AC 2009-257: DEVELOPING METRICS TO EVALUATE INSTRUCTIONAL SCHOLARSHIP IN ENGINEERING

Richard Taber, National Academy of Engineering

Elizabeth Cady, National Academy of Engineering

Norman Fortenberry, National Academy of Engineering

Developing Metrics to Evaluate Instructional Scholarship in Engineering

Abstract

If valid and reliable means to assess instructional scholarship are identified, and they are accepted by the engineering community, then greater attention would be devoted to scholarly teaching by engineering faculty and departments. With this goal in mind, an ad hoc committee completed a study on the development and implementation of metrics for scholarly teaching or "instructional scholarship" within the discipline of engineering. The committee sought to identify new options (with respect to choices of existing metrics, processes for evaluation of metrics, and agents to perform the evaluation of metrics) for evaluating scholarly teaching and to assess broadly the options identified in terms such as their validity, reliability, and ease-of-use by engineering faculty. The intent is to contribute to greater acceptance of instructional scholarship within engineering disciplines. The committee examined specific choices for metrics of the scholarship of teaching, schemes for the evaluation of selected metrics, and agent(s) who will evaluate the selected metrics. This paper summarizes the committee report.

Introduction

Scholarship of teaching [1] is often compared with the scholarships of discovery and synthesis. Shulman [2] further categorized the scholarship of teaching as discovery scholarship within the educational domain [3] and scholarly teaching as teaching that (a) focuses on learning outcomes and teaching practices, (b) originates with knowledge of pedagogy and course content, and (c) includes self-reflection, discussions with peers, and participation in peer evaluation [4].

When engineering faculty members attend to the different ways in which students learn, the students become more engaged and also learn more course content and connections between engineering concepts. Unfortunately, scholarly teaching is not as easily assessed as traditional engineering research, which has well-defined metrics, so faculty members have little motivation to spend time on scholarly teaching [4, 5]. On the other hand, metrics to evaluate this behavior do exist [4, 6], so the committee worked to develop processes for administrators and faculty members to evaluate the metrics. The goal was to make the evaluation of scholarly teaching as easy for a promotion and tenure committee to complete as the evaluation of discovery and synthesis scholarship (e.g., number of publications, the prestige of the journal in which they appear). Although these research metrics suffer from several problems, they are identified and used, whereas metrics of teaching are not well-known.

Several factors motivated the effort to encourage scholarly teaching by engineering faculty. First, as with education in other fields, economic factors have led to governmental oversight for public educational institutions in order to show taxpayers that funds provided to higher education improve student learning [7]. Second, the ABET, Inc. accreditation standards changed to focus on student learning outcomes rather than the means and processes used to achieve them [8]. These new criteria directly challenged engineering faculty to maintain a high standard and continuous cycle of curriculum development, student assessment, and improvement that furthers institutional educational goals as well as supplementing faculty members' technical expertise [9]. Third, work in cognitive science and educational research, specifically Boyer [1] and governmental and National Academies reports [10, 11], became better known and respected among both academic and industry engineers. Finally, the use of rigorous engineering education research rather than trial-and-error course reform became popular in undergraduate institutions as a means to improve engineering education [12].

The National Academy of Engineering (NAE) assembled a committee composed of faculty professional development experts, engineering educators, and individuals knowledgeable in teaching assessment to examine methods by which faculty and administrators could evaluate scholarly teaching. The committee sought to identify new options (with respect to choices of existing metrics, processes for evaluation of metrics, and agents to perform the evaluation of metrics) for evaluating scholarly teaching and to assess broadly the options identified in terms such as their validity, reliability, and ease-of-use by engineering faculty [13]. The intent is to contribute to greater acceptance of instructional scholarship within engineering disciplines. The committee examined specific choices for metrics of the scholarship of teaching, schemes for the evaluation of selected metrics, and agent(s) who will evaluate the selected metrics. Each set of choices was made from a diversity of options; the culling process involved the development of process for evaluation of options within each of the set of choices to be made. The committee also sought the inputs of engineering faculty in the process of making choices. A November 2007 workshop brought together 25 experts in institutional administration, engineering education, and teaching and learning assessment in order to begin investigating existing metrics, research on assessment of learning, and the components and process of teaching effectiveness. The workshop included four presentations discussing the potential strategies for workshop attendees to create possible metrics, ways to assess them, and ideas of who should assess them. Attendees then discussed the presentations within breakout groups, which then reported back to the group. After a plenary discussion that attempted to bring about consensus, the committee began work on the final report. The final report, *Developing Metrics for Assessing Engineering Instruction: What Gets Measured is What Gets Improved*[13], is scheduled for public release in April 2009.

The committee assumed that teaching and learning will both improve with an available system to evaluate the effectiveness of teaching. This process also assumes faculty members' ability and motivation to improve their teaching, which implies their acceptance of feedback that will enable that change. Faculty members at every level, including future faculty members, should also develop the skills necessary to evaluate teaching effectiveness. Finally, administrators and others performing the evaluations are assumed to act fairly and objectively in all aspects of the system [13].

The report outlines general characteristics of acceptable metrics, beginning with the idea that if teaching effectiveness is rewarded by promotion and tenure committees it will be valued, and if teaching effectiveness is valued it will be rewarded [13]. Metrics and the overall system used to evaluate engineering faculty members must not diverge from existing institutional policies and organization, because promotion and tenure reviews occur at a university-wide level. Faculty evaluation for a unit in a larger institution must agree with the larger value system of the institution. The first step must be to ascertain the *faculty role model* of the institution. That is, what various professional roles faculty members are expected to play and how much each contributes to the overall tenure and promotion evaluation of the faculty. Each of the three traditional roles should be assigned a minimum and maximum weight. For example, teaching should contribute between 20% and 60% of the evaluation, scholarly or creative activities should contribute between 30% and 70%, and service should count for 10% to 15% of evaluation decisions [13].

In addition, system development should originate with engineering deans and department heads so as to take advantage of their connections to both individual faculty and administrators. Once these efforts have been initiated, the faculty at large should be involved in defining and determining the faculty role model. Although the final preparatory work of developing a codified faculty role model should be accomplished by a committee, all involved stakeholders should have input as to the final model. This process will be iterative and the final model and methods of assessing faculty should be continuously evaluated [13].

According to Arreola, Theall, and Aleamoni [14], effective teachers have five basic skills. First, they must have expertise in the content of the course. In addition, they must be skilled in designing engaging instructional experiences and must be able to deliver them. They also must be able to assess student learning in order to provide feedback that aids their progress. Finally, effective teachers must be able to manage all aspects of a course. Given the broad scope of abilities necessary to effective teaching, its evaluation must also encompass those aspects. The individuals who develop the evaluative system must agree on fundamental aspects of teaching effectiveness, including learning, course design, outcome assessments, and pedagogies [13]. The wide variety of skills necessary to effective teaching requires multiple measures from multiple sources, although not all sources will be equally useful in evaluating various aspects and some sources should be weighted more heavily than others. For example, the individual faculty being reviewed can provide feedback on his/her content expertise, design and delivery of instructional components, assessment, and management skills, although that element should not carry as much weight as the ratings from others. Experts in the instructor's field can weigh in on content expertise, instructional design and delivery, and assessment. Students can provide information about instructional aspects (i. e., design, delivery, assessment). Finally, the department chair or another supervisor can evaluate the assessment and course management abilities [13].

Evaluative tools are difficult to develop or adapt to engineering, but several campuses have individuals with expertise in psychometrics who could help engineering faculty members adapt an existing tool or develop and validate a new one. These tools will aid in data collection, but evaluators must also manage the data in order to analyze it. One method of organizing the large amount of data that will emerge is to develop or adapt tools that use the same scale, such as a 5 or 7-point Likert-scale. Given a specific faculty role model and ratings from several sources using a standard scale, evaluators can arrive at a quantifiable number using a weighted average of the scores provided by student ratings, peer ratings, department head evaluations, and self-assessment [13].

These tools should also be available to the faculty, either during professional development activities or in a central location, so they can know the areas in which they may need to improve. Central to this idea is the availability of professional development activities that target the skills that will be evaluated. This assumes that evaluation will include both summative and formative components, although faculty will use the formative feedback to improve their effectiveness while committees will use the summative metrics to judge promotion and tenure [13]. Metrics should also be collected longitudinally from a wide variety of stakeholders using multiple response methods (e. g., exams, teaching evaluations). The adaptable nature of this type of system allows for progress toward a variety of institutional goals and allows faculty members to experiment with new learning activities without fear of negative repercussions [13]. However, it also requires a skill set that engineering faculty and administration may need to acquire with the help of on-campus psychometric experts, ideally so both faculty and administrators can execute the various aspects of the evaluation without undue time and inconvenience. Finally, a good system of evaluation needs occasional reviewing from internal and external reviewers [13].

Summary and Conclusions

The report outlines both stipulations and recommendations for action. Although faculty members do attend programs designed to improve their teaching, resources and other time commitments limits participation. In addition, developing methods of assessing instructional effectiveness based on cognitive science and education research will produce quantitative metrics to be utilized during the promotion and tenure process, which will motivate faculty members to work to improve their teaching skills [13]. Development of this evaluation scheme should include substantial participation from both faculty members and administrators, and the final assessment process should be transparent to both faculty members who will be evaluated and to administrators and faculty members who will evaluate them [13]. Finally, the data collected for promotion and tenure summative evaluation should be separate from formative data collected by faculty members in an effort to improve their teaching through professional development activities [13].

The committee encourages evaluators, including department heads, engineering deans, and institutions, to follow four recommendations. First, teaching effectiveness should be evaluated using information gathered from several stakeholders (e. g., students, peers) regarding course content, organization, and delivery as well as student learning assessment. Second, department heads and engineering deans should promote the adoption of these metrics throughout both the school of engineering and the larger institution. Third, similar to building a department of individuals with content expertise and research skills, these leaders should form a team of expert evaluators. Fourth, they should utilize the resources developed by both on-campus teaching and learning centers and stand-alone programs focused on improving teaching abilities [13].

The committee also encourages the leaders within engineering (e. g., NAE, ABET, Inc., ASEE) to assist those institutions and individuals who are developing and using metrics of teaching effectiveness. These national organizations should strive to develop base models of evaluation that institutions can adapt to their situation [13].

The development of a thoughtfully designed and agreed-upon method of evaluating teaching effectiveness that are based on research on effective teaching and learning would provide administrators and faculty members with the wherewithal to use quantitative metrics in promotion and tenure decisions. Such metrics would also provide individual faculty members with an incentive to invest time and effort in developing their instructional skills, because they would be favorably reflected in advancement decisions. Developing metrics for evaluating instructional effectiveness should be undertaken with the understanding that all faculty and the administration will have significant input into the design of the evaluation system, as well as feedback from the results. The assumptions, principles, and expected outcomes of the evaluation method should be explicit (and repeated frequently) to those who will be subject to evaluations, as well as those who will participate in administering the evaluations [13].

1. Boyer, E.L. (1990) *Scholarship reconsidered: Priorities of the professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.
2. Shulman, L. (1993) Teaching as community property: Putting an end to pedagogical solitude. *Change, 25 (6)*, 6-7.
3. Streveler, R.A., Borrego, M., & Smith, K.A. (2006) *Moving from the "scholarship of teaching and learning" to "educational research:" An example from engineering*. In *To improve the academy (vol. 25)*. Anker Publishing Company: Bolton, MA.
4. National Research Council. (2002). *Evaluating and improving undergraduate teaching in science, technology, engineering, and mathematics*. Washington, DC: National Academies Press.
5. Wankat, P.C., Felder, R.C., Smith, K.A., & Oreovicz, F.S. (2002) *The scholarship of teaching and learning in engineering*. In *Disciplinary styles in the scholarship of teaching and learning: Exploring common ground*, M.T. Huber and S.P. Moreale, Editors. American Association for Higher Education and Carnegie Foundation for the Advancement of Teaching: Washington, DC.
6. Arreola, R.A. (2000) *Developing a comprehensive faculty evaluation system*. 2nd ed. Bolton, MA: Anker Publishing Company.
7. United States Department of Education. (2006). *A test of leadership: Charting the future of U. S. higher education*. Washington, DC: USDEd.
8. Accreditation Board for Engineering and Technology. (1995). *Vision for change: A summary report of the ABET/NSF/industry workshops*. 95-VIS. Baltimore, MD: Accreditation Board for Engineering and Technology.
9. ABET, I. (2008) *2008-09 Criteria for evaulating engineering programs*. Available online at http://www.abet.org/Linked%20Documents-UPDATE/Criteria%20and%20PP/E001%2008-09%20EAC%20Criteria%2011-30-07.pdf.
10. National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academies Press.
11. National Research Council. (1999) *How people learn: Bridging research and practice*. M.S. Donovan, J.D. Bransford, and J.W. Pellegrino, Editors. National Academies Press: Washington, DC.
12. Lohmann, J. (2005) Building a community of scholars: The role of the Journal of Engineering Education as a research journal. *Journal of Engineering Education, 94*, 1-4.
13. National Academy of Engineering. (2009). *Developing metrics for assessing engineering instruction: What gets measured is what gets improved*. Washington, DC: National Academies Press.
14. Arreola, R.A., Theall, M., & Aleamoni, L.M. (2003) Beyond scholarship: Recognizing the multiple roles of the professoriate. *AERA Convention*. Chicago, IL.