

A Methodology for Selecting Program Outcomes Assessment Methods

Gregory F. Starzyk, JD, MPM, CPC, Assoc. DBIA
College of Architecture and Environmental Design
California Polytechnic State University
San Luis Obispo, California

Benefits of program outcomes assessment include fostering academic introspection; providing information for recruitment, planning and accreditation; improving institutional, college and department image; enhancing funding opportunities; energizing relationships with alumni, professional associations, and trade groups; and attracting better students. This paper puts forth a selection criteria matrix as a methodology for selecting assessment methods.

Key Words: Outcomes Assessment, ACCE, Educational Objectives, Learning Outcomes, Criteria

Introduction

As legislatures wrestle with a tenacious and enduring financial crisis the pressures on the higher education system are arriving from every direction with increasing intensity and persistence. Construction management departments and their faculty are not immune to these pressures. In this environment it behooves educators to demonstrate unequivocally the values that their programs deliver to all of their stakeholders: the construction industry, trade and professional associations, private sources of funding, accrediting bodies, alumni, parents, the general public, students and others. Value may not be self-evident. It must be demonstrated.

A benefit that should not be overlooked is that systematic outcomes assessment - when it evidences the fulfillment of our duties as educators, when it demonstrates value delivered beyond expectations, when it delivers ever-improving quality – shifts the burden of accountability back to those who have the obligation to support our worthy endeavors.

> "There is a compelling public stake in education. As educators, we have a responsibility to the publics that support or depend on us to provide information about the ways in which our students meet goals and expectations. But that responsibility goes beyond the reporting of such information. Our deeper obligation - to ourselves, our students, and society - is to improve. Those to whom educators are accountable have a corresponding obligation to support such attempts at improvement" (Association of American Colleges and Universities & the Council for Higher Education Accreditation, 2008).

This paper puts forth a selection criteria matrix as a methodology for selecting program outcome assessment methods.

Definition and Purpose of Outcomes Assessment

[Outcomes] assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development (Marchese, T.J., 1987). The overriding purpose of [outcomes] assessment is to understand how educational programs are working and to determine whether they are contributing to student growth and development (Marchese, T.J. 1987).

Terminology

The American Council for Construction Education (the "ACCE") has a preferred assessment taxonomy. Yet §9.3.2 of the ACCE's new additions, effective July 12, 2012, to Document 103 *Standards and Criteria for Baccalaureate*

and Associate Programs, anticipates that differing terminology may be in place at educational institutions. Appendix A, Terminology Matrix, compares California Polytechnic State University at San Luis Obispo (Cal Poly) terminology with ACCE terminology. In the matrix, emboldened terms highlighted in red have been added to illustrate commonality in meaning between differing terms. Cal Poly's terms (underlined in the first column on the left) will be used throughout. A comparison of Cal Poly terminology with that of the Accreditation Board for Engineering and Technology (ABET) is beyond the scope of this paper.

Figure 1 "Outcomes Assessment Element Hierarchy" provides a graphic depiction of the entire outcomes assessment process. There are six levels in the hierarchy. Six ACCE terms are provided to align Cal Poly Terminology with §9.3.1 of the ACCE's new additions, effective July 12, 2012, to Document 103 and one ACCE term "Assessment Tool" to align Cal Poly Terminology with §9.3.4 of Document 103.

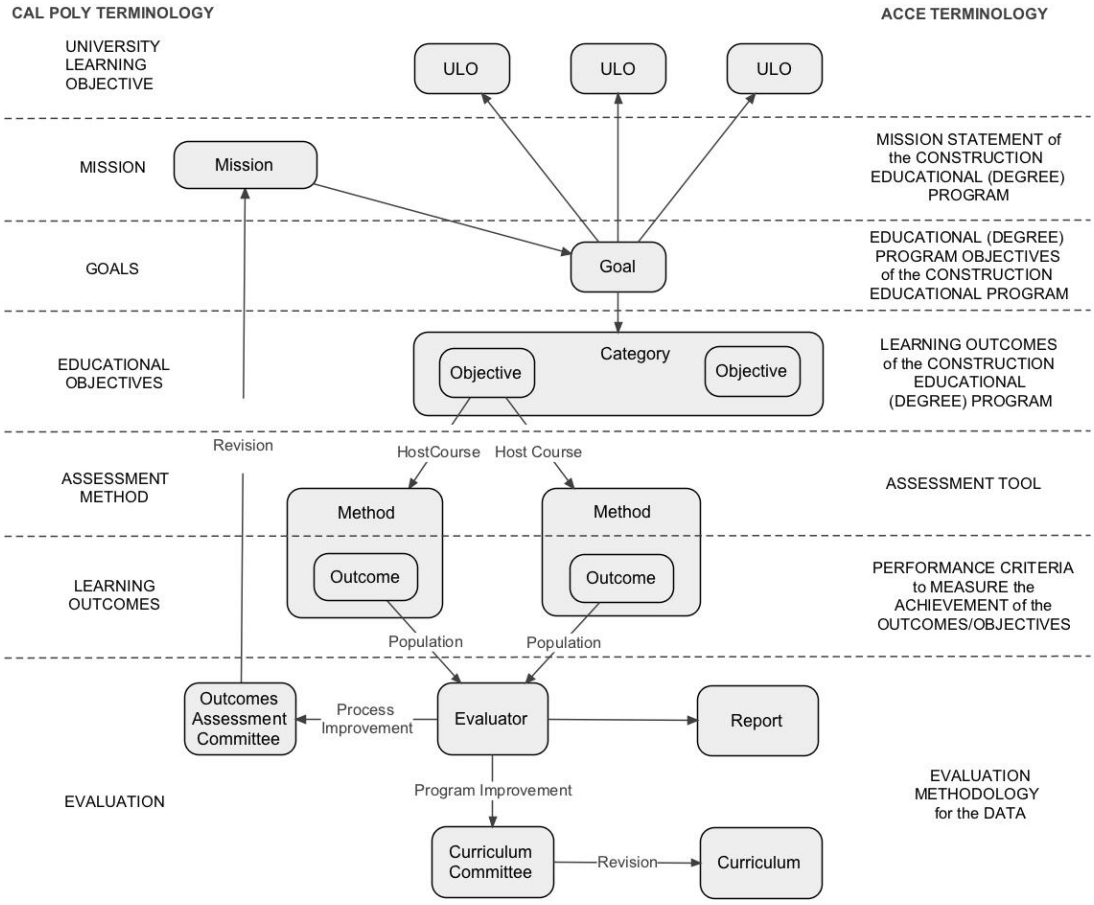

Fig. 1: Outcomes Assessment Element Hierarchy

Assessment Method Selection Criteria

Assessment methods may be ranked by the complexity of the particular learning outcome being measured (see Figure 2), as follows:

- *Integrated Development Outcomes* (combining skill, cognitive and/or affective outcomes, e.g. group dialogue, public speaking, relationship building, group problem-solving, knowledge integration).
- *Occupational Success Outcomes* (affective subset, e.g. dependability, initiative, self-motivation), and
- *Affective Outcomes* (values and attitudes, e.g. exhibiting personal discipline, managing emotions, displaying integrity, clarifying purposes, practicing ethics, providing leadership);

- *Cognitive Outcomes* (thinking skills, e.g. topic-driven outcomes);
- *Skill Outcomes* (psychomotor skills, e.g. drawing skills, BIM dexterity, jobsite mobility);

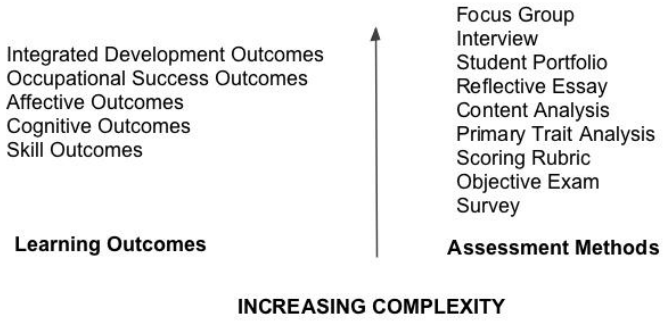

Fig. 2: Learning Outcomes and Assessment Methods Complexity Scale

An assessment method can be as simple or as complex as the outcome that it is testing. Complex assessment methods can be used to test simple outcomes. But simple assessment methods should not be used to test complex outcomes. The assessment method should fit the learning outcome. The proposed assessment selection method poses fit as a threshold question. Figure 2 provides an approximate complexity scale for learning outcomes and assessment methods.

The Assessment Method Selection Process

The process of selecting an assessment method begins by asking faculty to identify the criteria that they consider the most important in an assessment instrument. Develop a list of the criteria that matter. The ability of any method to address a particular assessment question is the most important consideration (Allen, et. al., 2002). Multiple-choice exams are useful for assessing knowledge and understanding of course content. A multiple-choice exam, however, is less useful for testing the more complex integrated development outcomes, such as adeptness in social interactions. For that, peer evaluations, observations by faculty, or employer feedback are more useful.

Typical criteria are, as follows:

Relevance: Will the method satisfactorily answer the assessment question?
Utility: Is the method fit for the purpose and useful for answering the assessment question?

Does the method have validity? Will the instrument measure what we want it to measure? It is never easy to answer this question. Gathering evidence of a particular instrument's validity is a time-consuming and challenging task. Validity is less of a definable measure than it is an overall evaluative judgment.

Construct: Is there congruence between the meaning of the underlying construct and the elements of the test instrument? (Results should differ between groups who are expected to differ. Exiting senior should be more ready to perform in construction than entering freshmen.)
Criterion: Includes concurrent validity and predictive validity. Concurrent validity is the dependability of the relationship between the scores on an instrument and the characteristic of interest. (Is the work in a portfolio the student's own work or is it actually someone else's work or her team's work?) Predictive validity is the dependability of the relationship between the scores on an instrument and a particular future outcome. (Will students who score highly on an "ability to work in groups" assessment instrument demonstrate that ability in future employer evaluations?)
Content: Does the content of the instrument focus upon the curriculum or other area that we are interested in assessing? Does it provide evidence that a particular learning vehicle was successful? (Are students learning how to do cost estimating in a particular class in the curriculum or are they learning it in their independent preparation for an Associated Schools of Construction (the "ASC") student competition?)

Is the instrument reliable? An instrument is reliable when what is actually being measured – such as knowledge, performance or attitudes - accounts for the variance in its scores. (Banta, et. al., 1996)

Error: The score variance on unreliable instruments is due to measurement error or the administration and scoring of the instrument

Consistency: Are consistent responses produced over time? Instruments are inconsistent when they are poorly constructed (unclear words and ambiguities in the questions), the length of the instrument both for testing and grading exceeds the time available for it, or different raters do not agree on the meaning of items in its rating scale.

Motivation: Error is often attributable to the motivation of student, faculty, alumni or others responding to the instrument Choose testing instruments that student, alumni and employers will find valuable and that will encourage them to cooperate.

Course-Embedment: Lengthy tests taken outside of class have to overcome motivation problems. In contrast, exams that are taken in class are taken seriously, particularly because they are graded. This is the strongest argument for "course-embedded" assessment.

External Input: Former students, alumni and employers may or may not be motivated to respond to surveys, interviews or other assessment instruments. Their level of participation and the motivation of those who participate must be carefully considered.

Does balance exist? There is a tradeoff between reliability and validity. The more complex the outcome the more effort we must put into the development of an authentic testing method (Eder, 2001). Authenticity increases an instrument's validity. The more authentic we make an instrument, the more unique it is going to be. Unique instruments suffer from a lack of widespread use and the standardization from which reliability arises. We are challenged to design instruments that achieve a suitable balance between reliability and validity.

What are the costs? There are at least two types of costs that are attributable to assessment activities: money costs, and opportunity costs.

Money Costs: Standardized national tests cost money. Surveys may involve mailing, telephone or internet charges. There is a cost for printing testing materials. Faculty may have to be trained to rate and grade consistently. Trainers of faculty may have to be hired.

Opportunity Costs: Opportunity costs are a particularly important concern. Faculty will invest considerable time to develop test instruments, implement them, grade them, administer them, evaluate them and improve upon them Faculty can only work on outcomes assessment at the expense of not working on something else. Students, alumni, and employers also participate in assessments at the expense of not doing something else. Is it worth the trade off?

The Assessment Method Selection Criteria Matrix

The assessment method criteria matrix is configured around four elements: Validity, Reliability, Balance and Cost.

Validity is the totality of Construct, Criteria and Content. It is scored on a 5-point scale where 0 is a low level of validity and 5 is a high level of validity.

Reliability is a combination of Error, Consistency and Motivation. Error should reflect measurement error. Frequently, measurement error occurs when the time required to take, administer and score the test exceeds the time available for taking, administering and scoring. Consistency should reflect the amount of resources invested in developing and improving the test. This would include investment in faculty training to score the test consistently. Course-embedment should rank highest in motivation and external input (as with employer surveys) should rank low in motivation. Reliability is scored on a 5-point scale where 0 is a low level of reliability and 5 is a high level of reliability.

Balance is the differential between the validity and reliability criteria, entered as a negative number. For example

when validity = 5 and reliability = 3, enter -2 for Balance. When validity and reliability are equal enter 0 for Balance.

Cost is money spent plus the opportunity cost of faculty doing assessment work when they could be doing something else. It is scored on a 5-point scale where 0 represents high cost and 5 represents low cost.

Assessment Method Criteria Matrices, Examples

Figure 3 is a criteria matrix for selecting a comprehensive exam. This particular example compares three different types of examination formats, administered to all graduating students: 1) The American Institute of Constructors & Constructor Certification Commission's (the "AIC") Level 1 Exam, Autonomous, where exam scores do not affect students' grades; 2) AIC Level 1 Exam, Embedded, where all or part of students' grades in a designated course depends on their exam scores; and 3) Faculty Designed, Autonomous, an exam that is comparable to the AIC Level 1 Exam, where all or part of students' grades in a designated course depend on their exam scores.

Please note that these criteria scores reflect the subjective judgment of the author. In practice, criteria matrices should be jointly prepared, in an integrative fashion, by members of a faculty committee.

	Validity	Reliability	Balance	Cost	Score
AIC Level I Exam, Autonomous	4	3	-1	5	11
AIC Level I Exam, Embedded	4	5	-1	5	13
Faculty Designed, Embedded	5	4	-1	1	9

Figure 3: Assessment Selection Criteria Matrix for Comprehensive Exams

The "AIC Level I Exam, Embedded" earned the highest score because of its low opportunity cost and high reliability. There is no opportunity cost because the exam is already developed. Its reliability is high because it is embedded so as to properly motivate students and it is a widely used, tested and maintained exam. The detailed analysis follows. On validity, the faculty developed exam scored higher on content because it can be more focused on the curricula than a national exam can ever be. All three assessments were scored highly on criterion, both in concurrent validity and predictive validity, and in construct. On reliability, the AIC Level I, Embedded exam scored highest in reliability because results should be consistent and it is properly motivated. The AIC Level I Exam, Autonomous scored lowest because a serious motivation problem can be anticipated. The Faculty Designed, Embedded exam also scored highly on motivation but was marked lower on error because development, testing, and improvement by the faculty cannot hope to equal development, testing and improvement of the AIC exam, at least over the near term. On balance, all three exams are equally balanced between validity and reliability. On cost, there would be a very large opportunity cost for the faculty to develop and maintain a comprehensive exam similar to the AIC Level 1 exam. Therefore, it received the lowest score on cost. The AIC exam gets a high score because of zero opportunity cost (except for a proctor).

Figure 4 is an assessment criteria matrix for selecting between different types of student portfolios, As an assessment instrument we have the choice of a single portfolio, scored at the end of the senior year and a longitudinal portfolio scored at various points in time. For this example an evaluation at the end of the sophomore year was chosen. We also have the choice of grading every portfolio or just a small sampling of the portfolios. For this example a stratified sample was chosen. A stratified sample consists of two or more small samples, each sample with a different demographic, such as gender or socio-economic background.

	Validity	Reliability	Balance	Cost	Score
Final Porfolio, All Students	3	3	0	2	8
Longitudinal Portfolio, All Students	3	3	0	1	7
Final Portfolio, Stratified Sample	3	3	0	4	10
Longitudinal Portfolio, Stratified Sample	3	3	0	3	9

Figure 4: Assessment Selection Criteria Matrix for Student Portfolios

"Final Portfolio, Stratified Sample" earned the highest score because, all other things being equal, it has the lowest opportunity costs. The detailed analysis follows. On validity, all of the methods have a concurrent validity problem because it may be difficult to determine the authorship of work in any student's portfolio. On reliability, all four methods have error and consistency problem. It will take a substantial amount of effort for faculty to agree on a scoring matrix and even more time in training and practice to evaluate and analyze the portfolios consistency and without error. Motivation should be fine as long as portfolios are embedded in a course for grade. (All portfolios should be graded, but a smaller random sample, or stratified samples, should receive additional analysis and evaluation for learning outcomes reporting purposes.) On balance, all of the methods are equally balanced between validity and reliability. On cost, the portfolio method has a high opportunity cost as development, evaluation and analysis will require a very substantial commitment of faculty time and the task falls outside of their current duties.

Conclusion and Recommendations

Criteria matrices should be jointly prepared, in an integrative fashion, by members of a faculty committee. Each matrix must be individualized to its institution. Faculty at each institution will have different judgments about assessment instruments. One institution may score the AIC exam highly because of its reliability and low cost. Another institution may attach low scores to the same AIC exam because validity is highly important: the faculty may be concerned that the AIC exam imposes strictures on their curriculum

Program outcomes assessment is not about the assessment of skills or the performance of any one student. Rather, it is collective measure of the quality of a construction management program. We should not lose sight of this. The purpose of program outcomes assessment is to provide a systematic method of program improvement. Program improvements arise from evaluating the learning outcomes for each assessment method against its goals and educational objectives to determine if those goals and educational objectives were achieved and if there is a validated need for improvement in any area of the curricula.

Process improvements arise from evaluating the learning outcomes for each assessment method against its goals and educational objectives to determine if the assessment method is valid. There should be two assessment methods intersecting with each educational objective. By having two intersecting methods, evaluators are better able to assess the validity and reliability of test results. On those evaluations, recommendations for changes to the process may be made

References

Allen, Mary; Noel, Richard, C.; Rienzi, Beth, M.; & McMillin, Daniel, J. (2002). *Outcomes assessment handbook.* California State University, Institute for Teaching & Learning. Long Beach, CA

Association of American Colleges and Universities & the Council for Higher Education Accreditation (2008). New Leadership for Student Learning and Accountability: A Statement of Principles, Commitments to Action. Retrieved May 25, 2012, http://www.aacu.org/resources /assessment/documents/New_Leadership_Statement.pdf

Banta, Trudy W. (2001). Assessing competence in higher education. In Palomba, Catherine A. and Banta, Trudy W. (Eds.) *Assessing student competence in accredited disciplines: pioneering approaches to assessment in higher education.* (pp.1-12). Sterling, Virginia : Stylus

Banta, Trudy W.; Lund, Jon B.; Black, Karen E.; and Oblander, Frances W. (1996). Assessment in practice: putting principles to work on college campuses. San Francisco : Jossey-Bass

Eder, Douglas J. (2001). Accredited programs and authentic assessment. In Palomba, Catherine A. and Banta, Trudy W. (Eds.) *Assessing student competence in accredited disciplines: pioneering approaches to assessment in higher education.* (pp.199-216) Sterling, Virginia : Stylus

Marchese, T.J. (1987). Third down, ten years to go. *AAHE* Bulletin. 40. 3-8

Palomba, Catherine A. and Banta, Trudy W. (1999). *Assessment essentials: planning, implementing, and improving assessment in higher education.* San Francisco : Wiley; Jossey-Bass

APPENDIX A
Terminology Matrix

CAL POLY'S TERMINOLOGY		ACCE TERMINOLOGY	
Assessment	"Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development." (Palomba & Banta, 1999) "Assessment is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance." (Banta, 2001)	Assessment	process used to identify, collect, and prepare data to evaluate the achievement of learning outcomes and educational (degree) program objectives.
Evaluation	The use of assessment findings (evidence/data) to judge program effectiveness; used as a basis for making decisions about program changes or improvement.	Evaluation	Process of interpreting the meaning of the data accumulated through assessment practices. Evaluation determines the extent to which learning outcomes or educational (degree) program objectives are being achieved.
Goals	Goals are the general aims or purposes of a program and its curriculum. Effective goals are **broadly stated**, meaningful, achievable and assessable. Goals provide a framework for determining the more specific educational objectives of a program, and should be **consistent with program** and institutional **mission.**	Educational (Degree) Program Objectives	These are **statements** describing educational (degree) program accomplishments **in support of its mission**.
Educational Objectives	These include the **knowledge, skills**,	Learning Outcomes	These are statements that describe the **skills and knowledge that**

	abilities, capacities, attitudes or dispositions **students are expected to acquire as a result of completing your academic program**. Objectives are sometimes treated as synonymous with outcomes, though outcomes are usually more detailed, behavioral in nature, and stated in precise operational terms.		**students are expected to know by the time of graduation** and that support the educational (degree) program objectives.
Learning Outcomes	These are operational statements describing specific student behaviors that evidence the acquisition of desired knowledge, skills, abilities, capacities, attitudes or dispositions. Learning outcomes can be usefully thought of as behavioral **criteria** for determining whether students are **achieving** the educational objectives of a program, and, ultimately, whether overall program goals are being successfully met. Outcomes are sometimes treated as synonymous with objectives, though objectives are usually more general statements of what students are expected to achieve in an academic program.	Performance **Criteria**	Measurable **achievements** identifying the performance(s) required to meet the learning outcome.