

What the Words Mean! A Glossary of Assessment Terms

Adaptive Assessment: is an innovative, online form of assessment in which items are presented in a sequence that is dependent on the correctness of the examinee's response to the preceding item. Through this process, each examinee is administered a unique set of items that provides an accurate measure of his or her achievement level.

Alternative Assessment: Any kind of assessment other than traditional norm-referenced or criterion-referenced pencil-and-paper tests.

Anecdotal Records: Informal documenting of student achievement, behavior and learning as it occurs.

Assessment: The practice of collecting information about what learners know and can do.

Aptitude Test: A standardized test intended to measure a student's innate ability to learn, given before receiving instruction.

Criterion-Referenced Test (CRT): A test including questions based on what the student was taught designed to measure how much the student learned from that instruction. Can use multiple choice or essay questions, or combination of both.

Continuum: A visual representation of the stages of learning development in a particular area; specific descriptors provide a framework for assessing individual growth.

Diagnostic or Pre-Assessment: The collection of information about a student's entry-level skills before a unit, course, semester or year.

Disaggregated Data: Data broken down into smaller sub-groups: i.e. gender, ethnic, SES, etc.

Essay Tests: Students are required to write paragraphs or themes as responses to a question. Short essay tests require students to recall, explain, or apply specific information they have learned in their own words.

Evaluation: The process of assigning meaning to the data collected on learners. Evaluation may focus on either what the learner can do or what the learner cannot do.

Feedback: Direct information provided reflecting strengths and needs of a student or building based on an assessment.

Formative Assessments: checkpoints of learning; snapshots; used to inform instruction and essential learning; periodic collection of information to provide feedback concerning progress toward achieving learning goals; provides the information need to adjust teaching and learning while they are happening (inform along the way) Assessment for learning!

Frequency Distribution: A listing of the number of people who obtain each score or fall into each range of scores on a test.

Grades: Symbols (letters, numbers, or words) that represent a value judgment concerning the relative quality of a student's achievement during a specified period of instruction.

Grade Equivalent Scores: The grade level at which a student performs on a standardized test.

Interim Assessments: An assessment that occurs multiple times (typically 3-4 times) throughout the academic year rather than just at the end. Interim assessments usually are administered between formative and summative assessment. **Purposes:** predicting a student's ability to succeed on a large scale assessment, such as the Kansas Summative Assessment; evaluation of a particular education program or pedagogy or diagnosing gaps in student's learning.

Item Analysis: A process used to diagnose particular strengths and weaknesses of students.

Kansas Assessment "All" Students: Data is grouped and reported for all students tested (including gen. ed., gifted, sped)

Kansas Assessment "General Education" Students: Data is grouped and reported for general education students and gifted students; all others are excluded from this data.

Multiple Choice Tests: Direct question or incomplete statement (called the stem) followed by two or more possible answers (called responses), only one of which is to be selected.

Normal Curve Equivalent: Standard score with a range from 1 to 99 with a mean of 50 and a standard deviation of about 21.

Norm: Distribution of scores obtained from the norm group. It is the midpoint of performance of the students in the norm group: 50 percent of them score above the norm and 50 percent below.

Norm Group: A random group of students that originally take a test under the auspices of the test developer to determine the baseline and establish the percentiles of performance.

Norm-Referenced Tests: An assessment instrument that relates the scores of each student to the scores of those in the norm group. A norm-referenced test tells how each student and group of students measure up against the original group that took the test, by distributing the range of scores along a continuum.

Objective Tests: Tests whose items are easily scored and analyzed, given to large numbers of students, take very little time to administer and score, and are free of bias in scoring and in requiring unrelated skills such as writing.

Observation Checklist: A record keeping device for teachers to use to keep track of the degree to which each student has demonstrated a targeted behavior, action, skill or procedure.

Open-ended Questions: Calls for students to answer by writing a statement that may vary in length. Answer may have a specific answer or may be "open-ended" with no pre-determined answer.

Percent Correct: The percentage of items answered correctly. Percent correct scores range from 0 to 100 and are calculated by dividing the number of points accumulated by the total number of score points possible.

Percentile Rank: Range from a low of 1 to a high of 99, with 50 denoting average performance. The percentile rank corresponding to a given score indicates the percentage of a reference group obtaining scores equal to or less than that score. For example, if a student scores at the 65th percentile, it means that he or she performed better on the test than 65 percent of the norm group. Percentile marks are not equal interval scores.

Performance Assessment: Students are required to "perform" a task rather than simply answer questions. It is judged against established criteria. Also called "authentic" assessment because it requires students to demonstrate what they know, can do, or were taught, instead of just selecting an answer written by someone else.

Portfolio: A purposeful collection of student work that tells the story of a student's efforts, progress, or achievement. (Examples: showcase, best work, working portfolio, assessment, and subject area).

Raw Score: The number of test items correctly answered by a student.

Reliability: When a student's performance remains the same on repeated measures.

Rubric: Indicators of different levels of a criterion being used to assess a performance (or scoring guides that use specific written criteria to distinguish among levels of student proficiency on a common task).

Running Records: Used to record systematic notes about students' reading; an on-going assessment; running records can be used with students at any grade level.

Sampling: A way to get information about a large group by examining only a small number of the group (the sample), or by giving all members small segments of the test. When conducted properly, the results are considered highly reliable.

Self-Reflection: Students/teachers are asked to reflect on, make a judgment about, and then report on their own behavior and performance; usually the reports are made in response to specific questions or statements designed to elicit feedback on selected aspects of their experience, behavior, feelings, and so on.

Standard Score: Indication of how far each student is above or below the mean in a way that allows comparison of scores from different tests, regardless of the size of the class or the number of items on the test. Found by subtracting the mean from the student's raw score and dividing by the standard deviation.

Standard of Excellence (building standard): Established in 1994 by the Kansas State Board of Education for Reading, Math and Writing. These are not minimums; these are standards of excellence. New standards were set in 2006.

Standard of Excellence (student standards): Established in 1997 by the Kansas State Board of Education for Reading, Math and Writing. These are not minimums; these are standards of excellence. New standards were set in 2000 based on student performance levels. Five performance levels have been determined: unsatisfactory, basic, satisfactory, proficient and advanced. In 2003 the descriptors for the performance levels changed to: unsatisfactory, basic, proficient, advanced and exemplary. In 2006 new standards were set and changed to: academic warning, basic, meets standard, exceeds standard, and exemplary.

Stanine Score: Standard score with a mean of 5 and a standard deviation of 2.

Standardized Tests: Tests prepared for nationwide use to provide information on students' levels of performance relative to others at their age or grade.

Summative Assessment: The collection of information at the end of an instructional unit to judge the final quality and quantity of student achievement and/or the success of the instructional program.
Assessment of Learning

Teacher Made-Tests: Written or oral assessments of student achievement that are (a) not commercially produced or standardized and (b) designed specifically for the teacher's students.

True-False Test Item: A fact, statement, definition, or principle followed by two responses (true or false).

Quartile: After scores are determined, the distribution may be broken down into four groups: Quartile 1 = 0 – 25, Quartile 2 = 26 – 50, Quartile 3 = 51 – 75 and Quartile 4 = 76 – 99 (100).

Validity: Validity means that the test actually measures what it was designed to measure, all of what it was designed to measure, and nothing but what it was designed to measure.