

AC 2007-127: DIRECT ASSESSMENT MEASURES

Joseph Shaeiwitz, West Virginia University

Joseph A. Shaeiwitz received his B.S. degree from the University of Delaware and his M.S. and Ph.D. degrees from Carnegie Mellon University. His professional interests are in design, design education, and outcomes assessment. Joe is an associate editor of the Journal of Engineering Education, and he is a co-author of the text Analysis, Synthesis, and Design of Chemical Processes (2nd ed.), published by Prentice Hall in 2003.

Daina Briedis, Michigan State University

DAINA BRIEDIS is a faculty member in the Department of Chemical Engineering and Materials Science at Michigan State University. Dr. Briedis has conducted research in bioadhesion; she is currently studying development of integrated approaches to using computation tools to support technical problem solving throughout the curriculum. She is active nationally and internationally in engineering accreditation, is an ABET IDEAL Scholar, and is a member of the ABET Board. She leads the assessment and evaluation efforts in her program.

Direct Assessment Measures

Introduction

Engineering programs have recently completed or are in the process of preparing for their second accreditation visit under the assessment-based criteria. Based on our combined experience of over 30 visits as both program evaluators and team chairs, it appears that many programs are struggling to identify valid measures for their program outcomes to come in full compliance with the requirements of Criterion 3 of the Engineering Criteria. This is substantiated by evidence of the relatively large number of citations for shortcomings relative to some aspect of this criterion.¹ One cause of this is that many programs rely very heavily on surveys and similar indirect, or “soft,” measures of these outcomes. We believe that there is too much reliance on these indirect assessment measures and programs should endeavor to make direct assessment a cornerstone of their program improvement processes.

Indirect assessment measures include end-of-course surveys, graduation surveys, and alumni surveys in which an evaluation is based on opinion or self-reporting. These assessment measures are necessary but not sufficient, and some may be more appropriate for evaluation of objectives rather than demonstration of outcomes. Direct assessment by the faculty is necessary to provide an objective measure of students’ achievement of educational outcomes. All quality assessment plans should include direct assessment measures.

ABET is taking a harder line on the need for direct assessment as part of demonstrating that students have the required skills, knowledge, and attitudes comprised in the eleven ABET-designated outcomes. ABET has received significant negative feedback from engineering programs about the inconsistencies in program evaluator findings for compliance with Criterion 3. While it is true that, in the past, programs may have passed through accreditation visits with only indirect assessment (surveys, faculty opinions, *ad hoc* data) as evidence of student achievement in outcomes, it is likely that this will no longer be the case. Programs relying **only** on indirect measures for outcomes assessment will likely be cited with shortcomings in future accreditation visits.

This paper deals mainly with assessment of program *outcomes*—the knowledge, skills and attributes that students should demonstrate by the time of graduation. Therefore, the focus is campus- or curriculum-based assessment. This paper does not address *program educational objectives*, which describe career and professional accomplishments of program alumni. The former usually requires different evaluation and assessment tools than the latter; although, some overlap does exist. In addition, we will also refer to *course* objectives, which are not to be confused with program educational objectives.

Therefore, given this context and the imminent importance of using direct assessment methods, this paper provides a review of direct assessment measures. These methods include, but are not limited to, instructor end-of-course assessments, use of targeted assignments (assigned problems, exam questions, projects), capstone examinations (including the FE Exam), student portfolios, and use of capstone experiences.

The Case For and Against Indirect Assessment Measures

Indirect assessment measures, such as surveys, focus groups, and interviews, are an important part of any assessment plan. They are an excellent method for obtaining results on program objectives from alumni several years after graduation. However, they are supportive of but insufficient for assessing program outcomes for several reasons. Surveys provide an uncalibrated snapshot of what students *believe* they have learned. It can be argued that relying exclusively on indirect or self-assessment measures for program outcomes is tantamount to allowing students to grade their own exams and projects and assign their own grades. Professional expertise and judgment are just as important in assessing program outcomes as they are in assigning grades.

Most experienced faculty members will, at one time, have either said or heard a colleague say something like: “These students will not admit to having learned anything in previous classes, even though I know it was taught.” Conversely, we probably have all experienced the student who claims to be able to do everything but who repeatedly fails to prove it. In either case, these comments clearly demonstrate the inherent inconsistency with relying on student self-assessment to evaluate program outcomes.

Heavy or exclusive reliance on indirect assessment measures is not specifically prohibited by ABET. In fact, at this time, the Engineering Criteria do not specifically state which assessment methods are to be used.² However, it could be argued that heavy or exclusive reliance on indirect assessment measures does not allow a program to demonstrate achievement of program outcomes. So, while indirect, self-assessment measures may be a piece of an assessment plan—and in some instances may be the only way to measure some aspects of program outcomes,³ a more robust method is needed to ensure valid measures of program outcomes.

The Case for Direct Assessment Measures

Faculty members routinely evaluate student performance when grading exams, when grading projects, and when assigning final grades. Graded assignments and final grades differ from assessment because assessment is similar to a distributed parameter model while assignments and final grades are similar to a lumped parameter model. Program assessment is focused on providing demonstration of student knowledge or skill directly linked to a specific program outcome.⁴ Grades depend on a faculty member’s expectations for a course, course objectives, and the particular grading policy. Grades are often norm-based and reflect a student’s standing relative to others within a class or on a test. Assessment is based on a student’s absolute performance on a specific measurable aspect of an outcome

For example, on an exam, the average might be 80%, but it is possible that the entire class missed the same 20% question. In this caricature case, achievement of whatever program outcome or course objective that “missed” question was supposed to measure is in doubt, yet the 80% average is reasonably good performance relative to overall course content. Therefore, composite scores on graded assignments and composite course grades do not provide the level of detail desired for assessment of program outcomes.

Review of Direct Assessment Measures

This section describes some typical direct assessment opportunities, along with some advantages and disadvantages of each. A more comprehensive discussion of the advantages and disadvantages of all assessment measures is available.⁵

End-of-Course Assessments

All course syllabi must have course objectives. One of the instructor's jobs is to determine the success in meeting these course objectives. This is usually accomplished by mapping individual problems assigned into course objectives so that student performance on each problem provides evidence regarding achievement of its corresponding course objective. If the course objectives are mapped into program outcomes, this approach provides evidence for student achievement in the outcome areas specifically covered in that course.⁶

This is relatively quick and easy, and may be done routinely by instructors when they review problem assignments or exams to see on what problems students did well and on what problems students did poorly. This can be a component of an assessment plan, but we believe that it should not be the only direct assessment measure in a high-quality assessment plan.

Targeted Assignments

Another direct assessment measure is to identify certain assignments and/or problems throughout the curriculum that are key determinants of achievement of specific program outcomes. Student performance on these assignments is used to assess these specific outcomes. This is similar to end-of-course assessments, but a major difference is that the entire faculty must be involved in determining which assignments in which classes are used. Having multiple problems/assignments for a specific outcome is recommended. It is also recommended that these multiple experiences be integrated throughout the curriculum, so that student progress toward achieving an outcome can be measured. For example, having oral and/or written communication experiences throughout the curriculum, with feedback after each experience, is not only sound pedagogy, but it provides a method to demonstrate student progress toward mastery of these skills.

With a wide variety of targeted assignments and the involvement of the entire faculty, this would be a sufficient direct assessment method. A clear advantage of this method is that the effort required by instructors is incremental to the effort required to grade the targeted assignments. However, a strategy to require the different instructors to perform the assessment reliably is needed. This is discussed later.

Capstone Experiences

Capstone experiences are where students are supposed to apply what they have previously learned to a comprehensive, usually design-oriented, problem. Therefore, this is a very logical place to assess what students have learned. Furthermore, since these experiences are usually done in teams and they usually involve written reports and oral presentations, the professional

skills (teamwork, communication, global/societal context, life-long learning, contemporary issues) can be assessed similarly. Laboratory experiences may also fall into this category, especially if the laboratory class comes after the basic material is taught in the curriculum. More details of one program's use of capstone experiences are available.⁷

A major advantage of using capstone experiences is that they are already required in the curriculum and that only incremental effort is required to perform assessment while assignments and projects are graded. However, just as for targeted assignments, a methodology for different instructors to perform the assessment reliably is needed.

Capstone Examinations

There are two types of capstone examinations, the FE Exam and department-generated comprehensive examinations. On the surface, the topic-specific information provided by most state boards can provide data on achievement of subject mastery. However, most program outcomes, at least in chemical engineering, typically do not list subjects (*i.e.*, fluid mechanics, thermodynamics) in their outcomes. Furthermore, there are differences of opinion as to whether the FE Exam provides sufficient information for assessment of program outcomes.^{8,9} Another problem is that most disciplines, chemical engineering included, do not require the exam, and the numbers of student taking the exam may not be statistically significant.

In our experience as program evaluators, some programs that have tried to institute their own capstone exam have not been pleased with the results. In these cases, the exam was not part of a course—just taking the exam was the only course requirement, and there was no performance requirement on the exam for graduation. As a result, students did not take the exam seriously. On the other hand, it could be argued that if students really learned the material, they would not need to prepare to perform satisfactorily. It is also possible that the exams given were not designed appropriately for the desired purpose. Therefore, extreme caution is recommended before implementing a program-level, comprehensive outcomes assessment exam.

Portfolios

Portfolios are a collection of student work that demonstrates the skills they have learned. An appropriate collection of such work throughout the curriculum can be used to demonstrate achievement of program outcomes by an individual student as well as demonstrate the progress the student made while proceeding toward achieving the outcome. By choosing a random, but representative, sample of students from which to collect portfolios, achievement of outcomes can be demonstrated.

This sounds far easier than it actually is. It is not sufficient merely to collect the portfolio. (Note: This is why course notebooks alone are not valid assessment measures.) The portfolio must be assessed. This is one of the disadvantages of portfolios. In several of the previously discussed methods, assessment was performed while assignments were graded, which is an excellent tool for minimizing the time required of faculty. Here, the assessment is done when the portfolio is complete, often by those not involved in the instruction in the courses from which the

portfolio components were taken. This requires a time commitment that most faculties are not willing to make.

However, if such a time commitment were to be made, this is an excellent assessment measure. One of the best examples of portfolios is the Rose Hulman electronic portfolio (RosE-portfolio), in which students submit work electronically into a data base and write a brief reflection on the work. Work is submitted in categories related to program outcomes. Then, a group of faculty members assesses the portfolios.¹⁰ Once again, a methodology for different faculty members to perform the portfolio assessment reliably is needed.

Developing Performance Criteria and Rubrics

Outcomes must be measurable. This is required by ABET criteria and is a means by which a program can truly benchmark the status of student learning. For effective assessment of program outcomes, performance criteria are needed. Constructing this set of descriptors to make outcomes *measurable* is one of the most important steps in implementing and sustaining an effective assessment process and may also be one of the more difficult and time-consuming. However, when designed well, a good set of measures provides for consistency and reliability of direct assessment.

For example, suppose a program outcome is that students will be able to communicate effectively in written reports. Program faculty should understand for themselves, in the context of their own programs, what this outcome will “look like” when students are demonstrating its traits (performance criteria). Performance criteria define what it means to be able to communicate effectively. One such performance criterion might be that students will prepare reports that are written appropriately. Another performance criterion might be that students will follow a prescribed format in preparing a written report. One can then define attributes of a performance criterion. For example, under appropriate writing, correct grammar, correct spelling, and correct punctuation might be included. Under prescribed format, the content of a cover memo, abstract, introduction, results, discussion, etc., can be described. Table 1 shows a sample rubric for written reports. Similar rubrics for oral presentations, designs, and laboratory are available.¹¹

As mentioned above, developing rubrics takes initial effort. Once developed, rubrics must be tested, and deviations between raters of the same instrument should not exceed one level. It is rare that a rubric will prove satisfactory without modifications after testing, and periodic re-evaluation and modification are also necessary. Once performance criteria and/or attributes are defined, the best way to proceed is to define what superior performance is and what not acceptable performance is, *i.e.*, the extremes should be defined first. Precisely defined wording is necessary. For example “not acceptable” or “excellent” is not appropriate since these words can mean different things to different people. Definitions of “not acceptable” or “excellent” are needed. Then, the middle ground can be filled in. It is up to the user as to how many levels of achievement are defined. We suggest a minimum of three and a maximum of five. Additional suggestions and instructions are also available.^{12, 13}

Table 1: Rubric for Written Presentations

Attribute	1-Not proficient	2-Progressing to proficiency	3-Proficient	4-Superior proficiency	Score
Report Mechanics					
Organization	inappropriate content of most sections of report	some content in inappropriate section of report	content appropriate to all sections of report	unique organization enhances readability and/or understandability of report	
Complete Story Told	no story told, very incomplete	aspects of complete story missing	complete story told	additional material enhances quality of report	
Aesthetics	unacceptable – e.g., tables and figures cannot be read/understood, fonts difficult to read	some portions are sloppy and difficult to read	text, tables, figures readable and understandable	text, tables, figures so clear and understandable as to enhance report impact	
Format	so many format errors as to make report useless	some format errors	format followed	unique format aspects that enhance report impact	
Spelling	any spelling errors	only spelling errors are different spellings for same pronunciation	no spelling errors		
Grammar and Punctuation	too many grammar and punctuation errors	grammar and punctuation errors	only a very few minor grammar or punctuation errors	no grammar or punctuation errors	
Length	far too long or too short	too long or too short	appropriate length		

Content		simply says “here it is”	includes key results and recommendations	so clear and complete as to enhance impact of report	
Cover Memo – problem stated, conclusion summarized	not present	problem stated somewhat, significant results not included	problem clearly stated, key results clearly stated	so clear and complete as to enhance impact of report	
Abstract/Executive Summary – problem stated, conclusion summarized	problem not stated, conclusion not summarized, only stated what did	problem stated, perhaps poorly, no or limited discussion of constraints	problem clearly stated, constraints clearly discussed	so clear and complete as to enhance impact of report	
Introduction – problem stated, constraints explained	problem not stated, constraints not explained, contains results/conclusions	unclear theoretical development	clear theoretical development	so clear and complete as to enhance impact of report	
Theoretical development – appropriate detail	no theoretical development	unclear or incomplete	clearly stated, for design report, PFD, stream table, equipment summary, utility summary all present	so clear and complete as to enhance impact of report	
Results – presented, methodology clear, problem solved	missing results, for design report PFD and stream table not first	unclear explanations, for design report, few optimizations	clear, for design report, optimizations presented and explained	so clear and complete as to enhance impact of report	
Discussion – physical explanations provided	no physical explanations provided				

Conclusions/Recommendations – present, significance explained, no new ideas	not stated, includes ideas not already discussed in report	weak, includes ideas not already discussed in report, wishes rather than recommendations	clear, clearly follow report discussion, meaningful recommendations	so clear and complete as to enhance impact of report	
References – provided as appropriate	incorrect format, not cited in report		correct format, all cited in report		
Appendix	appendix does not contain material that should be in text, text does not contain material that should be in appendix	appropriate locations, appendix not clear and easy to follow	appropriate locations, appendix indexed, clear and easy to follow	so clear and complete as to enhance impact of report	
Effective use of software to prepare written report	makes no use of or incompetent use of software features	poor use of or uses very few software features	uses appropriate software features correctly	uses unusual software features and/or uses software features uniquely to enhance report effectiveness	

Rubrics for *grading* can also be developed. In this case, there would probably either be five levels (A through F) or three levels (A, C, F, with interpolation used). Numerous examples of such rubrics are available.¹⁴ An advantage of using rubrics such as these for grading is that students are provided with a clear understanding of what is required to earn a specific grade, which is most useful for projects and similar assignments that, on the surface, are graded more subjectively. Grading rubrics can also be developed to evaluate problems on exams.

Conclusion

Demonstration of effective assessment processes is necessary for accreditation, and the bar for standards for compliance with the new criteria has been raised subsequent to first-round accreditation visits. It is most significant that, as part of learning about best practices, new program evaluators are being trained to look for direct assessment in outcomes assessment processes. Faculty who have attended assessment workshops or training programs are also becoming savvy in understanding that indirect measures alone are insufficient for effective outcomes assessment; those programs interested in best practices will be intrinsically motivated to use direct methods as a key element of their assessment processes.

If programs still plan on seeking ABET accreditation in the future, assessment will be required, and if it must be done, why not gather valuable information? The authors strongly believe that all high-quality assessment plans should include direct assessment methods to complement surveys and similar indirect assessment methods.

Bibliography

1. Briedis, D., "Enabling Faculty in Effective Program Evaluation and Assessment," workshop conducted at the "Best Assessment Processes Symposium," February 27-29, 2006, Rose-Hulman Institute of Technology, Terre Haute, IN.
2. *Criteria for Accrediting Engineering Programs*, ABET, Inc., <http://www.abet.org>.
3. Rogers, G., "Assessment 101: Direct and Indirect Assessments," Community Matters, ABET, Inc., August, 2006.
4. Nichols, James O., "The Departmental Guide and Record Book for Student Outcomes Assessment and Institutional Effectiveness," Agathon Press, NY (1995).
5. Prus, J. and R. Johnson, "A Critical Review of Student Assessment Options," in *Assessment and Testing: Myths and Realities* (T. H. Bers and M. L. Miller, eds.), New Directions for Community Colleges, no. 88, Jossey-Bass, San Francisco, 1994, pp. 69-83.
6. Winter, H. H., "Using Test Results for Assessment of Teaching and Learning," *Chemical Engineering Education*, **36** (3), 188-191, 2002.
7. Shaeiwitz, J. A., "Mining Capstone Engineering Experiences for Program Assessment Results," *International Journal of Engineering Education*, **18** (2), 193-198 (2002).
8. Watson, J. L., "An Analysis of the Value of the FE Examination for the Assessment of Student Learning in Engineering and Engineering Science Topics," *Journal of Engineering Education*, **87** (3), 305-311 (1998).
9. Nirmalakhandan, N., D. Daniel, and K. White, "Use of Subject-Specific FE Exam Results in Outcomes Assessment," *Journal of Engineering Education*, **93** (1), 73-77 (2004).
10. <http://www.rose-hulman.edu/REPS/>
11. <http://www.che.cemr.wvu.edu/ugrad/outcomes>
12. <http://edweb.sdsu.edu/webquest/rubrics/weblessons.htm>

13. <http://edweb.sdsu.edu/webquest/rubrics/rubrics.html>
14. Waalvord B. E. and V. J. Anderson, *Effective Grading. A Tool for Learning and Assessment*, Jossey-Bass, San Francisco, 1998.