
ASSESSMENT TERMINOLOGY

1. DIRECT VERSUS INDIRECT

Direct measures of student learning *intrinsically* demonstrate mastery of a particular skill or competency. Winning a debate competition provides direct evidence of effective oral communication; publishing a research article in a peer-reviewed journal demonstrates mastery of research and writing skills; obtaining a job after graduation as an assistant editor in a publishing house demonstrates mastery of writing skills.

Indirect measures of student learning are perceptions, behaviors, or attitudes that should, in theory, be evident in proportion to the extent that mastery of a particular learning outcome has been achieved. But they do not *intrinsically* demonstrate mastery. A student self-report on a questionnaire that they have "very much" improved their skills in effective writing; alumni saying that their college years were responsible for enhancing their quantitative reasoning skills; a transcript indicating completion of multiple lab science courses may demonstrate mastery of research methodology.

In reality these are a continuum, not a dichotomy: all measures of mastery involve a subjective element, a "leap of faith" that the test site provides and appropriate context to assess mastery. In educational domains, direct measurement is usually an unattainable goal that is approximated to varying degrees.

2. TRIANGULATION

Because all measures of student learning are flawed, the use of multiple methods for assessing student mastery is advisable, so as to better "triangulate" on the truth. The term triangulation comes from seismology: seismographs can assess *distance* from an earthquake epicenter, but not *direction* to the epicenter; thus, one detector can substantially narrow the range of possible epicenter locations, but only by using distance estimates from at least *three* geographically separated detectors can the true location of the epicenter be determined. Student self-report, course taking behavior, focus groups with graduating seniors, measures of time-on-task in relation to a particular learning outcome are all (individually) imperfect measures, but taken together, if they all suggest the same interpretation, confidence in that interpretation is greatly increased.

3. QUANTITATIVE VERSUS QUALITATIVE

Quantitative assessment methodologies utilize tools to measure student learning numerically, so that results can be manipulated statistically. Qualitative research uses texts – samples of student work, interview or focus group transcripts – as the raw "data" for assessing student learning. Each approach has its strengths and weaknesses, and the strongest assessment plans employ both (see above). Quantitative techniques are often best for large groups or programs, and for testing competing hypotheses about student learning, though in assessment work they can produce a false sense of precision; qualitative research can provide richer content, particularly when assessing outcomes in areas that are more holistic in nature (e.g., understanding diverse cultures, ethical reasoning) but, though it is often assumed to be "easier", it is just as difficult for non-specialists to do well, and is often more suited to *generating* good hypotheses about student learning than it is with *testing* them or demonstrating that learning has occurred.

4. EMBEDDED ASSESSMENT

Assessment techniques that are integrated with coursework or other learning experiences, but are designed in such a way that they yield direct evidence of mastery of learning outcomes. A senior thesis project, evaluated

according to a rubric that is aligned with departmental learning goals, can provide direct evidence of student mastery of those goals; in some contexts simply being more explicit about grading in a manner that is aligned with departmental or course learning goals can be considered embedded assessment; so-called "signature assignments" such as externally evaluated exhibitions of student work, service learning experiences, and the like are examples of "embedded assessment".

5. SAMPLING

When assessing large programs (many students) or numerous examples of student work, *sampling* – selecting and evaluating a subset of students or work samples – can reduce the assessment workload. The key is that the sampling must be done systematically (usually with the guidance of a research methodologist), to ensure that the sample being evaluated truly represents the larger group to which conclusions will be generalized.

6. PORTFOLIOS

Portfolios of student work can provide direct evidence of student learning, particularly if they are organized in a way that permits assessment of improvement in terms of departmental learning goals over time. In order to be considered "assessment", however, academic departments must also systematically evaluate the content of student portfolios against a set of departmental learning goals, and use the results of those evaluations to inform program changes and improvements.

7. VALUE-ADDED VERSUS STANDARDS-BASED

This is an institutional as well as a departmental issue, and it refers to whether student learning is assessed at the student level by comparing mastery at program completion with mastery at program entry (value-added), or relative to a standard at program completion only (standards-based). The usual answer is, "both".

8. BENCHMARKS

Performance is always assessed relative to a standard, whether is internal or external (see above). While it is probably not worth obsessing over *precisely* where to set the benchmark for "success", since the primary goal is *continuous improvement*, departments and institutions do vary here in ways that have important implications for curricular design and resource allocation. Using an "exceeds minimum standards" benchmark, for example (as opposed to value-added), may prompt a department to focus more resources on students with below-average aptitude; using "graduate school admissions" as the benchmark for success, conversely, might encourage a department to dedicate relatively more mentoring and curricular resources to the preparation of "star" students; finally, a benchmark that emphasizes having the greatest *proportion* of graduates exceed a given performance benchmark may encourage a focus on the "middle 50%" of students in terms of aptitude, as the "stars" would exceed the threshold anyway, while lower-aptitude students might require a disproportionate amount of departmental resources on a per-student basis.

9. LEVELS OF ANALYSIS

Departmental learning goals exist within a broader framework of institutional learning goals, while at the same time they subsume course-specific learning goals. Ensuring that learning goals across these three levels of analysis are aligned is an important and sometimes difficult task.

10. COST TO BENEFIT RATIO

A favorable cost to benefit ratio is what defines "good" assessment. Focusing on assessment *quality* (high benefits) without realistically considering high costs (time, money, opportunity costs) required to produce it is no more likely to produce meaningful program improvements than is lower quality, low-cost (e.g., self-report surveys, enrollment counts) assessment work if those costs are unsustainable.

"Good enough" assessment is indeed often good enough in that it can provide accurate guidance for program improvements, particularly when multiple methods are used (see #2). Assessment work should go beyond the anecdote, the "gut sense", and conventional wisdom, but peer-reviewed scholarly research is in most circumstances not the model for assessment work with a favorable cost to benefit ratio. The perfect should not be the enemy of the good.

The cost / benefit ratio also relates to the degree of standardization versus customization in an institution's overall assessment processes. Quantitative data allow for economies of scale in reporting and analysis across academic departments, but only if they are willing to accept some degree of standardization in those reports. For example, it may take x hours to produce a quantitative data analysis in support of assessment for one department, but it may take just 3x hours to develop a template that will generate the same report for all 35 academic departments, though this reduced institutional cost in staff time will involve reducing the benefit somewhat for any single department, as the reports are less customized to their unique needs.