait time after asking questions for students to think of a response.
The instructor uses eye contact effectively.
The instructor provides clear and comprehensive explanations when required.

[14] G. R. Sell and N. Chism, p 8.

Instructor-Student Rapport
The instructor encourages students to ask questions and express their opinions.
The instructor gives clear and understandable responses to students' questions.
The instructor seems genuinely concerned about the students' learning.
The instructor is actively helpful when students need assistance.
The instructor is skillful at promoting interaction among students.
The instructor is able to involve everyone in the class.
The instructor listens carefully to student questions and comments.
The instructor knows when students seem confused.
The instructor provides clear, relevant and understandable responses to student questions.
The instructor periodically checks to make sure everyone understands what has been covered.
The instructor is able to involve everyone in the class, not just the most outspoken students.
The instructor is interested in students as individuals.
The instructor listens carefully to student questions and comments.
The instructor holds students' attention.

Subject Matter and Instruction
The instructor stimulates interest in the subject matter.
The instructor relates various topics of the course to each other.
The instructor uses real-life anecdotes and examples to illustrate abstract ideas.
The instructor creates a classroom atmosphere conducive to learning.
The instructor seems enthusiastic about teaching the material.
The instructor makes effective use of props, visual aids, illustrations and examples.
The instructor demonstrates command of the subject matter.

Establishing evaluation standards and norms requires additional precautions: "Teachers regarded as excellent by some observers and poor by others should be rated by as many observers as possible. . . . Norms should be calculated at the campus level for some elements of any evaluation form used in promotion procedures. . . . Departments or subject areas might find it useful also to calculate their own norms, particularly if they have developed their own evaluation forms, but it is desirable that any norms used be recalculated at frequent intervals to assure that the system of evaluation is being responsive to change."[15]

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 skills necessary to effectively evaluate student learning.

Suggested criteria for evaluating teaching and course materials[16]

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?

[15] Hildebran, *et al*. p. 38.
[16] R. Miller (l987). *Evaluating Faculty for Promotion and Tenure*. (San Francisco: Jossey-Bass).

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 lively?
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 papers are graded fairly.
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-homeexaminations.
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 in their papers and projects the principles learned in the course?
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 overall 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 varied?
3. Do students use the library for the course?
4. Would audiovisual or television services 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 department goals?
4. Is it appropriate to the course level?

Once a set of evaluation criteria has 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 to ensure 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[17]

Sources of information	**Percentage of Total Evaluation**
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 improvement of practice should be a major criterion by which a tenure candidate is evaluated. Just what that means and how it can be accomplished, documented and evaluated is the subject of the last chapter.

---

[17] Adapted from Peter Cohen and Wilbert McKeachie, p. 152.

Chapter V—Improving Practice: Case Examples

Improvement of teaching is the result of an integration of refinements in how teachers think about teaching: What are their beliefs about how they can influence student learning? What have they come to know and understand, based on experience and experimentation in action? To what degree do they see teaching as an important part of their scholarly work? Reflection helps teachers develop their professional knowledge, which will in turn affect their future instructional planning. 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. Improvement may not always be indicated by a standard deviation in higher student evaluation scores. Improvement grows out of reflection on feedback about one's teaching. 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 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 aware of it, assimilate an adaptive attitude that places a low priority on teaching in order to survive. When such a thing happens, everybody loses: the candidate, the students, and the department that has made the investment in that teacher.

Reflection on practice typically takes place in the planning stage of teaching—when the teacher is identifying course objectives, writing examinations, designing individual class teaching plans and choosing readings. 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."[1] His basic premise is that 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 includes monitoring student interest and comprehension and adjusting teaching practice accordingly. Using misconceptions to guide

---

1 Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action* (New York: Basic Books), 276-78.

instruction is another reflective approach to developing one's teaching. 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 new ways.

Some people may observe and reflect as a matter of course, but given the heavy content orientation that graduate education tends to reinforce, chances are that most reflection on action is limited to trial and error or substantive thinking. Reflection on the process of learning the subject matter and what is psychologically necessary to teach it adequately is crucial 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 teacher may have come to the conclusion, based upon reflecting on teaching practice, that no amount of time spent exclusively on the subject matter and limited to the teacher's frame of reference will be very productive from a learning point of view. 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 from what I understand to be their level of comprehension to my own. From a planning point of view, this would involve a different set of questions for setting class objectives: 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 into it?

Regardless of their graduate training or experience, all teachers bring to their practice what might be called personal theories of teaching and learning (Cornett, Yeotis, Terwilliger, 1990). These theories may be only partly conscious, but the authors who study teacher thinking suggest that teachers' personal theories are what guide them in decisions about planning, course design, and classroom behavior. 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, just as we gravitate toward 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 about central stylistic tendencies can be thought of as indices of our instructional development. We can read books that may influence the way we think about teaching, but we learn how to do it and improve through trial and error practice—by, as one author describes it, "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), which is what 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 mundanely, "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 idea can be illustrated with a frequently used teaching technique we all are familiar with— "Are there any questions?" We usually ask this question 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 also with veterans, particularly when teaching something for the first time, the question proves ineffective. What may be forthcoming, rather than a flood of questions, 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 strategy is in need of change, much less pinpoint the specific problem. Acknowledgment may slowly dawn in me from repeated experience, 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, "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 have noticed 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 notes 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—such as "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 that I look down at my notes when I check for understanding might help me further conceptualize this situation. If I ask myself why I am looking at my notes at that point, I might begin to recognize that I use the time right after asking, "Are there any questions?" to plan my next move. 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 [1] D.