University at Buffalo
Curriculum, Assessment and Teaching Transformation

Glossary of Assessment Terms

LEARNING ASSESSMENT

A systematic process for gathering and processing information to describe the achievement of students with regard to identified learning outcomes (ILOs, PLOs, CLOs). This information helps program faculty understand the effectiveness of course and/or program delivery and guides necessary improvement in instruction, course, and program design.

Assessment for Accountability

Assessment of some unit (could be a department, program or entire institution) to satisfy stakeholders external to the unit itself. Results are often compared across units. Always summative. Example: to retain state approval, the achievement of a 90 percent pass rate or better on teacher certification tests by graduates of a school of education. (Leskes, 2018)

Assessment for Improvement

Assessment that feeds directly, and often immediately, back into revising the course, program or institution to improve student learning results. Can be formative or summative (see "formative assessment" for an example). (Leskes, 2018)

Assessment of Individuals

Uses the individual student, and his/her learning, as the level of analysis. Can be quantitative or qualitative, formative or summative, standards-based or value added, and used for improvement. Would need to be aggregated if used for accountability purposes. Examples: improvement in student knowledge of a subject during a single course; improved ability of a student to build cogent arguments over the course of an undergraduate career. (Leskes, 2018)

Assessment of Institutions

Uses the institution as the level of analysis. Can be quantitative or qualitative, formative or summative, standards-based or value added, and used for improvement or for accountability. Ideally institution-wide goals and objectives would serve as a basis for the assessment. Example: how well students across the institution can work in multi-cultural teams as sophomores and seniors. (Leskes, 2018)

Assessment of Programs

Uses the department or program as the level of analysis. Can be quantitative or qualitative, formative or summative, standards-based or value added, and used for improvement or for accountability. Ideally program goals and objectives would serve as a basis for the assessment. Example: how sophisticated a close reading of texts senior English majors can accomplish (if used to determine value added, would be compared to the ability of newly declared majors). (Leskes, 2018)

Authentic

Refers to assessment tasks that elicit demonstrations of knowledge and skills in ways that they are applied in the "real world." An "authentic assessment" task is also engaging to students and reflects the best current thinking in instructional activities. Thus, teaching to the task is desirable. (McTighe & Arter, 2018)

Benchmark

A standard or point of reference against which data is compared. Synonymous with criterion or norm.

Bias and Distortion

Factors, unrelated to the skill being assessed, that interfere with a valid inference regarding a student's true ability. For example, too much reading on a

mathematics test might result in a distorted vision of a student's mastery of
mathematics content. (McTighe & Arter, 2018)

Constructed Response Assessment

Assessment questions that require a student to produce a response rather than
selected it from a list. For example, essays, reports, oral presentations, reading
fluency, open-ended mathematics problems, etc. (McTighe & Arter, 2018)

Content Standards

Goal statements identifying the knowledge, skills, and dispositions to be developed
through instruction. (McTighe & Arter, 2018)

Criteria

Guidelines, rules, or principles by which student responses, products, or
performances are judged. (McTighe & Arter, 2018). Generic, or general, criteria is
that hwich can be used to score performance on a large number of related tasks;
e.g., a writing rubric that can be used to score critical thinking regardless of the
specific content. (McTighe & Arter, 2018)

Criterion-Referenced

An approach for describing a student's performance according to established
criteria; e.g., she typed 55 words per minute without errors. (McTighe & Arter,
2018)

Curriculum Mapping

A curriculum is a process that uses indexing to produce a diagram of a curriculum
learning outcomes. It may include multiple levels of learning outcomes,
demonstrate redundancies, gaps, and misalignments in content, measures of
assessment, and/or data collected.

Direct Assessment of Learning

Gathers evidence, based on student performance, which demonstrates the learning itself. Can be value added, related to standards, qualitative or quantitative, embedded or not, using local or external criteria. Examples: most classroom testing for grades is direct assessment (in this instance within the confines of a course), as is the evaluation of a research paper in terms of the discriminating use of sources. The latter example could assess learning accomplished within a single course or, if part of a senior requirement, could also assess cumulative learning. (Leskes, 2018)

Dispositions

Refers to the affective dimensions of students in school; e.g., motivation to learn, attitude toward school, academic self-concept, flexibility, persistence, and locus of control. Some scoring guides are designed to assess dispositions. These provide specific, observable indicators of the disposition being assessed. (McTighe & Arter, 2018)

Embedded Assessment

A means of gathering information about student learning that is built into and a natural part of the teaching-learning process. Often uses for assessment purposes classroom assignments that are evaluated to assign students a grade. Can assess individual student performance or aggregate the information to provide information about the course or program; can be formative or summative, quantitative or qualitative. Example: as part of a course, expecting each senior to complete a research paper that is graded for content and style, but is also assessed for advanced ability to locate and evaluate Web-based information (as part of a college-wide outcome to demonstrate information literacy). (Leskes, 2018)

Evaluation

Judgment regarding the quality, value, or worth of assessment results; e.g., "the information we collected indicates that students are reading as well as we would like." Evaluations are usually based on multiple sources of information. (McTighe & Arter, 2018)

External Assessment

Use of criteria (rubric) or an instrument developed by an individual or organization external to the one being assessed. Usually summative, quantitative, and often high-stakes. Example: GRE exams. (Leskes, 2018)

Formative Assessment

The gathering of information about student learning-during the progression of a course or program and usually repeatedly-to improve the learning of those students. Example: reading the first lab reports of a class to assess whether some or all students in the group need a lesson on how to make them succinct and informative. (Leskes, 2018)

"High Stakes" Use of Assessment

the decision to use the results of assessment to set a hurdle that needs to be cleared for completing a program of study, receiving certification, or moving to the next level. Most often the assessment so used is externally developed, based on set standards, carried out in a secure testing situation, and administered at a single point in time. Examples: at the secondary school level, statewide exams required for graduation; in postgraduate education, the bar exam. (Leskes, 2018)

Holistic Scoring

A scoring procedure yielding a single score based upon an overall impression of a product or performance. (McTighe & Arter, 2018)

Indirect Assessment of Learning

Gathers reflection about the learning or secondary evidence of its existence. Example: a student survey about whether a course or program helped develop a greater sensitivity to issues of diversity. (Leskes, 2018)

Learning Outcome (as in "learning target" or objective)

Statements of what we want students to know and be able to do. (McTighe & Arter, 2018). Can be at the institution (ILO), program (PLO), or course (CLO) levels.

Local Assessment

Means and methods that are developed by an institution's faculty based on their teaching approaches, students, and learning goals. Can fall into any of the definitions here except "external assessment," for which is it an antonym. Example: one college's use of nursing students' writing about the "universal precautions" at multiple points in their undergraduate program as an assessment of the development of writing competence. (Leskes, 2018)

Norm-Referenced

Describing a student's performance by comparison to other, similar students; e.g., she typed better than 80 percent of her classmates. (McTighe & Arter, 2018)

Performance Assessment

An assessment activity that requires students to construct a response, create a product, or perform a demonstration. Since performance assessments generally do not yield a single correct answer or solution method, evaluations of student products or performances are based on judgments guided by criteria. (McTighe & Arter, 2018)

Performance List

A scoring guide consisting of designated criteria, but without descriptive details. For example, a performance list for writing might contain six features—ideas, organization, voice, word choice, sentence fluency, and conventions. Unlike a rubric, a performance list merely provides a set of features without defining the terms or providing indicators of quality. (McTighe & Arter, 2018)

Performance Standard

An established level of achievement, quality of performance, or degree of proficiency. Performance standards specify how well students are expected to achieve or perform. (McTighe & Arter, 2018)

Primary Trait(s) Scoring

A scoring procedure by which products or performances are evaluated by limiting attention to a single criterion or a few selected criteria. These criteria are based upon the trait or traits determined to be essential for a successful performance on a given task. For example, a note to a principal urging a change in a school rule might have persuasiveness as the primary trait. Scorers would attend only to that trait. (McTighe & Arter, 2018)

Qualitative Assessment

Collects data that does not lend itself to quantitative methods but rather to interpretive criteria. (Leskes, 2018)

Quantitative Assessment

Collects data that can be analyzed using quantitative methods. (Leskes, 2018)

Reliability

The degree to which the results of an assessment are dependable and yield consistent results across raters (inter-rater reliability), over time (test-retest reliability), or across different versions of the same test (internal consistency or inter-form reliability). Technically, this is a statistical term that defines the extent to which errors of measurement are absent from an assessment instrument. (McTighe & Arter, 2018)

Rubric

A set of general criteria used to evaluate a student's performance in a given outcome area. Rubrics consist of a fixed measurement scale (e.g., 4-point) and a list of criteria that describe the characteristics of products or performances for each score point. Rubrics are frequently accompanied by examples (anchors) of products or performances to illustrate the various score points on the scale. (McTighe & Arter, 2018)

Scoring Guide

A generic term for a criterion-based tool used in judging performance. In this book, we are using scoring guide synonymously with criteria and rubric. (McTighe & Arter, 2018)

Selected Response Assessments

Assessment questions that ask students to select an answer from a provided list. For example, multiple-choice, matching, and true-false. (McTighe & Arter, 2018)

Standardized

A set of consistent procedures for constructing, administering, and scoring an assessment. The goal of standardization is to ensure that all students are assessed under uniform conditions so that interpretation of their performance is

comparable and not influenced by differing conditions. Both norm-referenced and criterion-referenced assessments can be standardized. (McTighe & Arter, 2018)

Standards

Sets a level of accomplishment all students are expected to meet or exceed. Standards do not necessarily imply high quality learning; sometimes the level is a lowest common denominator. Nor do they imply complete standardization in a program; a common minimum level could be achieved by multiple pathways and demonstrated in various ways. Examples: carrying on a conversation about daily activities in a foreign language using correct grammar and comprehensible pronunciation; achieving a certain score on a standardized test. (Leskes, 2018)

Summative Assessment

The gathering of information at the conclusion of a course, program, or undergraduate career to improve learning or to meet accountability demands. When used for improvement, impacts the next cohort of students taking the course or program. Examples: examining student final exams in a course to see if certain specific areas of the curriculum were understood less well than others; analyzing senior projects for the ability to integrate across disciplines. (Leskes, 2018)

Task

An assessment exercise involving students in producing a response, product or performance; e.g., solving a mathematics problem, conducting a laboratory in science, or writing a paper. Since tasks are associated with performance assessments, many are complex and open- ended, requiring responses to a challenging question or problem. However, there can be simple performance tasks, such as reading aloud to measure reading rate. Tasks don't have to be exclusively used as stand-alone activities that occur at the end of instruction; teachers can observe students working on tasks during the course of regular instruction in order to provide on- going feedback. (McTighe & Arter, 2018)

Task-Specific Criteria

A scoring guide or rubric that can only be used with a single exercise or task. Since the language is specific to a particular task (e.g., to get a "4" the response must have "accurate ranking of children on each event, citing Zaia as overall winner"), task-specific guides cannot be applied to any other task without modification. (McTighe & Arter, 2018)

Test

A set of questions or situations designed to permit an inference about what an examinee knows or can do in an area of interest. (McTighe & Arter, 2018)

Validity

An indication of how well an assessment measures what it was intended to measure; e.g., does a test of laboratory skills really assess laboratory skills or does it assess ability to read and follow instructions? Technically, validity indicates the degree of accuracy of predictions or inferences based upon an assessment measure. (McTighe & Arter, 2018)

Value added

The increase in learning that occurs during a course, program, or undergraduate education. Can either focus on the individual student (how much better a student can write, for example, at the end than at the beginning) or on a cohort of students (whether senior papers demonstrate more sophisticated writing skills-in the aggregate-than freshmen papers). Requires a baseline measurement for comparison. (Leskes, 2018)

References

Leskes, A. (2018). Beyond confusion: An assessment glossary. Peer Review, 4(2/3). Retrieved from https://www.aacu.org/publications-research/periodicals/beyond-confusion-assessment-glossary

McTighe, J., & Arter, J. (2018). Glossary of assessment terms. Jay McTighe & Associates Consulting. Retrieved from https://www.jaymctighe.com/downloads/Glossary-of-Assessment-Terms.pdf