

Assessment Terminology

Prepared for: Assessing Learning. Should the Tail Wag the Dog? Assessing Learning Conference -- 9/28-30, 1995 <http://www.newhorizons.org/strategies/assess/terminology.htm>

Accountability

The demand by a community (public officials, employers, and taxpayers) for school officials to prove that money invested in education has led to measurable learning. "Accountability testing" is an attempt to sample what students have learned, or how well teachers have taught, and/or the effectiveness of a school's principal's performance as an instructional leader. School budgets and personnel promotions, compensation, and awards may be affected. Most school districts make this kind of assessment public; it can affect policy and public perception of the effectiveness of taxpayer-supported schools and be the basis for comparison among schools. Accountability is often viewed as an important factor in education reform. An assessment system connected to accountability can help identify the needs of schools so that resources can be equitably distributed. In this context, accountability assessment can include such indicators as equity, competency of teaching staff, physical infrastructure, curriculum, class size, instructional methods, existence of tracking, number of higher cost students, dropout rates, and parental involvement as well as student test scores. It has been suggested that test scores analyzed in a disaggregated format can help identify instructional problems and point to potential solutions.

Achievement Test

A standardized test designed to efficiently measure the amount of knowledge and/or skill a person has acquired, usually as a result of classroom instruction. Such testing produces a statistical profile used as a measurement to evaluate student learning in comparison with a standard or norm.

Action Research

School and classroom-based studies initiated and conducted by teachers and other school staff. Action research involves teachers, aides, principals, and other school staff as researchers who systematically reflect on their teaching or other work and collect data that will answer their questions. It offers staff an opportunity to explore issues of interest to them in an effort to improve classroom instruction and educational effectiveness. (Source: Bennett, C.K. "Promoting teacher reflection through action research: What do teachers think?" *Journal of Staff Development*, 1994, 15, 34-38.)

Affective

Outcomes of education involving feelings more than understanding; likes, pleasures ideals, dislikes annoyances, values.

Alternative Assessment

Many educators prefer the description "assessment alternatives" to describe alternatives to traditional, standardized, norm- or criterion-referenced traditional paper and pencil testing. An alternative assessment might require students to answer an open-ended question, work out a solution to a problem, perform a demonstration of a skill, or in some way produce work rather than select an answer from choices on a sheet of paper. Portfolios and instructor observation of students are also alternative forms of assessment.

Analytic Scoring

A type of rubric scoring that separates the whole into categories of criteria that are examined one at a time. Student writing, for example, might be scored on the basis of grammar, organization, and clarity of ideas. Useful as a diagnostic tool. An analytic scale is useful when there are several dimensions on which the piece of work will be evaluated. (See Rubric.)

Aptitude Test

A test intended to measure the test-taker's innate ability to learn, given before receiving instruction.

Assessment

The Latin root *assidere* means to sit beside. In an educational context, the process of observing learning; describing, collecting, recording, scoring, and interpreting information about a student's or one's own learning. At its most useful, assessment is an episode in the learning process; part of reflection and autobiographical understanding of progress. Traditionally, student assessments are used to determine placement, promotion, graduation, or retention.

In the context of institutional accountability, assessments are undertaken to determine the principal's performance, effectiveness of schools, etc. In the context of school reform, assessment is an essential tool for evaluating the effectiveness of changes in the teaching-learning process.

Assessment Literacy

The possession of knowledge about the basic principals of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment. Increasingly, familiarity with alternatives to traditional measurements of learning.

Assessment Task

An illustrative task or performance opportunity that closely targets defined instructional aims, allowing students to demonstrate their progress and capabilities.

Authentic Assessment

Evaluating by asking for the behavior the learning is intended to produce. The concept of model, practice, feedback in which students know what excellent performance is and are guided to practice an entire concept rather than bits and pieces in preparation for eventual understanding. A variety of techniques can be employed in authentic assessment.

The goal of authentic assessment is to gather evidence that students can use knowledge effectively and be able to critique their own efforts. Authentic tests can be viewed as "assessments of enablement," in Robert Glaser's words, ideally mirroring and measuring student performance in a "real-world" context. Tasks used in authentic assessment are meaningful and valuable, and are part of the learning process.

Authentic assessment can take place at any point in the learning process. Authentic assessment implies that tests are central experiences in the learning process, and that assessment takes place repeatedly. Patterns of success and failure are observed as learners use knowledge and skills in slightly ambiguous situations that allow the assessor to observe the student applying knowledge and skills in new situations over time.

Benchmark

Student performance standards (the level(s) of student competence in a content area.)

An actual measurement of group performance against an established standard at defined points along the path toward the standard. Subsequent measurements of group performance use the benchmarks to measure progress toward achievement.

Examples of student achievement that illustrate points on a performance scale, used as exemplars. (See Descriptor, Cohort.)

Cohort

A group whose progress is followed by means of measurements at different points in time.

Competency Test

A test intended to establish that a student has met established minimum standards of skills and knowledge and is thus eligible for promotion, graduation, certification, or other official acknowledgement of achievement.

Concept

An abstract, general notion -- a heading that characterizes a set of behaviors and beliefs.

Criterion Referenced Tests

A test in which the results can be used to determine a student's progress toward mastery of a content area. Performance is compared to an expected level of mastery in a content area rather than to other students' scores. Such tests usually include questions based on what the student was taught and are designed to measure the student's mastery of designated objectives of an instructional program. The "criterion" is the standard of performance established as the passing score for the test. Scores have meaning in terms of what the student knows or can do, rather than how the test-taker compares to a reference or norm group. Criterion referenced tests can have norms, but comparison to a norm is not the purpose of the assessment.

Criterion referenced tests have also been used to provide information for program evaluation, especially to track the success or progress of schools and student populations that have been involved in change or that are at risk of inequity. In this case, the tests are not used to compare teachers, teams or buildings within a district but rather to give feedback on progress of groups and individuals.

Curriculum Alignment

The degree to which a curriculum's scope and sequence matches a testing program's evaluation measures, thus ensuring that teachers will use successful completion of the test as a goal of classroom instruction.

Curriculum-embedded or Learning-embedded Assessment

Assessment that occurs simultaneously with learning such as projects, portfolios and "exhibitions." Occurs in the classroom setting, and, if properly designed, students should not be able to tell whether they are being taught or assessed. Tasks or tests are developed from the curriculum or instructional materials.

Cut Score

Score used to determine the minimum performance level needed to pass a competency test.

Descriptor

A set of signs used as a scale against which a performance or product is placed in an evaluation. An example from Grant Wiggins' Glossary of Useful Terms Related to Authentic and Performance Assessments is taken from "the CAP writing test where a 5 out of a possible 6 is described: 'The student describes the problem adequately and argues convincingly for at least one solution...without the continual reader awareness of the writer of a 6.'"

Descriptors allow assessment to include clear guidelines for what is and is not valued in student work. Wiggins

adds that "[t]he word 'descriptor' reminds us that justifiable value judgments are made by know how to empirically describe the traits of work we do and do not value." (emphasis his.)

Dimension

Aspects or categories in which performance in a domain or subject area will be judged. Separate descriptors or scoring methods may apply to each dimension of the student's performance assessment.

Essay Test

A test that requires students to answer questions in writing. Responses can be brief or extensive. Tests for recall, ability to apply knowledge of a subject to questions about the subject, rather than ability to choose the least incorrect answer from a menu of options.

Evaluation

Both qualitative and quantitative descriptions of pupil behavior plus value judgments concerning the desirability of that behavior. Using collected information (assessments) to make informed decisions about continued instruction, programs, activities. Exemplar Model of excellence. (See Benchmark, Norm, Rubric, Standard.)

Formative Assessment

Observations which allow one to determine the degree to which students know or are able to do a given learning task, and which identifies the part of the task that the student does not know or is unable to do. Outcomes suggest future steps for teaching and learning. (See Summative Assessment.)

Grade Equivalent

A score that describes student performance in terms of the statistical performance of an average student at a given grade level. A grade equivalent score of 5.5, for example, might indicate that the student's score is what could be expected of a average student doing average work in the fifth month of the fifth grade. This score allows for a theoretical or approximate comparison across grades. It ranges from September of the kindergarten year (K.O.) to June of the senior year in high school (12.9) Useful as a ranking score, grade equivalents are only a theoretical or approximate comparison across grades. In this case, it may not indicate what the student would actually score on a test given to a midyear fifth grade class.

High Stakes Testing

Any testing program whose results have important consequences for students, teachers, schools, and/or districts. Such stakes may include promotion, certification, graduation, or denial/approval of services and opportunity. High stakes testing can corrupt the evaluation process when pressure to produce rising test scores results in "teaching to the test" or making tests less complex.

Holistic Method

In assessment, assigning a single score based on an overall assessment of performance rather than by scoring or analyzing dimensions individually. The product is considered to be more than the sum of its parts and so the quality of a final product or performance is evaluated rather than the process or dimension of performance. A holistic scoring rubric might combine a number of elements on a single scale. Focused holistic scoring may be used to evaluate a limited portion of a learner's performance.

I. Q. Tests

The first of the standardized norm-referenced tests, developed during the nineteenth century. Traditional psychologists believe that neurological and genetic factors underlie "intelligence" and that scoring the performance of certain intellectual tasks can provide assessors with a measurement of general intelligence. There is a substantial body of research that suggests that I.Q. tests measure only certain analytical skills, missing many areas of human endeavor considered to be intelligent behavior. I. Q is considered by some to be fixed or static; whereas an increasing number of researchers are finding that intelligence is an ongoing process that continues to change throughout life.

Item Analysis

Analyzing each item on a test to determine the proportions of students selecting each answer. Can be used to evaluate student strengths and weaknesses; may point to problems with the test's validity and to possible bias.

Journals

Students' personal records and reactions to various aspects of learning and developing ideas. A reflective process often found to consolidate and enhance learning.

Mean

One of several ways of representing a group with a single, typical score. It is figured by adding up all the individual scores in a group and dividing them by the number of people in the group. Can be affected by extremely low or high scores.

Measurement

Quantitative description of student learning and qualitative description of student attitude.

Median

The point on a scale that divides a group into two equal subgroups. Another way to represent a group's scores with a single, typical score. The median is not affected by low or high scores as is the mean. (See Norm.)

Metacognition

The knowledge of one's own thinking processes and strategies, and the ability to consciously reflect and act on the knowledge of cognition to modify those processes and strategies.

Multidimensional Assessment

Assessment that gathers information about a broad spectrum of abilities and skills (as in Howard Gardner's theory of Multiple Intelligences).

Multiple Choice Tests

A test in which students are presented with a question or an incomplete sentence or idea. The students are expected to choose the correct or best answer/completion from a menu of alternatives.

Norm

A distribution of scores obtained from a norm group. The norm is the midpoint (or median) of scores or performance of the students in that group. Fifty percent will score above and fifty percent below the norm.

Norm Group

A random group of students selected by a test developer to take a test to provide a range of scores and establish the percentiles of performance for use in establishing scoring standards.

Norm Referenced Tests

A test in which a student or a group's performance is compared to that of a norm group. The student or group scores will not fall evenly on either side of the median established by the original test takers. The results are relative to the performance of an external group and are designed to be compared with the norm group providing a performance standard. Often used to measure and compare students, schools, districts, and states on the basis of norm-established scales of achievement.

Normal Curve Equivalent

A score that ranges from 1-99, often used by testers to manipulate data arithmetically. Used to compare different tests for the same student or group of students and between different students on the same test. An NCE is a normalized test score with a mean of 50 and a standard deviation of 21.06. NCEs should be used instead of percentiles for comparative purposes. Required by many categorical funding agencies, e.g., Chapter I or Title I.

Objective Test

A test for which the scoring procedure is completely specified enabling agreement among different scorers. A correct-answer test.

On-Demand Assessment

An assessment process that takes place as a scheduled event outside the normal routine. An attempt to summarize what students have learned that is not embedded in classroom activity.

Outcome

An operationally defined educational goal, usually a culminating activity, product, or performance that can be measured.

Percentile

A ranking scale ranging from a low of 1 to a high of 99 with 50 as the median score. A percentile rank indicates the percentage of a reference or norm group obtaining scores equal to or less than the test-taker's score. A percentile score does not refer to the percentage of questions answered correctly, it indicates the test-taker's standing relative to the norm group standard.

Performance-Based Assessment

Direct, systematic observation and rating of student performance of an educational objective, often an ongoing observation over a period of time, and typically involving the creation of products. The assessment may be a continuing interaction between teacher and student and should ideally be part of the learning process. The

assessment should be a real-world performance with relevance to the student and learning community. Assessment of the performance is done using a rubric, or analytic scoring guide to aid in objectivity. Performance-based assessment is a test of the ability to apply knowledge in a real-life setting. Performance of exemplary tasks in the demonstration of intellectual ability.

Evaluation of the product of a learning experience can also be used to evaluate the effectiveness of teaching methods.

Stiggins (1992) defines performance-based assessment as the use of performance criteria to determine the degree to which a student has met an achievement target. Important elements of performance-based assessment include clear goals or performance criteria clearly articulated and communicated to the learner; the establishment of a sound sampling that clearly envisions the scope of an achievement target and the type of learning that is involved (use of problem-solving skills, knowledge acquisition, etc.) Attention to extraneous interference (cultural biases, language barriers, testing environment, tester biases) and establishment of a clear purpose for the data collected during the assessment before the assessment is undertaken, keeping in mind the needs of the groups involved (teachers, students, parents, etc.) (from an article by Richard J. Stiggins, "The Key to Unlocking High-Quality Performance Assessments. "Assessment: How Do We Know What They Know? ASCD, 1992.

Performance Criteria

The standards by which student performance is evaluated. Performance criteria help assessors maintain objectivity and provide students with important information about expectations, giving them a target or goal to strive for.

Portfolio

A systematic and organized collection of a student's work that exhibits to others the direct evidence of a student's efforts, achievements, and progress over a period of time. The collection should involve the student in selection of its contents, and should include information about the performance criteria, the rubric or criteria for judging merit, and evidence of student self-reflection or evaluation. It should include representative work, providing a documentation of the learner's performance and a basis for evaluation of the student's progress. Portfolios may include a variety of demonstrations of learning and have been gathered in the form of a physical collection of materials, videos, CD-ROMs, reflective journals, etc.

Portfolio Assessment

Portfolios may be assessed in a variety of ways. Each piece may be individually scored, or the portfolio might be assessed merely for the presence of required pieces, or a holistic scoring process might be used and an evaluation made on the basis of an overall impression of the student's collected work. It is common that assessors work together to establish consensus of standards or to ensure greater reliability in evaluation of student work. Established criteria are often used by reviewers and students involved in the process of evaluating progress and achievement of objectives.

Primary Trait Method

A type of rubric scoring constructed to assess a specific trait, skill, behavior, or format, or the evaluation of the primary impact of a learning process on a designated audience.

Process

A generalizable method of doing something, generally involving steps or operations that are usually ordered and/or interdependent. Process can be evaluated as part of an assessment, as in the example of evaluating a student's performance during pre-writing exercises leading up to the final production of an essay or paper.

Product

The tangible and stable result of a performance or task. An assessment is made of student performance based on evaluation of the product of a demonstration of learning.

Profile

A graphic compilation of the performance of an individual on a series of assessments.

Project

A complex assignment involving more than one type of activity and production. Projects can take a variety of forms, some examples are a mural construction, a shared service project, or other collaborative or individual effort.

Quartile

The breakdown of an aggregate of percentile rankings into four categories: the 0-25th percentile, 26-50th percentile, etc.

Quintile

The breakdown of an aggregate of percentile rankings into five categories: the 0-20th percentile, 21-40th percentile, etc.

Rating Scale

A scale based on descriptive words or phrases that indicate performance levels. Qualities of a performance are described (e.g., advanced, intermediate, novice) in order to designate a level of achievement. The scale may be used with rubrics or descriptions of each level of performance.

Reliability

The measure of consistency for an assessment instrument. The instrument should yield similar results over time with similar populations in similar circumstances.

Rubric

Some of the definitions of rubric are contradictory. In general a rubric is a scoring guide used in subjective assessments. A rubric implies that a rule defining the criteria of an assessment system is followed in evaluation. A rubric can be an explicit description of performance characteristics corresponding to a point on a rating scale. A scoring rubric makes explicit expected qualities of performance on a rating scale or the definition of a single scoring point on a scale.

Sampling

A way to obtain information about a large group by examining a smaller, randomly chosen selection (the sample) of group members. If the sampling is conducted correctly, the results will be representative of the group as a whole. Sampling may also refer to the choice of smaller tasks or processes that will be valid for making inferences about the student's performance in a larger domain. "Matrix sampling" asks different groups to take small segments of a test; the results will reflect the ability of the larger group on a complete range of tasks.

Scale

A classification tool or counting system designed to indicate and measure the degree to which an event or behavior has occurred.

Scale Scores

Scores based on a scale ranging from 001 to 999. Scale scores are useful in comparing performance in one subject area across classes, schools, districts, and other large populations, especially in monitoring change over time.

Score

A rating of performance based on a scale or classification.

Scoring Criteria

Rules for assigning a score or the dimensions of proficiency in performance used to describe a student's response to a task. May include rating scales, checklists, answer keys, and other scoring tools. In a subjective assessment situation, a rubric.

Scoring

A package of guidelines intended for people scoring performance assessments. May include instructions for raters, notes on training raters, rating scales, samples of student work exemplifying various levels of performance.

Self-Assessment

A process in which a student engages in a systematic review of a performance, usually for the purpose of improving future performance. May involve comparison with a standard, established criteria. May involve critiquing one's own work or may be a simple description of the performance. Reflection, self-evaluation, metacognition, are related terms.

Senior Project

Extensive projects planned and carried out during the senior year of high school as the culmination of the secondary school experience, senior projects require higher-level thinking skills, problem-solving, and creative thinking. They are often interdisciplinary, and may require extensive research. Projects culminate in a presentation of the project to a panel of people, usually faculty and community mentors, sometimes students, who evaluate the student's work at the end of the year.

Standardized Test

An objective test that is given and scored in a uniform manner. Standardized tests are carefully constructed and items are selected after trials for appropriateness and difficulty. Tests are issued with a manual giving complete guidelines for administration and scoring. The guidelines attempt to eliminate extraneous interference that might influence test results. Scores are often norm-referenced.

Standards

Agreed upon values used to measure the quality of student performance, instructional methods, curriculum, etc.

Subjective Test

A test in which the impression or opinion of the assessor determines the score or evaluation of performance. A test in which the answers cannot be known or prescribed in advance.

Summative Assessment

Evaluation at the conclusion of a unit or units of instruction or an activity or plan to determine or judge student skills and knowledge or effectiveness of a plan or activity. Outcomes are the culmination of a teaching/learning process for a unit, subject, or year's study. (See Formative Assessment.)

Validity

The test measures the desired performance and appropriate inferences can be drawn from the results. The assessment accurately reflects the learning it was designed to measure.