THE FUTURE OF ENGINEERING EDUCATION

Part 5. Assessing Teaching Effectiveness and Educational Scholarship

Richard M. Felder • *North Carolina State University* • *Raleigh, NC 27695-7905*
Armando Rugarcia • *Iberoamericana University* • *Puebla, Mexico*
James E. Stice • *University of Texas* • *Austin, TX 78712-1062*

The first four papers in this series[1-4] offered a number of ideas for effective teaching and preparing faculty members to teach. An inevitable question is, how does one determine whether or not a faculty member's teaching is effective? Another important question is, how does one determine whether or not an instructional *program*—such as that of an engineering department—is effective?

The instructional component of the mission of every educational institution is to produce graduates with satisfactory levels of knowledge, skills, and attitudes.[1] The specific knowledge, skills, and attitudes may differ from one department to another and the definition of satisfactory may differ from one institution to another, but the instructional mission is invariant. In engineering, the basis of a department's accreditation is the extent to which the department is fulfilling this mission. An instructor may be a brilliant lecturer with student ratings at the top of the charts, but if his or her teaching is not furthering the instructional mission of the department, that teaching cannot be considered effective.

To appraise programmatic teaching effectiveness, we must answer the following questions:[5,6]

1. ***Educational goals***. *What are the published goals of the instructional program? Does the faculty know what they are? Does the faculty generally agree with them?*
2. ***Performance criteria***. *Are the criteria that will be used to evaluate faculty performance measurable and clearly tied to the goals? Does the faculty know what they are? Does the faculty generally agree with them?*
3. ***Assessment process***. *What assessment data will be collected? How and when and by whom will they be collected and analyzed? Are available resources (including faculty time) adequate to permit their collection and analysis?*
4. ***Evaluation process***. *How will conclusions about teaching effectiveness be inferred from the data, and by whom? What type of feedback will be provided to the faculty, and when and by whom will it be provided?*

The answers to these questions should be based on the university mission statement and program accreditation requirements, with additional criteria and procedures contributed by the program administration and faculty.

An additional factor enters into the appraisal of an individual faculty member's teaching performance—namely, the extent to which he or she is contributing to the improvement of education. We refer to this performance factor as *educational scholarship*. It encompasses developing or systematically improving teaching methods and methods of assessing learning outcomes, writing textbooks and courseware, and publishing scholarly papers and monographs and giving work-

Richard M. Felder *is Hoechst Celanese Professor (Emeritus) of Chemical Engineering at North Carolina State University. He received his BChE from City College of New York and his PhD from Princeton. He has presented courses on chemical engineering principles, reactor design, process optimization, and effective teaching to various American and foreign industries and institutions. He is coauthor of the text* Elementary Principles of Chemical Processes *(Wiley, 2000).*
Armando Rugarcia *graduated from the Universidad Iberoamericana (UIA) in 1970 and went on to earn his MS in chemical engineering from the University of Wisconsin in 1973 and his Doctorate in Education from West Virginia University in 1985. He has been a full-time professor of engineering at UIA since 1974 and was Chair of the Chemical Engineering Department there from 1975 to 1980. He was also Director of the Center for Teaching Effectiveness at UIA from 1980 until 1986. He has written four books on education, one on process engineering, and more than 130 articles.*
James Stice *is Bob R. Dorsey Professor of Engineering (Emeritus) at the University of Texas at Austin. He received his BS degree from the University of Arkansas and his MS and PhD degrees from Illinois Institute of Technology, all in chemical engineering. He has taught chemical engineering for 44 years at the University of Arkansas, Illinois Tech, the University of Texas, and the University of Wyoming. At UT he was Director of the Bureau of Engineering Teaching Center and initiated the campus-wide Center for Teaching Effectiveness, which he directed for 16 years.*

shops and seminars on education-related topics. For individual faculty performance evaluation (as opposed to instructional program evaluation), the questions listed above should therefore be augmented by these:

5. ***Educational scholarship***. *What evidence of scholarly contributions to education will be collected? How and by whom will the evidence be evaluated?*

In this paper we suggest options for answering most of these questions. We first propose principles of instructional assessment and summarize common violations of these principles. Then we elaborate on how to assess the effectiveness of both teaching and educational scholarship, leaving the evaluation process (determining what qualifies as satisfactory performance) to be determined by institutional norms and values.

In this paper we . . .propose principles of instructional assessment and summarize common violations of these principles. Then we elaborate on how to assess the effectiveness of both teaching and educational scholarship, leaving the evaluation process . . . to be determined by institutional norms and values.

SEMANTIC NOTES

In the educational literature, the two terms *assessment* and *evaluation* are constantly encountered. They are sometimes used interchangeably as synonyms for appraisal of instructional effectiveness; sometimes assessment denotes the appraisal of individual teaching and evaluation the appraisal of teaching programs;[5,6] and sometimes assessment denotes collecting and analyzing data that reflect on teaching quality and evaluation denotes interpreting the assessment outcomes and drawing conclusions about teaching quality.[7] Unless otherwise noted, we will use the latter definitions in our discussions.

An important distinction is that between *formative assessment,* which has improvement of teaching as its objective, and *summative assessment*, which produces information that can be used to make decisions about instructional personnel or programs. Formative assessment is (or should be) an important part of institutional programs to help faculty members become more effective as teachers, a topic discussed in the preceding paper in this series.[4] This paper concerns summative assessment.

CRITERIA FOR EFFECTIVE COURSE INSTRUCTION

Evaluation of either programmatic teaching effectiveness or individual faculty member performance involves assessing the quality of instruction in individual courses. Extensive research supports the use of the following criteria as a basis for the assessment:[8-15]

1. The course contributes toward published program goals.
2. The course has clearly stated measurable learning objectives.[2]
3. The assignments and tests are tied to the learning objectives and are fair, valid, and reliable.[2]
4. Appropriate methods have been devised to monitor the effectiveness of the instruction.
5. The learning environment is appropriate.[2,3]
6. The instructor has appropriate expertise in the course subject.
7. The instructor communicates high expectations of students and a belief that they can meet those expectations, interacts extensively with them inside and outside class, conveys a strong desire for them to learn and motivates them to do so.
8. The instructor seeks to provide an education in the broadest sense of the word, not just knowledge of technical content.[1]
9. The instructor integrates teaching with research.
10. The instructor continually attempts to improve the course by updating the content and/or making use of new instructional materials and methods (including applications of instructional technology).
11. The students achieve the learning objectives.

More details are given by Woods.[15]

ASSESSMENT AND EVALUATION OF TEACHING EFFECTIVENESS

An assessment plan should involve assembling several types of evidence to determine the degree to which the foregoing criteria are being met. Among the possibilities are the following:

- *Learning outcomes assessments: student performance on standardized tests, comparisons of student performance with performance of control groups, evaluations of student products by external reviewers.*
- *Student end-of-course ratings.*
- *Student surveys, focus groups, or interviews directed at specified criteria.*
- *Retrospective student ratings of courses and instructors (e.g., pre-graduation ratings by seniors).*
- *Alumni ratings of courses and instructors.*
- *Peer ratings of classroom instruction, learning objectives, assignments and tests.*
- *Evaluations submitted by external referees.*
- *Self-evaluations by instructors.*

The assessment data may be collected for individual faculty members in *teaching portfolios* (or *teaching dossiers*), which may be evaluated by a review team to provide an effective assessment of instructional effectiveness. The portfolios assembled by all members of a department collectively provide a partial basis for evaluating the effectiveness of the department's instructional program. More would have to be done to demonstrate that the program graduates meet specified criteria related to their knowledge, skills, and attitudes (such as those specified as Outcomes 3a–3k of ABET Engineering Criteria 2000).

Assessment of Learning

The ultimate assessment of teaching is assessment of learning. Teaching that does not satisfy institutional, departmental, and individual instructors' learning objectives cannot be considered effective, regardless of what other assessment measures may indicate.

The past decade has seen a growing realization that the traditional assessment tool used in undergraduate engineering education for most of the past century—the written examination on material covered in lectures and readings—provides an inadequate measure of the knowledge, skills, and attitudes that engineering schools wish to impart to their students. Driven in large part by the impending adoption of Engineering Criteria 2000 as the accreditation system for all U.S. engineering departments, a large and constantly growing body of work on the systematic assessment of specified learning outcomes has arisen. A full review of this literature is well beyond the scope of this paper; what follows is a brief summary of the principal ideas.

Assessment—whether of learning or teaching, whether for individual courses or entire instructional programs—can only be done meaningfully in the light of clearly stated goals and measurable objectives. In the case of assessment of learning, the requirements are explicit statements of the knowledge, skills, and attitudes that the students are supposed to acquire (the goals) and of what the students must do to demonstrate that the goals have been met (the objectives). The following assessment tools may be used as part of that demonstration. The terms in parentheses indicate the categories of objectives that the specified tools may be used to assess, including outcomes specified by Engineering Criteria 2000.

❒ **Complete tests and individual test items** (*knowledge, conceptual understanding, engineering problem-solving skills*). Tests given in engineering courses may provide good measures of relative learning among students in a particular class, but they are frequently unsuitable for assessment of true conceptual understanding and problem-solving skills. They may also provide misleading results. For example, if tests are too long for most students to finish (a situation that unfortunately characterizes many engineering tests), students who work sloppily but quickly may earn much higher grades than students who work accurately but slowly. The most meaningful assessment is provided when the test results may be compared with established norms or with results from comparison groups, such as another class taught in parallel to the one in question by a different instructor and/or using a different instructional method. The nationally normed Fundamentals of Engineering (FE) examination has the potential to provide a basis for assessment.[16]

❒ **Laboratory reports, design project reports, live or videotaped oral presentations, research proposals** (*knowledge, conceptual understanding, analysis, creative thinking, critical thinking, experimental design, identification of engineering problems, teamwork, written and oral communication skills, professional or social awareness, lifelong learning skills*). The usual drawback of reports as assessment instruments is subjectivity in their evaluation. One way to improve their effectiveness is to use detailed checklists in evaluating the reports, tying the checklist items to specific learning objectives. Even greater assessment validity is provided by using several independent raters who reconcile their ratings after completing their checklists.

❒ **Resumes, letters, memos** (*written communication skills, professional or ethical awareness*). An effective way to prepare students to function as professionals is to ask them to engage in common professional activities and provide them with feedback on their efforts. For example, periodically ask engineering students to prepare resumes and to write letters and memos dealing with common hypothetical

situations, such as reporting a result to a supervisor, asking for an interview with a prospective employer, persuading a client or a prospective client to purchase a product or service, or recommending an action to a superior or a subordinate in a situation that has ethical implications.

❒ **Critiques of technical reports, papers, letters, and memos** (*analysis, critical thinking, written communication skills*). It is often easier to see weaknesses in someone else's work than in one's own. Having students critique one another's first drafts of written documents and revise their own documents based on the feedback they get helps them develop critical thinking skills, especially if the critiques are collected and graded. The papers handed in to the instructor are generally much better than they would have been without the preliminary feedback, and the grading job of the instructor is consequently much less burdensome.

❒ **Self-evaluations, learning logs, journals** (*any skills or attitudes*). Surveying or interviewing students is a direct way to obtain their impressions of how much their skills have improved as a consequence of their education. The validity of such data is greatest if the data are consistent with results obtained by other means, or if the same data are available for comparison groups subjected to different forms of instruction. Student learning logs or journals can be rich indicators of the degree of acquisition of selected skills and attitudes, but trained evaluators are needed to make such inferences and the process can be extremely time- and labor-intensive.

❒ **Other classroom assessment techniques** (*any skills or attitudes*). The classic reference on classroom research by Angelo and Cross[17] suggests a large variety of techniques for assessing knowledge, recall, understanding, and ability to apply learned information; skills in analysis and critical thinking, synthesis and creative thinking, and problem solving; and self-awareness as learners. While the usual applications of these techniques are formative, any of them may also be used for summative assessment.

A comprehensive picture of student learning is provided by assembling *student portfolios*—longitudinal records of student learning assessment results. Panitz[18] describes uses of portfolios for both formative and summative purposes at different schools. Some instructors allow students to determine how much weight should be assigned to different course components, assemble the portfolios themselves, indicate the grade they think they have earned, and write a statement indicating how the portfolio contents justify the grade. Others set up competency matrices of one type or another. One format consists of rows for different student products in the portfolio and columns for specific learning outcomes or objectives, with marks to show which products demonstrate which outcomes or the levels (A, B, C,...) at which the objectives are satisfied. Rogers and Williams[19] describe a web-based portfolio system created at the Rose-Hulman Institute of Technology. Students enter work that they believe demonstrates their progress toward meeting specific performance criteria and state justifications for their claims, and faculty raters evaluate the entries.

Student Ratings of Instruction

The most common method—and in many programs, the only method—of assessing instructional quality is to collect student ratings at the end of each course. The rating form is often haphazardly designed, and the results may be difficult to interpret with any degree of objectivity. In part because of these defects, many faculty members discount the validity and value of student ratings. Commonly heard criticisms are that ratings do not correlate with quality of learning and the easiest teachers get the highest ratings.

In fact, more than a thousand research studies of student ratings have been performed, and the results collectively show that ratings are reliable, stable over time, and positively correlated with results obtained using other forms of teaching assessment, including assessment of learning outcomes.[20-23] Contrary to popular assertions, they are not affected appreciably by the instructor's personality or gender or the time of day a class is offered.[21] Difficult courses that do not require unreasonable expenditures of time and effort are rated somewhat more favorably than courses that lack challenge.[22] Some studies show positive correlations between ratings and grades, but it is not clear whether the higher grades in the more highly rated courses reflect inappropriately easy grading or superior learning. The positive correlations observed between ratings and learning outcomes suggest that the latter may be a strong contributing factor.

Their validity notwithstanding, student ratings should not be the only method used to assess instructional quality. There are several important aspects of teaching that students lack the knowledge and perspective to judge fairly, including the currency and importance of the course content, the instructor's understanding of the subject, and the appropriateness of the assignments, tests, and grading policies.[22-24] Many institutions use non-standardized assessment instruments and fail to take into account extraneous factors such as class size, course level, and whether courses are required or elective, making the results for different faculty members difficult or impossible to compare.

Nevertheless, course-end student ratings are an essential component of instructional quality assessment. As long as they are to be collected, certain steps should be taken to maximize their effectiveness.[25]

❒ **Collect ratings of the effectiveness of the course and the instructor in a few critical aspects.** The most commonly used format is probably the five-point Likert scale (e.g., 1=strongly disagree, 2=disagree, 3=neutral, 4=agree,

5=strongly agree) applied to items related to the quality of teaching and learning in the course. The following items have been shown to be related to teacher effectiveness as measured by mean student performance on examinations:[26]

- *Each class period was carefully planned in advance.*
- *The instructor presented the material clearly.*
- *The professor made students feel free to ask questions, disagree, express their ideas, etc.*
- *The professor used examples from his/her own research or experience.*
- *This course has increased my knowledge and competence.*

Other questions might be asked related to acquisition of specific skills included in the course goals (*e.g.,* critical or creative thinking, writing, teamwork, etc.). Since a standardized form is desirable for summative assessment, however, the items chosen should be small in number and general enough to apply to different courses and instructors within a single discipline and across disciplines. (For formative assessment, items may be included on any aspect of the instruction on which the instructor wishes feedback.) The form should not contain questions about things the students are not equipped to evaluate, such as the instructor's knowledge of the subject.

❒ **Collect overall course-end ratings of instruction**. "Rate the instruction you received in this course on a scale from 1 to 5, with 1 being the highest response." Ratings of this sort are most effective when the numbers on the response scale are clearly defined. Definitions like "excellent," "above average," "fair," etc., are subjective and ambiguous, and when they are used a very broad performance range tends to be lumped into "above average." Greater discrimination is obtained by giving descriptions of the characteristics of instructors in each category, making it clear that very few instructors are likely to fall into the extreme categories.

One approach is to use a norm-referenced system, wherein 5 means that the instructor is one of the three best teachers the student has ever had (or is in the top 1% or the top 5%), 1 signifies one of the three worst teachers (or the bottom 1% or 5%), and 2, 3, and 4 represent different percentile ranges (*e.g.,* bottom 20%, middle 60%, and top 20%). The problem with this system is that it penalizes faculty members in departments with a large number of excellent instructors. A better approach calls on students to base their overall rating on the average of their ratings of individual characteristics of the course and instructor (previous bullet). For example, the students could be asked to total their ratings of the individual items, and ranges for the total could be given on the form that translate to overall ratings of 5, 4, 3, 2, and 1. The ranges corresponding to the highest and lowest overall ratings should be relatively narrow (*e.g.,* a total that would yield an average rating in the range 4.75 to 5.0 might correspond to an overall rating of 5; 3.75 - 4.75 to a rating of 4; 2.25 - 3.75 to a 3; 1.25 - 2.25 to a 2; and 1 - 1.25 to a 1). If this system were used, instructors who get 5 would clearly be worthy of nomination for an outstanding teacher award and instructors who get 1 would clearly have very serious problems with their teaching.

❒ **Administer and collect course-end ratings in a single class session rather than counting on students to return them later**. Results of evaluations for which the return rate is less than a minimal percentage should be regarded with deep suspicion: the recommended minimum is 50% (classes of 100 or more), 66% (50-100), 75% (20-50), and 80% (<20).[27] The environment used for gathering the data should include student anonymity and absence of the instructor from the room.

❒ **Interpret ratings collected over a period of at least two years**. One semester of low ratings (or high ratings, for that matter) does not provide a valid measure of an instructor's teaching effectiveness.

❒ **Periodically collect retrospective student evaluations in addition to course-end ratings.** Ratings from seniors and alumni of how well individual instructors helped them acquire knowledge and develop skills are powerful indicators of teaching effectiveness. These retrospective ratings help identify the relatively small percentage of instructors whose students only appreciate their effectiveness as teachers years after taking their courses. For faculty members at research universities, ratings from former research advisees attesting to the degree to which professors promoted their intellectual curiosity and research skills should also be sought.

Peer Ratings

Peer ratings can contribute significantly to the evaluation of teaching if they are well designed and conducted, but the common practice of having untrained faculty members sit in on a lecture and make notes on whatever happens to catch their attention yields results that are neither reliable nor valid.[28] To be effective, summative peer ratings should include the features described below.[29,30]

❒ **Who should do the reviewing?** Reviewers should be good teachers who have received training on what to look for in a classroom and who recognize that different styles of teaching can be equally effective. Training dramatically increases the likelihood that evaluations from different reviewers will be consistent with one another (reliability) and with accepted standards for good teaching (validity).

❒ **How should classroom observations be performed?** At least two reviewers should conduct at least two class visits during a semester, preceding each visit with a brief meeting at which the instructor provides information about the class to be observed. The reviewers independently complete stan-

dardized rating checklists after each observation and soon afterwards visit with the instructor to discuss their observations and invite responses. After all individual observations and reviews have been completed, the reviewers compare and reconcile their checklists to the greatest extent possible and write a summary report to be placed in the instructor's teaching portfolio or personnel file.

❒ **What should the lecture observation checklist contain?** The checklist is a collection of statements about the observed classroom instruction with which the reviewers indicate their levels of agreement or disagreement, adding explanatory comments where appropriate. Statements such as the following might be included:[31]

- **Organization**. The instructor (a) begins class on time, (b) reviews prior material, (c) previews the lecture content, (d) presents material in a logical sequence, (e) summarizes main points at the end of the period, (f) ends class on time.
- **Knowledge**. The instructor (a) has a good understanding of the course material, (b) integrates ideas from current research and engineering practice into the lectures, (c) answers questions clearly and accurately.
- **Presentation**. The instructor (a) speaks clearly, (b) holds the students' attention throughout the period, (c) highlights important points, (d) presents appropriate examples, (e) encourages questions, (f) seeks active student involvement beyond simple questioning, (g) *attains* active student involvement, (h) explains assignments clearly and thoroughly.
- **Rapport**. The instructor (a) listens carefully to student comments, questions, and answers and responds constructively, (b) checks periodically for students' understanding, (c) treats all students in a courteous and equitable manner.

Many other statements could be included, some of which might be particularly applicable to laboratory or clinic settings. Examples of validated observation instruments are given in a recent book edited by Seldin.[31]

❒ **How should instructional materials be rated?** Examination of instructional objectives, lecture notes, assignments, tests, and representative student products may provide a better picture of teaching effectiveness than classroom observation. Trained observers can judge whether (a) the objectives cover a suitable range of knowledge and skills, (b) the course content is sufficiently comprehensive and current, (c) the assignments and tests are appropriately rigorous, fair, and consistent with the stated objectives. As with classroom observation, the ratings should be done by two or more independent observers using a validated checklist and reconciled to arrive at a consensus rating.

The Teaching Portfolio

The *teaching portfolio* (or *teaching dossier*) is a device used for assessing the teaching effectiveness of an individual faculty member, as opposed to effectiveness of instruction in a single course or of an instructional program. The portfolio is a summary of teaching assessment data, including self-assessment. Most authors who discuss portfolios[32-38] do so in the context of formative assessment and recommend customizing the portfolio to fit the strengths and objectives of the individual faculty member. In keeping with the theme of this paper, we will confine our discussion to summative assessment, which requires using a standard format to provide evaluative consistency.

A recommended format for a summative portfolio consists of several parts:

❒ **Preamble**. Context of the portfolio, time period covered, and outline of the contents.

❒ **Reflective statement of teaching philosophy, goals, and practices**. The instructor's answers to such questions as: "What is my mission as a teacher?" "What skills and attitudes should I be helping my students develop?" "What methods am I using in and out of class to fulfill my mission and enable my students to develop the desired skills and attitudes?" "What am I doing to motivate and equip them to succeed, academically, professionally, and personally?"

❒ **Summary of teaching and advising responsibilities**. Titles, levels, contact hours, and class sizes for all courses taught over the past five years, annotated with brief comments about the way each course is taught. Number of students advised and comments about the nature of the advising. Comments should relate explicitly to the reflective statement and to published institutional and departmental goals.

❒ **Representative instructional materials and student products.** Illustrative assignment statements and tests with grade distributions. Copies of outstanding and typical graded assignments, tests, and project reports. Discussion of the materials in the context of the reflective statement.

❒ **Evidence of teaching effectiveness**. Results of student ratings in the context of average departmental ratings for the same courses over the past six years. Results of retrospective senior and alumni ratings and peer ratings. Results of learning assessments, including student performance on standardized tests. Data from instruments that assess approaches to and attitudes toward learning such as the Lancaster Approaches to Studying Questionnaire and the Course Perceptions Questionnaire[3,15,39] and the Perry or King/Kitchener Inventory.[3,6,15] Reference letters from students and alumni. Implications of the evidence in the context of the reflective statement.

❒ **Efforts to improve teaching effectiveness**. Steps taken to keep knowledge of course content and effective instructional methods up-to-date: workshops, seminars, and conferences attended, papers read, networking done. Steps taken to obtain student feedback and to monitor and improve the

learning environment and quality of classroom instruction.

❒ **Teaching innovations**. New courses developed and changes made to existing courses. New instructional materials generated, teaching strategies adopted, and methods used to motivate and empower students. Copies of publications or presentation abstracts describing innovations. Discussion of the innovations in the context of the reflective statement.

❒ **Evidence of effectiveness of advising and mentoring**. Successes of and recognition received by advisees. Reference letters from advisees. Implications of the evidence in the context of the reflective statement.

❒ **Awards and recognition**. Nominations for awards and awards received (include award criteria). Other recognition.

When the portfolio is used as part of the basis for personnel decisions (e.g. awarding of promotion or tenure or determining merit raises), it should be independently reviewed by at least two raters who have been trained in portfolio evaluation. Following a predetermined scheme, the raters should assign values to the quality of reflection and documentation, the instructor's commitment to high quality teaching and learning, and the instructor's teaching and advising effectiveness and (if appropriate) educational scholarship. The raters should compare and discuss their ratings, make any changes they believe to be appropriate, and arrive at a consensus rating. The individual and consensus ratings should be included in the portfolio to be used in the decision-making process.

Eventually, the department head must make a determination of teaching effectiveness based on his or her review of the assessment data. A form for guiding this review is available from the Kansas State University IDEA Center.[24]

ASSESSMENT AND EVALUATION OF EDUCATIONAL SCHOLARSHIP

In his landmark work *Scholarship Reconsidered*,[40] Ernest Boyer proposed that academics can pursue scholarly activities in four different arenas: *discovery* (advancement of the frontier of knowledge in a discipline), *integration* (making connections across disciplines, putting research discoveries in broader contexts and larger intellectual patterns), *application* (applying the outcomes of discovery and integration to socially consequential problems), and *teaching* (helping students acquire knowledge and develop skills). Boyer argued that these four areas are all equally vital to the mission of the research university and that universities should therefore recognize and reward them all equally.

The publication of *Scholarship Reconsidered* intensified an ongoing discussion about the role of teaching in the evaluation of faculty performance at research universities. Among the focal questions of the discussion are "What is educational scholarship?" and "How can you assess its quality?" The following discussion is taken largely from a recent article that addresses these questions.[41]

What is educational scholarship?

Boyer lists the elements that make teaching a scholarly activity:[40]

1. **Subject knowledge**. The scholarly instructor has a deep conceptual understanding and a broad awareness of the current state of knowledge of the subject being taught.
2. **Pedagogical knowledge**. The scholarly instructor can formulate analogies, metaphors, and images that build bridges between his or her understanding of the subject and the knowledge and level of experience of the students. The instructor is also familiar with a variety of effective instructional methods and the research base that confirms their effectiveness.
3. **Commitment to continuing growth as an educator**. The scholarly instructor is committed to continuous improvement of his or her disciplinary and pedagogical knowledge. Indications of such a commitment are books read, journals subscribed to, and seminars, workshops, and conferences attended.

A fourth element might be added to this list:[41]

4. **Involvement in development, assessment, and dissemination of innovative instructional methods and materials**.

Instructors who keep their subject knowledge current, learn about and implement effective teaching methods, and continue to work on improving their teaching may be said to be effective teachers, worthy of being nominated for whatever rewards the institution offers for teaching effectiveness, but they are not necessarily educational scholars. To qualify for that title, we propose that they must also undertake the activities associated with traditional disciplinary research: innovation and rigorous assessment and evaluation of the innovations. In educational scholarship as in disciplinary scholarship, the fruits of the labor might be products (e.g., textbooks or instructional software) or processes (e.g., new or improved methods for motivating students, promoting their intellectual development, or assessing their learning). Also as in disciplinary scholarship, making results available to the professional community for evaluation, replication, and adoption is a necessary component of educational scholarship.

The improving climate for educational scholarship

In the past, even if engineering professors were inclined to do scholarly work in education there were barriers to their doing so successfully. Grants for engineering education research were in short supply and provided minimal funding. Engineering education journals did not require rigorous assessment as a condition for publication, and journals in

education and educational psychology that did so were not receptive to contributions of an applied nature from other disciplines. Engineering administrators and faculty peers called on to evaluate faculty performance reports were unfamiliar with the education literature and generally discounted all education-related papers, including those that adhered to good assessment practices and were published in journals with high standards. Campus awards for outstanding scholarship in teaching did not exist.

The climate for scholarship in engineering education has become considerably warmer in recent years. The National Science Foundation has provided millions of dollars of funding through its Division of Undergraduate Education and the Engineering Education Coalition program, and corporate foundations have also provided significant support to efforts to improve engineering education. The *Journal of Engineering Education* has become a first-rate vehicle for scholarly publications, and other high-quality refereed journals now accept papers on engineering education research.[15] National, regional, and—on some campuses—local awards for outstanding scholarship in engineering education are given. Unfortunately, many who rate faculty performance in engineering are still inclined to discount education-related activities as not worthy of being counted toward promotion, tenure, and merit raises, funded and published though they may be. Hopefully, this situation will also improve before too long as more and more professors are motivated to undertake serious efforts to study and improve engineering education—rigorously setting goals, developing measurable outcomes, gathering data about the effectiveness of their interventions in the classroom, and subjecting the data to rigorous analysis and interpretation.

How can educational scholarship be assessed and evaluated?

Earlier in this paper, we proposed that for teaching to qualify as a scholarly activity, the instructor should demonstrate a command of both subject and pedagogical knowledge, a commitment to continuing growth as an educator, and an involvement in innovation in teaching and dissemination of results. We further propose that assessment of an instructor's educational scholarship should consist of answering the following three questions:[41]

1. *Did the teaching qualify as a scholarly activity?*
2. *Was the teaching effective?*
3. *Were the innovative products and processes developed by the instructor well conceived, implemented, assessed and evaluated, and disseminated?*

The data obtained using the assessment tools described in the preceding sections of this paper and summarized in the section on the teaching portfolio should be adequate to assess the first two questions. To answer the third question, the same forms of evidence traditionally used in the assessment of disciplinary research may be gathered. Acceptable evidence includes the number and quality of conference presentations, invited seminars, books, monographs, and refereed publications; number of grants and contracts; citations of publications; referee comments on submitted manuscripts and grant proposals; internal and external reference letters and comments, and recognition and awards.

The following standards proposed by Glassick, *et al.*,[42] provide a good basis for evaluating the quality of educational innovations:

1. ***Clear goals***. *Is the basis of the work clearly stated, the questions addressed important in the field, and the objectives realistic and achievable?*
2. ***Adequate preparation***. *Does the scholar show an understanding of existing scholarship in the field, the necessary skills to do the work, and the ability to assemble the necessary resources?*
3. ***Appropriate methods***. *Were the methods used appropriate to the goals, applied effectively, and appropriately modified when necessary?*
4. ***Significant results***. *Were the goals achieved? Did the work contribute significantly to the field? Did it open areas for further exploration?*
5. ***Effective presentation***. *Was the work presented effectively and with integrity in appropriate forums?*
6. ***Reflective critique***. *Does the scholar critically evaluate his or her own work, bringing an appropriate breadth of evidence to the critique and using the critique to improve the quality of future work?*

Faculty members who meet these standards are clearly vital to both the educational and scholarly missions of the university. They merit advancement up the faculty ladder—tenure, promotion, and merit raises—no less than faculty members who meet institutional standards for disciplinary research.

SUMMARY

The assessment of teaching should done for a clearly defined purpose—to evaluate teaching effectiveness (summative assessment) or to improve it (formative assessment). It should be done in the context of published goals, measurable performance criteria, and agreed-upon forms of evidence. The evidence should come from a variety of sources, including learning outcomes assessments, student end-of-course ratings, student surveys, focus groups, or interviews, retrospective student evaluations of courses and instructors, alumni and peer evaluations, and self-assessments.

The ultimate measure of the effectiveness of teaching is

the quality of the resulting learning. As with any other area of assessment, meaningful assessment of learning requires prior formulation of learning goals and measurable objectives that address all desired knowledge, skills, and attitudes. Tools for assessing learning include tests and test items, written reports and proposals, oral presentations and interviews, student-generated critiques of work produced by others, student self-evaluations, learning logs and journals. The validity of inferences drawn from the data is increased if norms or control group responses are available for objective tests and test items and if multiple independent evaluations are submitted and reconciled for subjective judgments such as ratings of written project reports and oral presentations.

Student ratings of teaching are a valid and important source of evidence for teaching effectiveness, especially if they are averaged over at least a two-year period. Extensive research shows that student ratings correlate positively with both learning outcomes and ratings submitted by alumni and peers. They should not be the sole instrument used to evaluate teaching, however, since students are generally not qualified to judge aspects of instruction like the currency and importance of the course content, the depth of the instructor's knowledge, and the appropriateness of the assignments, tests, and grading policies. Peer ratings are the most appropriate source of such judgments.

The common approach to peer rating is for untrained faculty members to observe lectures and write about whatever catches their attention, an approach that yields information of doubtful value. For peer ratings of instruction in a course to be reliable and valid, the ratings should be obtained from at least two good teachers who have received training on what to look for in a classroom. The raters should use a checklist of items regarding specific aspects of the instruction and associated instructional materials (syllabi, handouts, assignments, and tests), and the independent ratings should be reconciled to arrive at a consensus rating.

A summative teaching portfolio may be assembled to evaluate the teaching effectiveness of an individual faculty member (as opposed to the effectiveness of teaching in a single course or an instructional program). The portfolio should contain a reflective statement of the faculty member's teaching and advising philosophy, goals, and practices; a summary of teaching and advising responsibilities; representative instructional materials and student products; assessment data that reflect on teaching and advising effectiveness; documentation of efforts to improve effectiveness; a summary of teaching innovations (new courses, instructional materials, and teaching methods developed, and education-related papers and presentations); and a list of teaching awards and award nominations. When the portfolio is used as part of the basis for personnel decisions, at least two independent evaluations of the portfolio should be performed by trained raters and reconciled.

Since the publication of *Scholarship Reconsidered*,[40] recognition has been growing that teaching can be a scholarly activity no less than disciplinary research, and that scholarship in teaching should play the same role in determining faculty advancement that disciplinary research has played for the past four decades. Following Boyer, we propose that the defining elements of scholarly teaching are mastery of subject knowledge, familiarity with both general and subject-specific pedagogy, and commitment to continuing personal growth as an educator, and we propose the additional element of involvement in development, assessment, and dissemination of innovative instructional materials and methods. The innovations should reflect an awareness of the current state of the art of engineering education, and analysis and evaluation of the results should adhere to the same standards of rigor customarily applied to traditional disciplinary research.

Assessment of the quality of a faculty member's educational scholarship should be based on the answers to three questions: (1) Did the faculty member's teaching qualify as a scholarly activity? (2) Was his/her teaching effective? (3) Were his/her innovations well conceived, implemented, assessed and evaluated, and disseminated? The faculty member's subject knowledge, pedagogical knowledge, commitment to continuing personal growth, and involvement in innovation (the elements of scholarly teaching) and the effectiveness of the teaching can be judged from the material assembled in a teaching portfolio. The quality and impact of educational innovations can be inferred from the same forms of evidence used to evaluate disciplinary research (number and quality of books, papers, and presentations; literature citations; number of research grants and contracts; reference letters; and recognition and awards). Faculty members who meet or exceed institutional standards for educational research merit the same recognition and opportunities for advancement as faculty members who excel in disciplinary research.

ACKNOWLEDGMENTS

Thanks for valuable suggestions during the preparation of this paper to Chris Knapper of Queen's University, Kingston, and Paola Borin, Dale Roy, Erika Kustra, Barb Love, Heather Sheardown, and Phil Wood of McMaster University. Thanks also for helpful reviews of the paper to Brian Baetz of McMaster University, Suzanne Kresta of the University of Alberta, Rich Noble of the University of Colorado, and Dendy Sloan of the Colorado School of Mines. Special thanks go to Don Woods of McMaster, whose ideas had a profound influence on the content of this paper and the thinking of its authors.

REFERENCES

1. Rugarcia, A., R.M. Felder, D.R. Woods, and J.E. Stice, "The Future of Engineering Education: Part 1. A Vision for a New Century," *Chem. Eng. Ed.,* **34**(1), 16 (2000)
2. Felder, R.M., D.R. Woods, J.E. Stice, and A. Rugarcia, "The Future of Engineering Education: Part 2. Teaching Methods that Work," *Chem. Eng. Ed.,* **34**(1), 26 (2000)
3. Woods, D.R., R.M. Felder, A. Rugarcia, and J.E. Stice, "The Future of Engineering Education: Part 3. Developing Critical Skills," *Chem. Eng. Ed.,* **34**(2), 108 (2000)
4. Stice, J.E., R.M. Felder, D.R. Woods, and A. Rugarcia, "The Future of Engineering Education: Part 4. Learning How to Teach," *Chem. Eng. Ed.,* **34**(2), 118 (2000)
5. Alverno College, *Assessment at Alverno College*. Alverno College Publications, Milwaukee (1985)
6. Woods, D.R., *Problem-Based Leaning: How to Gain the Most from PBL*. Woods, Waterdown, Ontario, Canada (1994). Distributed by McMaster University Bookstore, Hamilton Ontario, Canada
7. Rogers, G.M., and J.K. Sando, *Stepping Ahead: An Assessment Plan Development Guide*, Rose Hulman Institute of Technology, Terre Haute, IN (1996)
8. Feldman, K.A., "The Superior College Teacher's from the Students' View," *Res. in Higher Ed.,* **5,** 43 (1976)
9. Cross, K.P., "On College Teaching," *J. Eng. Ed.,* **82**(1), 9 (1993)
10. Chickering, A.W., and Z.F. Gamson, "Seven Principles for Good Practice in Undergraduate Education," *AAHE Bulletin*, 3 (March 1987)
11. Eble, K.E., "Project to Improve Teaching," *Academe*, **4,** 3 (1970)
12. Woods,D.R., *Ideas for Teachers to Improve Learning*, 3rd ed., Chemical Engineering Department, McMaster University, Hamilton, Ontario, Canada (1999)
13. Ramsden, P., and N.J. Entwistle, "Effects of Academic Departments on Students' Approaches to Studying," *British J. of Ed. Psychology,* **51**, 368 (1981)
14. Terenzini, P.T., and E.T. Pascarelli, "Living with Myths: Undergraduate Education in America," *Change*, 28 (Jan–Feb 1994)
15. Woods, D.R., *Motivating Teachers to Improve Learning*. Department of Chemical Engineering, McMaster University, Hamilton Ontario, Canada, in press.
16. Watson, J.L., "An Analysis of the Value of the FE Examination for the Assessment of Student Learning in Engineering and Science Topics," *J. Eng. Ed.,* **87**(3), 305 (1998)
17. Angelo, T.A., and K.P. Cross, *Classroom Assessment Techniques: A Handbook for College Teachers*, 2nd ed., Jossey-Bass, San Francisco, CA (1993)
18. Panitz, B., "The Student Portfolio: A Powerful Assessment Tool," *ASEE Prism,* **5**(7), 24 (1996)
19. Rogers, G.M., and J. Williams, "Building a Better Portfolio," *ASEE Prism,* **8**(5), 30 (1999).
20. Felder, R.M., "What Do They Know, Anyway?" *Chem. Eng. Ed.*, **26**(3), 134 (1992). Available on-line at < *http://www2.ncsu.edu/unity/lockers/users/f/felder/public/Columns/Eval.html* >
21. Cashin, W.E., "Student Ratings of Teaching: The Research Revisited," *IDEA* Paper No. 32, *IDEA* Center, Kansas State University (1995). Available on-line at <*www.idea.ksu.edu*>.
22. McKeachie, W., "Student Ratings: The Validity of Use," *American Psychologist,* **52**(11), 1218 (1997)
23. Marsh, H.W., and L.A. Roche, "Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility," *American Psychologist,* **52**(11), 1187 (1997)
24. Hoyt, D.P., and W.H. Pallett, "Appraising Teaching Effectiveness: Beyond Student Ratings," IDEA Paper No. 36, IDEA Center, Kansas State University (1999). Available on-line at <*www.idea.ksu.edu*>. Forms for peer rating of instructional materials and for department head review of teaching assessment data are Appendices B and D of this paper.
25. Felder, R.M., "What Do They Know, Anyway? II. Making Evaluations Effective," *Chem. Eng. Ed.,* **27**(1), 28 (1993). Available on-line at < *http://www2.ncsu.edu/unity/lockers/users/f/felder/public/Columns/Eval2.html* >
26. McKeachie, W.J., *Teaching Tips: Strategies, Research, and Theory for College and University Professors*, 10th ed., Houghton Miflin, Boston, MA (1999)
27. Theall, M., and J. Franklin, eds., *Effective Practices for Improving Teaching.*, New Directions for Teaching and Learning No. 48, Jossey-Bass, San Francisco, CA, p. 89 (1991)
28. Felder, R.M., and R. Brent, "It Takes One to Know One," *Chem. Eng. Ed.,* **31**(1), 32 (1997). Available on-line at < *http://www2.ncsu.edu/unity/lockers/users/f/felder/public/Columns/Peer-rev.html* >
29. Keig, L., and M.D. Waggoner, *Collaborative Peer Review: The Role of Faculty in Improving College Teaching.* ASHE-ERIC Higher Education Report No. 2, Washington, DC, George Washington University (1994)
30. *Peer Observation of Classroom Teaching.* Report #CTL-15, Center for Teaching and Learning, University of North Carolina at Chapel Hill, Chapel Hill, NC (1994)
31. Seldin, P., *Changing Practices in Evaluating Teaching,* Anker Publishing Co., Bolton, MA (1999)
32. Woods, D.R., *Ideas for Teachers to Improve Learning*, 3rd ed., Chemical Engineering Department, McMaster University, Hamilton, Ontario, Canada (1999)
33. Felder, R.M., and R. Brent, "If You've Got It, Flaunt It: Uses and Abuses of Teaching Portfolios," *Chemical Eng. Ed.,* **30**(3), 188 (1996). Available on-line at < *http://www2.ncsu.edu/unity/lockers/users/f/felder/public/Columns/Portfolios.html* >
34. Seldin, P., *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*, 2nd ed., Anker Publishing Co., Bolton, MA (1997)
35. Seldin, P., *Successful Use of Teaching Portfolios*, Anker Publishing Co., Bolton, MA (1993)
36. Shore, B.M., S.F. Foster, C.K. Knapper, G.G. Nadeau, N. Neill, and V. Sims, *Guide to the Teaching Dossier, Its Preparation and Use*, 2nd ed., Canadian Association of University Teachers, Ottawa, Ontario, Canada (1986)
37. Diamond, R.M., *Preparing for Promotion and Tenure Review: A Faculty Guide*, Anker Publishing Co., Bolton, MA, (1995)
38. Weeks, P., "Developing a Teaching Portfolio." Available online at <*www.tals.dis.qut.edu.au/TalSSS/TALDU/papers/portfolio.html*>
39. Ramsden, P., *The Lancaster Approaches to Studying and Course Perceptions Questionnaires: Lecturer's Handbook*, Educational Methods Unit, Oxford Polytechnic, Oxford (1983)
40. Boyer, E., *Scholarship Reconsidered: Priorities of the Professoriate*, Carnegie Foundation for the Advancement of Teaching, Princeton, NJ (1990)
41. Felder, R.M., "The Scholarship of Teaching," *Chem. Eng. Ed.,* **34**(2), 144 (2000)
42. Glassick, C.E., M.T. Huber, and G.I. Maeroff, *Scholarship Assessed: Evaluation of the Professoriate,* Jossey-Bass, San Francisco, CA (1997) ❐