

AC 2008-153: EVALUATING INSTRUCTIONAL SCHOLARSHIP IN ENGINEERING

Norman Fortenberry, National Academy of Engineering

Norman Fortenberry is the founding director of the Center for the Advancement of Scholarship on Engineering Education (CASEE) at the National Academy of Engineering. CASEE is a collaborative effort dedicated to achieving excellence in engineering education--education that is effective, engaged, and efficient. CASEE pursues this goal by promoting research on, innovation in, and diffusion of effective models of engineering education.

Tylisha Baber, Michigan State University

At the time this paper was written, Dr. Tylisha Baber was serving as a National Academies Christine Mirzayan Science and Technology Policy Fellow. She earned a B.S. degree in chemical engineering from North Carolina State University and a Ph.D. in chemical engineering from Michigan State University. Tylisha's dissertation focused on the design and implementation of a biomass conversion process for improving the fuel properties of biodiesel. She is currently an adjunct assistant professor in the Department of Mechanical and Chemical Engineering at North Carolina A&T State University.

Evaluating Instructional Scholarship in Engineering

Abstract

There is a window of opportunity to engage the engineering community in a discussion of the metrics by which to measure scholarly teaching. We place emphasis on those summative metrics that could be used by administrators in judging new faculty rather than on strategies that would formatively help new faculty to document their teaching results and improve their practice. Our view is that unless a discussion is held on the potential evaluative metrics, it will be difficult to achieve the desired level of attention to the importance of high quality teaching. Therefore, it is in the long-term interests of new faculty that this precursor discussion targeting administrators be held.

Teaching is a multifaceted activity that is best evaluated using multiple measurement techniques and criteria. In general, there are six key steps in the development of a highly reproducible instructional evaluation system:

1. Determine the purpose of the evaluation;
2. Define the aspects/dimensions of teaching to be evaluated;
3. Identify valid sources of data or evidence for each aspect of teaching being evaluated;
4. Specify the criteria, or measuring instrument, by which the aspects will be judged;
5. Analysis and interpretation of data by skilled, trained personnel; and
6. Set weights, or scoring mechanism, for each aspect of teaching being evaluated.

The first five of these steps are examined in this paper; the sixth step is left for future work.

In this paper, the sources, types, reliability and validity of data used for summative evaluations, including peer review, student ratings, and self-assessment, were examined.

I. Background

Boyer describes at least four forms of scholarship: discovery, synthesis, application, and teaching^[1]. Our focus here is on the scholarship of teaching, discussion of which is often done with comparisons to the scholarships of discovery and synthesis. It is generally accepted that evaluation of discovery and synthesis within academic departments depends heavily on “counting” publications in refereed journals. The major debates center on the significance of the publications and the quality of the journals^[(2), p. 40-41].

The scholarship of teaching has been refined by Shulman into “scholarship of teaching and learning” and “scholarly teaching”^[3]. The former is essentially the scholarship of discovery within the domain of education^[4]. Our focus here is on the latter, “scholarly teaching” which is distinguished from “teaching” by its focus on teaching practice and learning outcomes, grounded in disciplinary content and pedagogic knowledge, reflective critique, commitment to communication to peers and openness to peer evaluation^[(2), pp. 87-88]. Scholarly teaching holds the promise of enhanced student learning through rigorous faculty attention to learning. Because tenure and promotion depend upon evaluations of scholarship and because compared to evaluation of the scholarships of discovery and synthesis scholarly teaching is more difficult to assess, faculty may perceive few incentives to devote their efforts to scholarly teaching^{[(2), p 14], [5]}.

Our assumption is that if valid and reliable means/mechanisms by which to achieve this outcome were identified and if they were accepted by the engineering community, then greater attention would be devoted to scholarly teaching by engineering faculty and departments. Our assumption is consistent with the NRC assertion, in the context of science instruction, that, “embracing and institutionalizing effective [teaching] evaluation practices can advance the recognition and rewarding of teaching scholarship.”^[(2), p. 17]. Achieving such recognition and rewards would be of value to new faculty. Therefore, it is in the long-term interests of new faculty that this precursor discussion targeting administrators be held.

Our particular focus is scholarly teaching by engineering faculty. The NRC has recognized that there are distinct differences in the ways the various disciplines approach teaching and learning^[(2), p. 5]. Three linked events have prompted changes in the instructional environment for engineering faculty and created an opportunity for increased attention to evaluation of teaching:

- In the late 1990’s the system of accreditation for undergraduate engineering degrees shifted from a focus on process to a focus on student learning outcomes^[6].
- The engineering community became increasingly aware of the broad response in the academic community to Boyer’s authorship of *Scholarship Reconsidered*^[11] and efforts by the US Department of Education along with the National Research Council to a) define rigorous education research^[7] and b) gather evidence and highlight the significance of recent findings on how people learn and the significance of those findings for education^[8], and
- Engineering education began a transition from cycles of “reform” to exploring use of a rigorous research base to inform the shape and content of the undergraduate engineering curriculum^[9].

As a result of these events, the engineering profession has substantially increased its level of practical experience and sophistication in scholarly approaches to teaching^[5].

Most importantly, the engineering accreditation system possesses a criterion that explicitly speaks to the instructional quality of faculty^[10]. Specifically, engineering faculty contributing to a program of study leading to a degree are expected (by implication, collectively) to possess the knowledge and skills required in order to a) properly guide a program of study, and b) develop and implement processes for the evaluation, assessment, and continuing improvement of the program, its educational objectives and outcomes. Beyond proper grounding in the content knowledge of the discipline, this criterion implies the examination of faculty pedagogic content knowledge that is a hallmark of scholarly teaching^[(2), pp. 16].

This changed environment was the focus of discussion at the June 2006 annual meeting of the American Society for Engineering Education. The opening plenary focused on “Advancing Scholarship in Engineering Education: Launching a Year of Dialogue”^[11].

Given these recent events, there is a window of opportunity to engage the engineering community in a discussion of the metrics by which to measure scholarly teaching. We place emphasis on those summative metrics that could be used by administrators in judging new faculty rather than on strategies that would formatively help new faculty to document their teaching results and improve their practice. Our view is that unless a discussion is held on the

potential evaluative metrics, it will be difficult to achieve the desired level of attention to the importance of high quality teaching.

At present, there is, relatively little agreement on how the quality of teaching should be evaluated for summative purposes^[12]. As currently performed, teaching evaluations often result in skewed profiles of a faculty members' teaching effectiveness. Thus, in order to maintain focus on quality education, a standardized evaluation system of scholarly teaching must be developed. Peer review, student ratings, and self assessment are the three most common sources of data used to measure teaching effectiveness^[13]. This paper reviews key criteria and guidelines for each type of evidence that can be taken into consideration when making summative judgments about the quality of teaching.

II. Reliability and Validity

Evaluations for summative purposes are only useful if the data they provide are reproducible. Standard criteria for determining reproducibility include reliability and validity. Reliability means consistency – an evaluation always must give similar scores, ranking, or ratings for similar tests, regardless of the evaluator or the evaluated^[14]. The keys to ensuring reliability are: 1) explicit statements of criteria, and 2) training evaluators to apply the criteria. Both are necessary for a reliable evaluation system^[15]. Validity addresses whether the interpretation, uses, and actions based on assessment results are appropriate. Validity is generally considered in at least two aspects—face validity and content validity. Face validity is an assessment of whether the survey items are clear and understandable to a general population similar to the one targeted to take the survey. Content validity refers to whether the items make sense to experts in survey research (to make sure that they don't confound multiple issues within a single question) as well as to the specific population targeted by the survey (that is, the survey items make sense in context). Research findings indicate the importance of developing evaluation systems that demonstrates reliability, validity, and that promote fairness^[16-19]. Thus, criteria for evaluating teaching will be more useful in the tenure and promotion process if they can reliably and consistently measure a specific competency both for the same instructor over time and between instructors.

III. Peer Review

Within research communities, peer review has been the long-accepted method of judging the quality of scholarly activity. In recent years, teaching has been re-conceptualized in terms of the notion of scholarship, which has brought about the need to include the views of peers in decisions about what quality teaching is. The peer review of teaching is a process in which faculty peers review a colleague's teaching effectiveness and performance through two major activities- direct classroom observations and course material evaluation. Peer review can be used for either developmental purposes to improve teaching (formative) or for use in decision-making about merit, promotion, and/or tenure (summative). Both formative and summative observations can be based on the same observation instruments.

A. Direct Classroom Observation

Types of Measuring Instruments

Classroom observation is a process in which a classroom observer records the performance of an instructor and an expert analyzes and interprets the records using some conceptual framework. Planning and implementing a systematic approach to observation reduces bias and improves reliability. The three most common instruments used in peer observation of teaching are checklists, rating scales, and open-ended narratives (written analysis)^[20-22].

Checklists

Checklists are structured evaluation instruments that can be used to determine whether particular behavioral, physical, or environmental characteristics are present during classroom instruction. Typically, desirable behaviors are described briefly and an observer checks whether each behavior is observed during a class session. Checklists are similar to rating scales with the exception of rubric scoring. The categories within a checklist are interrelated in that some of the same or similar kinds of questions occur in each category, enabling consistency in feedback across different forms of delivery and aspects of teaching^[23]. Checklists are often viewed as too restrictive by both observers and those observed^[20]. Furthermore, they only indicate that a behavior has been observed with no immediate feedback about its effectiveness. They can, however, help standardize observation procedures making the review more reliable. Common categories of checklist items used as criteria for evaluating classroom teaching performance and effectiveness include^[24-35]:

- Preparation & Organization
- Content & Knowledge
- Interaction
- Rapport & Sensitivity
- Presentation (Enthusiasm, Clarity)
- Assessment of Student Learning/Performance/Achievement

Rating Scales

Rating scales add the dimension of indicating relative effectiveness of a given behavior to checklists. The two categories of rating scales that have been widely used in instructor evaluation are (1) general ratings that are all-inclusive regarding the merit of the instructor and (2) specialized ratings that consider traits, qualities, or specific behaviors which are believed to reflect the teaching ability of the instructor^[36]. The former is considered to be subjective, failing to provide suitable consideration and weighting to all the important factors contributing to quality teaching. The latter is often criticized with regard to the method of selecting and weighting the factors. In other words, there is often disagreement about which traits, behaviors, and qualities are essential to quality teaching and the extent to which each contributes^[36]. Thus, to avoid ambiguity and to ensure reliability, all rating sheets need to include detailed behavior descriptors and an explanation of the scale points^[20].

Open-Ended Narratives

The written analysis format affords an open-ended opportunity for a peer not only to select what to observe, but also, how to interpret the information and structure the evaluation. The down-side to written appraisals is that peers can comment on very limited behaviors, focus only on one criteria, or reflect only the observer's personal approach to teaching^[20]. Thus, for summative purposes, the blank sheet observation is not reliable and therefore is not recommended^[26].

Summary of Evaluation Instruments

It is suggested that a combination of checklists, rating forms, and written analysis be adopted and used for both formative and summative observations^[20]. Seldin^[37] recommends a combination of two instruments. All instruments should be pertinent to the teaching situation, yield accurate insights into the classroom environment, and obtain data typical of classroom behavior.

General Guidelines for the Three-Step Observation Process

It is recommended that peer classroom observations be used as simply one part of the larger picture with regard to evaluation of teaching effectiveness^[9]. If peer observations are to be used for summative purposes, then a formal structure is required in order to ensure fairness and reliability. Furthermore, a fair and effective peer evaluation program requires a considerable investment in faculty time. The minimum amount of time suggested is 4 to 6 hours per evaluation, including a minimum of 1-hr for the pre-observation meeting, 1-hr for the observation meeting, 1-hr for the post-observation, and 1-hr to write the final report^[38]. Based on research studies and the experience of institutions in which peer review of teaching is practiced, the following elements seem to be essential^[39-41]:

- Peer observers should be neutral and well-trained.
- Observers must use standardized observation reports to ensure reliability.
- Teams of at least two colleagues should perform the observations and should make their observations independently of each other (i.e. they should not compare notes until after the observation is completed).
- During the actual review, observers should carefully record their observations, impressions, and questions.
- At least three observation visits in a course for a single semester/quarter are necessary for adequate sampling.
- Procedures should require pre-observation and post-observation conferences between the observer(s) and the instructor.
- Data should be analyzed according to a single, coherent, public, agree-upon conceptual framework. The framework must be linked to student learning.

The first step in the peer classroom observation process is that a pre-observation meeting should be held. The purpose of this meeting is to establish a mutual understanding of the procedures to be used during the actual observation, which promotes a good and effective relationship between all parties involved. A pre-observation meeting also serves to provide context for the peer observation with respect to the review itself, such as identifying the focus of the review, the aspects of teaching to be reviewed, the criteria to be used, and the review instruments to be used,

as well as the context for the class being observed, such as the goals and learning objectives, when and where the review will take place, and the type of students enrolled in the class. [13, 23, 32, 34, 38, 42, 43]

Once the purpose of the review is identified, the actual classroom observation is conducted. During the actual review, observers should carefully record their observations, impressions, and questions. Observations should include indication of when they occurred and their framing context. Observations should be limited to the aspect(s) of teaching agreed upon during the pre-observation meeting. The observer's notes should be converted to a more narrative form as soon as possible after class. [13, 23, 43]

The final stage of the classroom observation process is the post-observation meeting, which allows for feedback and discussion. This meeting is often best initiated by the person being observed by stating their observations and reflections on the class first. Observers should provide honest, non-judgmental, descriptive feedback in order to reduce defensiveness on the part of the observed. The effectiveness of feedback may be enhanced by use of the "praise sandwich" approach in which strengths are highlighted at the beginning and end of the meeting and areas for further development are discussed in the middle of the meeting. The use of specific examples during feedback promotes reflection about achievable changes in teaching strategies. The observer should also write a descriptive summary report for the reviewing authority. [13, 23, 34, 44]

B. Evaluation of Course Materials

In addition to classroom peer observation, faculty members involved in summative review also examine course materials. Review of course material involves critiquing the non-classroom aspects of a colleague's approaches to teaching and learning. All course materials should be organized in a teaching portfolio or dossier. Course material can be broken down into three main categories [45-47]:

- Materials that illustrate course organization, structure, goals, policies, and procedures (e.g., syllabi, textbooks, handouts, outlines, overheads, slides, reading and reference lists, computer-assisted teaching tools, statements of teaching philosophy, reflection papers and other supporting resources).
- Materials that illustrate course content. (e.g., assignments, projects, exams, presentations).
- Materials that assess student performance and learning outcomes (e.g., samples of student work, student evaluations, student interviews).

A fourth category can be related to the efforts to improve teaching:

- Materials that illustrate contributions to curriculum development. (e.g., workshops, seminars, courses, publications, description of teaching innovations, teaching awards).

C. Who Should Evaluate?

An important consideration in selecting a reviewer is the purpose of the review. If the review is summative, then important considerations in selecting reviewers include professorial rank, objectivity, credentials, and experience. For reviewing course content, reviewers should be from the same discipline or sub-discipline^[42]. Possible options include an academic peer (internal, external), administrator, and a non-university individual (business/industry). Effective peer observation requires training in observational and analytical skills. Training helps the faculty member observe correctly by focusing on specified teaching criteria. Faculty colleagues serve as the best judge for the following dimensions of teaching quality^[37, 45, 47-51]:

- Mastery of course content.
- Selection of course content
- Goals, content, design and organization of the course.
- Appropriateness of course objectives.
- Appropriateness of instructional materials.
- Appropriateness of evaluative devices to foster and measure student learning (i.e., exams, written assignments).
- Appropriateness of methodology used to teach specific content areas (i.e., pedagogical strategies).
- Commitment to teaching and concern for student learning.
- Student achievement, based on performance on exams, projects, presentations and reports.
- Support of departmental instructional efforts.
- Professional and ethical behavior.

Articles on peer evaluation of teaching suggest that all of these aspects can be used during the peer review process. A review of these aspects can be facilitated by a teaching portfolio.

D. Weaknesses

The process of peer observation of classroom teaching is heavily dependent upon the goodwill and professionalism of all parties involved. However, perhaps the greatest fear of faculty members is that colleagues will criticize their teaching based on difference in theoretical perspectives or methodical issues in a way that could infringe upon academic freedom^[52]. The major weaknesses of peer observation of classroom teaching are^[20]:

- Data is often biased due to previous data, personal relationships, or peer pressure
- Peer relationships may suffer
- Possible bias due to observer's preference for own teaching methods
- Peer observations often tainted by reputation of instructor
- Colleagues tend to be generous in ratings

IV. Student Ratings of Teaching Effectiveness

Quantitative student ratings of teaching are used more than any other method in the evaluation of quality teaching^[53, 54]. However, many faculty members discredit the validity and value of student ratings. Common criticisms include that student ratings do not correlate with quality of learning, students only give high ratings to courses with low standards, and students do not really appreciate a good course until they are out of college. Nearly 30 years of extensive and rigorous research studies of student ratings have been conducted and the results collectively show that ratings are reliable, valid, stable over time, and positively correlated with results obtained using other forms of teaching assessment, including assessment of learning outcomes and peer reviews^[55-58]. Thus, the major validity problem is the misuse or misinterpretation of the ratings by faculty or administrators^[59]. Even though student ratings are essential in assessing teaching quality, they should be used in combination with multiple sources of data if a fair judgment of all of the components of college teaching is to be made.

A. Criteria or Dimensions of Teaching Effectiveness

Student rating forms are multidimensional- meaning they measure several aspects of teaching. There is debate about how many or which dimensions should be used for personnel decisions. Marsh's SEEQ (Students' Evaluations of Educational Quality) form contains nine dimensions^[49], including:

- Learning/value
- Enthusiasm
- Organization
- Group interaction
- Individual rapport
- Breadth of coverage
- Exams/grades
- Assignments
- Workload

Both Centra^[61] and Braskamp & Ory^[41] identified six factors commonly found in student rating forms:

- Course organization and planning
- Clarity, communication skills
- Instructor student interaction, rapport
- Course difficulty, workload
- Grading and examinations
- Student self-rated learning

Theall & Franklin^[62] provided a similar set of features that students are qualified to rate:

- Frequencies of instructor behaviors
- Amount of work required
- Difficulty of the material
- Clarity of lectures

- Value of readings and assignments
- Clarity of the instructor's explanations,
- Instructor's availability and helpfulness

A study conducted by French^[63] found that the 10 items which contributed most to student overall-judgment at the University of Washington were:

- Interprets abstract ideas and theories clearly
- Gets students interested in the subject
- Has increased student's skills in thinking
- Has helped broaden student's interests
- Stresses important material
- Makes good use of examples and illustrations
- Motivates students to do their best work
- Inspires class confidence in his or her knowledge of the subject
- Has given students new viewpoints or appreciations
- Is clear and understandable in his/her explanations

Crawford and Bradshaw^[64] identified the four most common characteristics that students felt best described the most effective instructor: (1) thorough knowledge of subject matter, (2) well planned and organized lectures, (3) enthusiastic, energetic, lively interest in teaching, and (4) student-oriented, friendly, willing to help students. Components of effective teaching were also identified by Cohen^[65] from analysis of student ratings, including skill, rapport, structure, difficulty, interaction, and feedback. Feldman categorized student rating items into as many as 28 dimensions^[67].

There are a number of features related to teaching effectiveness for which students are unqualified to make informed judgments. Cashin^[67] listed 26 specific considerations which he regarded as relevant to teaching effectiveness; students were unqualified to provide valid feedback for 11 of these, including subject expertise, course design, and curriculum development. Similarly, Cohen and McKeachie^[48] identified 10 criteria of teaching effectiveness in which colleagues, not students, could assess. Moreover, Keig and Waggoner^[68] characterized Cohen and McKeachie's criteria for which student are unable to judge validly into three categories: (1) the appropriateness of course goals, content, and organization of course design, (2) methods and materials used in delivery, and (3) evaluation of student work.

B. Evaluating Instruments

The most important consideration in teaching evaluation is the use of multiple methods involving multiple sources of data. Below is a list of 7 ways to evaluation teaching performance based on student ratings^[47, 69, 70].

- End-of course rating forms and written comments
- Alumni letters and surveys
- Focus-group interviews, exit interviews, and surveys of students
- Retrospective student ratings of courses and instructors (e.g. pre-graduation ratings by seniors)

- Mid-course and periodic student feedback.
- Evaluation of student learning.
- Minute papers

The three most well-known student rating forms are the Student Instructional Report (SIR)^[61], the Student Evaluation of Educational Quality (SEEQ)^[60], and the Instructional Development and Effectiveness Assessment (IDEA)^[56].

C. Possible Sources of Bias

Factors that may introduce systematic error or bias in student ratings of teaching should be taken into consideration when analyzing the validity of the ratings. Cashin^[71] suggests that bias in student ratings should be restricted to variables that are not related to or not a function of the instructor's teaching effectiveness. Sources of bias that are correlated with student ratings should be controlled by using appropriate comparative data. The following are biasing factors that may influence student ratings^[57-61].

Variable Type			
Instructor Faculty Rank Expressiveness	Student Motivation Expected Grades Interest Preparation	Course Level of Course Academic Field Work Load/Difficulty	Administrative Non-Anonymous Ratings Purpose of the Ratings Instructor's Presence

In general, the following variables suspected of biasing student ratings tend to show little or no relationship to the ratings^[56, 58, 72].

Variable Type			
Instructor Age Teaching Experience Gender Race or Ethnicity Personality Research Productivity	Student Age Classification Gender GPA Personality	Course Class Size Time of Day	Administrative Time During the Term

D. Guidelines and Recommendations

Research supports the fact that student ratings are statistically reliable, valid, and relatively free from bias. They are an important source of evidence for teaching effectiveness. The following is a collection of recommendations and guidelines from the literature of how student ratings should be constructed, administered, and interpreted^[72-75].

- Student ratings should be supplemented by peer observations, alumni ratings, self-evaluations, and portfolios containing descriptions of course materials, teaching methods, samples of students' work, and other evidence of teaching effectiveness.

- Provide comparative data, preferably for all the dimensions.
- Statistically control bias factors (extraneous factors) that can influence the ratings.
- Use the Likert 5-point scale: (1) poor, (2) satisfactory, (3) good, (4) very good, (5) outstanding.
- Ratings should be anonymous.
- The instructor should leave the room while students complete their forms.
- The data should be interpreted with the assistance of someone knowledgeable about educational measurement.
- The form should be distributed during the last two weeks of the term, but not right before or after final exams.
- Results should not be given to professors until after students have received their final grade.
- To ensure creditability of the results, a minimum of 75% of the registered students in a class should complete the rating forms.
- The ratings should be gathered from several courses over several semesters to assure a reasonably accurate assessment of a professor over time.
- Several open-ended questions should be included on the rating form to allow students to respond in their own words.

V. Self Assessment and Teaching Portfolios

Reflective practice or self-assessment by faculty members is an essential component of a systemic evaluation system. Braskamp and Ory^[41] suggest the following as sources of evidence that self-reflective assessments can provide about one's teaching:

- The way a course has been designed
- The strengths and weaknesses experienced during the implementation of that design
- Changes or adjustments made during the implementation and the rationale behind those alterations
- Ways in which course assignments met or failed to meet course objectives and expectations
- Assessment issues including student learning outcomes and periodic course evaluations.

Self-assessment can easily be organized for review in a teaching portfolio or dossier. A teaching portfolio is a "factual description of a professor's major strengths and teaching achievements. It describes documents and materials which collectively suggest the scope and quality of a professor's teaching performance."^[76] The information contained in a portfolio can be divided into three categories^[77]:

- ***Self-generated material*** (a teaching philosophy statement; representative syllabi, instructional objectives, handouts, assignments, and tests; descriptions of educational innovations and evaluations of their effectiveness; textbooks and education-related papers published; instructional software developed; teaching workshops and seminars presented or attended).
- ***Teaching products*** (graded assignments, tests, and reports; scores on standardized tests; student publications or presentations on course-related work).

- **Information generated by others** (summaries of student, alumni, and peer evaluations; honors and awards; reference letters). Some items may be mandated, others may be included at the professor's option.

Portfolios can provide the framework for a systematic program of reflective analysis and peer collaboration leading to improvement of teaching and student learning. In fact, most researchers discuss portfolios in the context of formative assessment and recommend customizing them to emphasize the strengths and objectives of the individual faculty member^[76-80]. Teaching portfolios can also be used to present evidence about teaching quality for summative purposes. A summative teaching portfolio may be assembled to evaluate the teaching effectiveness of an individual faculty member. When the portfolio is used as part of the basis for personnel decisions, it should be independently reviewed by at least two raters who have been trained in portfolio evaluation^[69]. Moreover, it should not be an exhaustive compilation of all documents and materials pertaining to teaching performance. Rather, it should present selected information on teaching activities and provide evidence of their effectiveness. A recommended format for a summative portfolio consists of several parts, including^[69]:

- Preamble
- Reflective statement of teaching philosophy, goals, and practices
- Summary of teaching and advising responsibilities
- Representative instructional materials and student products
- Evidence of teaching effectiveness
- Efforts to improve teaching effectiveness
- Teaching innovations
- Evidence of effectiveness of advising and mentoring
- Awards and recognitions

VI. Summary

Teaching is a multifaceted activity that is best evaluated using multiple measurement techniques and criteria. In general, there are six key steps in the development of a highly reproducible instructional evaluation system:

1. Determine the purpose of the evaluation;
2. Define the aspects/dimensions of teaching to be evaluated;
3. Identify valid sources of data or evidence for each aspect of teaching being evaluated;
4. Specify the criteria, or measuring instrument, by which the aspects will be judged;
5. Analysis and interpretation of data by skilled, trained personnel; and
6. Set weights, or scoring mechanism, for each aspect of teaching being evaluated.

This paper examined the first five steps. The sources, types, reliability and validity of data used for summative evaluations, including peer review, student ratings, and self-assessment, were examined. The peer review process involves classroom observations and course material evaluations of an instructor by a well-trained colleague. The classroom observation process is a 3-step procedure, including the pre-observation meetings, observation, post-observation meetings. Evaluation of course materials is best conducted when the materials are organized in a teaching portfolio. Voluminous and rigorous studies on student ratings have shown that this source of data is statistically reliable, valid, and relatively free of bias. Even though studies show that student ratings are nearly free of bias, sources of bias should be controlled by using

appropriate comparative data. Teaching portfolios are evaluated as a means for faculty self-assessment of teaching effectiveness. A format for a summative portfolio is composed of 9 parts, which can be divided into three categories: self-generated material, teaching products, and information generated by colleagues.

Acknowledgements

This work was partially supported with funds provided under National Science Foundation (NSF) grant HRD-0441207 which supported Dr. Baber as a Christine Mirzayan Science and Technology Policy Graduate Fellow at the National Academy of Engineering. The views expressed are those of the authors and do not necessarily reflect those of the NSF. The Committee on Engineering Education of the National Academy of Engineering is engaged in a formal study of the evaluation of teaching, and this paper provided informal background to that work but did not involve the study committee members and is not indicative of the findings that will emerge from that effort.

Bibliography

1. Boyer, Ernest L. *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching, 1990.
2. National Research Council. *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics*, National Academy Press, Washington, DC, 2002.
3. Shulman, Lee. "Teaching As Community Property: Putting an End to Pedagogical Solitude." *Change*, vol. 25, no. 6, (1993): pages 6-7.
4. Streveler, Ruth A., Maura Borrego, and Karl A. Smith. "Moving From the "Scholarship of Teaching and Learning" to "Educational Research." An Example From Engineering." *To Improve the Academy*. Bolton, MA: Anker Publishing Company, Inc., In Press.
5. Wankat, Phillip C., Richard C. Felder, Karl A. Smith, and Frank S. Oreovicz. "The Scholarship of Teaching and Learning in Engineering." *Disciplinary Styles in the Scholarship of Teaching and Learning: Exploring Common Ground*. Mary Taylor Huber, and Sherwyn P. Moreale Washington, DC: American Association for Higher Education and Carnegie Foundation for the Advancement of Teaching, 2002.
6. Accreditation Board for Engineering and Technology. *Vision for Change: a Summary Report of the ABET/NSF/Industry Workshops*, 95-VIS, Accreditation Board for Engineering and Technology, Baltimore, MD, 1995.
7. National Research Council. *Scientific Research in Education*, National Academy Press, Washington, DC, 2002.
8. ———. *How People Learn: Bridging Research and Practice*, editors, M. Suzanne Donovan, John D. Bransford, and James W. Pellegrino, National Academy Press, 1999.
9. Lohmann, Jack. "Building a Community of Scholars: The Role of the *Journal of Engineering Education* as a Research Journal." *Journal of Engineering Education*, vol. 94, no. 1, (2005): pages 1-4.
10. ABET, Inc. *2004-2005 Criteria for Accrediting Engineering Programs*, Baltimore, MD: Accreditation Board for Engineering and Technology, Inc., 2003. Available at <http://www.abet.org/images/Criteria/E001%2004-05%20EAC%20Criteria%2011-20-03.pdf>
11. American Society for Engineering Education. "Advancing Scholarship Engineering Education: Launching a Year of Dialogue." Web page, February 2005 [accessed 9 May 2006].
12. England, J., P. Hutchings, and W.J. McKeachie. *The Professional Evaluation of Teaching* ACLS Occasional Paper No. 33 [cited Jan. 2008]; Available from: <http://www.acls.org/op33.htm>.

13. Milne, D. *Stepping Stones to Successful Peer Observation to Enhance Teaching & Learning*. 2005 [cited Jan. 2008]; Available from: <http://www.tedi.uq.edu.au/downloads/evaluations/StepStonesPeerObservation.pdf>.
14. Stodolsky, S.S., *Teacher Evaluation: The Limits of Looking*. Educational Researcher, 1984. **13**(9): p. 11-18.
15. Berman, E., *A Short Guide to Evaluating Teaching*. 2000, Assessment and Enrollment Research (University of Arizona).
16. McLaughlin, M.W. and R.S. Pfeifer, eds. *Teacher Evaluation: Improvement, Accountability, and Effective Learning* 1988, Teachers College Press: New York.
17. Wise, A.E., et al., *Case Studies for Teacher Evaluation: A Study of Effective Practices*. 1984, Santa Monica: The Rand Corporation.
18. Dwyer, C.A., *Psychometrics of Praxis III: Classroom Performance Assessments*. Journal of Personnel Evaluation in Education, 1998. **12**(2): p. 163-187
19. Shinkfield, A.J. and D.L. Stufflebeam, *Teacher Evaluation: Guide to Effective Practice*. 1995, Boston: Kluwer Academic Publishers.
20. *Preparing for Peer Observation: A Guidebook* [cited Jan. 2008]; Available from: <http://www.utexas.edu/academic/cte/PeerObserve.html>.
21. *NC State Peer Review Web Site: Observation for Summative (Evaluative) Purposes*. [cited Jan. 2008]; Available from: http://www.ncsu.edu/provost/peer_review/summativeobserve.htm.
22. *Peer Observation Guidelines and Recommendations*. [cited Jan. 2008]; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/guidelines/index.html>.
23. *Peer Review of Teaching*. [cited Jan. 2008]; Available from: <http://www.unisanet.unisa.edu.au/learningconnection/staff/practice/evaluationpeerreview-teaching.asp>.
24. *Classroom Observation Instruments: Classroom Observation Form - Open Ended* [cited Jan. 2008]; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments.html>.
25. *Classroom Observation Instruments: Classroom Observation Form - Scale*. [cited Jan. 2008]; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments.html>.
26. *Classroom Observation Instruments: Classroom Observation Worksheet*. [cited Jan. 2008]; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments.html>.
27. *Classroom Observation Instruments: A Made-to-Order Form for Instructional Observation*. [cited Jan. 2008]; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments.html>.
28. *Classroom Observation Instruments: Classroom Observation Report 2*. [cited; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments.html>.
29. *Classroom Observation Instruments: Faculty Observation Scale* [cited; Available from: <http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments.html>.
30. Minnesota, C.f.T.a.L.U.o. *Classroom Peer Observation* [cited Jan. 2008]; Available from: <http://www1.umn.edu/ohr/img/assets/17085/checklist.pdf>.
31. Milne, D. *Peer Observation of Teaching and Learning: Matrices to Identify and Support Effective Teaching and Learning at University*. 2005 [cited Jan. 2008]; Available from: <http://www.tedi.uq.edu.au/downloads/evaluations/PeerObservationTandL.pdf>.
32. *Peer Reviews*. [cited Jan. 2008]; Available from: http://www.sbm.temple.edu/foxteach/peer_reviews.html.
33. *Guidelines for Evaluating Teaching Performance of Candidates for Promotion and Tenure* [cited Jan. 2008]; Available from: http://oaa.osu.edu/eval_teaching/agecon.html.
34. *Guide to Peer Review of Teaching* [cited Jan. 2008]; Available from: http://www.utas.edu.au/tl/improving/peerreview/guide_peer_review.doc.
35. Felder, R.M., *It Takes One To Know One*. Chemical Engineering Education, 1997. **31**(1): p. 32-33.
36. Anderson, H.M., *A Study of Certain Criteria of Teaching Effectiveness*. Journal of Experimental Education, 1954. **23**(1): p. 41-71.
37. Seldin, P., *Changing Practices in Faculty Evaluation: A Critical Assessment and Recommendations for Improvement*. 1984, San Francisco: Jossey-Bass Publishers.
38. *Peer Review of Teaching: Observing Teaching*. [cited Jan. 2008; Available from: <http://teachingacademy.wisc.edu/archive/Assistance/MOO/observes.htm>.
39. (1994) *Peer Observation of Classroom Teaching*. For Your Consideration... **Volume**,
40. Wilde, J. *Assessment Strategies for Professional Development Activities*. 1996 [cited Mar. 2007]; Available from: <http://www.ncela.gwu.edu/pubs/eacwest/profdev/>.
41. Braskamp, L.A. and J.C. Ory, *Assessing Faculty Work : Enhancing Individual and Institutional Performance*. 1994, San Francisco: Jossey-Bass.

42. *Peer Review of Teaching: How do I conduct a review of a colleague's teaching?*. [cited Jan. 2008]; Available from: <http://teachingacademy.wisc.edu/archive/Assistance/MOO/reviewer.htm>.
43. CIDR, *Classroom Observation*. CIDR Teaching and Learning Bulletin 1998. **1**(4).
44. *Peer Review of Teaching: Characteristics of Reflective Feedback*. [cited Jan. 2008]; Available from: <http://teachingacademy.wisc.edu/archive/Assistance/MOO/listfb.htm>.
45. *Guidelines for Peer Review of Teaching*. [cited Jan. 2008]; Available from: <http://www.gmu.edu/facstaff/part-time/app-e.html>.
46. *Peer Review of Teaching: Summative Review of Course Materials*. [cited Jan. 2008]; Available from: http://www.ncsu.edu/provost/peer_review/summaterials.htm.
47. *Guidelines for Evaluating Teaching*. [cited Jan. 2008]; Available from: <http://www.crlt.umich.edu/tstrategies/guidelines.html>.
48. Cohen, P.A. and W. McKeachie, *The Role of Colleagues in the Evaluation of College Teaching*. *Improving College and University Teaching*, 1980. **28**(4): p. 147-154.
49. *Peer Review of Teaching: Summative Procedures*. [cited Jan. 2008]; Available from: http://www.ncsu.edu/provost/peer_review/sumprocedures.html.
50. *Peer Review of Teaching: Summary of Best Practices*. [cited Jan. 2008]; Available from: http://www.ncsu.edu/provost/peer_review/bestpractice.html.
51. *The Teaching Assessment and Evaluation Guide*. [cited Jan. 2008]; Available from: <http://www.yorku.ca/secretariat/senate/committees/scotl/tevguide.pdf>.
52. Sullivan, T.A., *Teaching Evaluation by Peers*. *Teaching Sociology*, 1995. **23**(1): p. 61-63.
53. Cashin, W.E., *Student Ratings of Teaching: Uses and Misuses, in Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*, P. Seldin, Editor. 1999, Anker Publishing Co.: Bolton.
54. Seldin, P., *Current Practices-Good and Bad-Nationally, in Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*, P. Seldin, Editor. 1999, Anker Publishing Co.: Bolton.
55. Felder, R.M., *What Do They Know Anyway?* *Chemical Engineering Education*, 1992. **26**(3): p. 134-135.
56. Cashin, W.E., *Student Ratings of Teaching: The Research Revisited*. IDEA Paper 1995(No. 32).
57. McKeachie, W.J., *Student Ratings: The Validity of Use*. *American Psychologist*, 1997. **52**(11): p. 1218-1225.
58. Marsh, H.W. and L.A. Roche, *Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility*. *American Psychologist*, 1997. **52**(11): p. 1187-1197.
59. Franklin, J. and M. Theall, *Who Reads Ratings: Knowledge, Attitude, and Practice of Users of Student Ratings of Instruction*, in *73rd Annual Meeting of the American Educational Research Association*. 1989: San Francisco, CA.
60. Marsh, H.W., *Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility*. *Journal of Educational Psychology* 1984. **76**(5): p. 707-754.
61. Centra, J.A., *A Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. 1993, San Francisco: Jossey-Bass.
62. Theall, M. and J. Franklin, *Looking for Bias in All the Wrong Places: A Search for Truth or a Witch Hunt in Student Ratings of Instruction?*, in *New Directions for Institutional Research*, M. Theall, P.C. Abrami, and L.A. Mets, Editors. 2001, Jossey-Bass: San Francisco. p. 45-56.
63. French, G.M., *College Students' Concept of Effective Teaching Determined by an Analysis of Teaching Ratings*. *Dissertation Abstracts*, 1957. **17**: p. 1380-1381.
64. Crawford, P.L. and H.L. Bradshaw, *Perception of Characteristics of Effective University Teachers: A Scaling Analysis*. *Educational and Psychological Measurement*, 1968. **28**: p. 1079-1085.
65. Cohen, P.A., *Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies*. *Review of Educational Research*, 1981. **51**(3): p. 281-309.
66. Feldman, K.A., *The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies Research in Higher Education*, 1989b. **30**(6): p. 583-645.
67. Cashin, W.E., *Defining and Evaluating College Teaching*. IDEA Paper, 1989(No. 21).
68. Keig, L. and M.D. Waggoner, *Collaborative Peer Review: The Role of Faculty in Improving College Teaching*. 1995, School of Education and Human Development (George Washington University): Washington DC.

69. Felder, R.M., *The Future of Engineering Education V. Assessing Teaching Effectiveness and Educational Scholarship*. Chemical Engineering Education, 2000. **34**(3): p. 198-207.
70. Lewis, K.G., *Gathering Data for the Improvement of Teaching*, in *New Directions for Teaching and Learning*, M. Theall and J. Franklin, Editors. 1994, Jossey-Bass: San Francisco. p. 65-81.
71. Cashin, W.E., *Student Ratings of Teaching: A Summary of the Research*. IDEA Paper, 1988(No. 20).
72. Jacobs, L.C. *Student Ratings of College Teaching: What Research Has to Say*. [cited April 2007]; Available from: <http://www.indiana.edu/~best/multiop/ratings.shtml>.
73. Cashin, W.E., *Student Ratings of Teaching: Recommendations for Use*. IDEA Paper, 1990(No. 22).
74. *Student Evaluation of Teaching 1994*, Center for Teaching and Learning (University of North Carolina at Chapel Hill).
75. Seldin, P., *The Use and Abuse of Student Ratings of Professors*. Chronicle of Higher Education, 1993. **39**(46): p. A40.
76. Seldin, P., *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*. 2nd ed. 1997, Bolton: Anker Publishing Company.
77. Felder, R.M., *If You've Got It, Flaunt It: Uses and Abuses of Teaching Portfolios*. Chemical Engineering Education, 1996. **30**(3): p. 188-189.
78. Seldin, P., *Successful Use of Teaching Portfolios*. 1993, Bolton: Anker Publishing Co.
79. Woods, D.R., *Ideas for Teachers to Improve Learning*. 3rd ed. 1999, Hamilton: McMaster University.
80. Shore, B.M., et al., *Teaching Dossier: A Guide to Its Preparation and Use*. 2nd ed. 1986, Ottawa: Canadian Association of University Teachers.
81. Diamond, R.M., *Preparing for Promotion and Tenure Review: A Faculty Guide*. 1995, Bolton: Anker Publishing Co.