ASSESSMENT GLOSSARY

Ability: Learning outcomes that are complex, multidimensional (knowledge/skills/attitudes), teachable, and transferable to other areas of life (e.g., critical thinking, communication—oral and written, quantitative reasoning, etc.).

Alternative Assessment: Alternative assessment is any type of assessment in which students create a response to a question. (This is different than assessments in which students choose a response from a list given, such as multiple choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performance assessments, oral presentations, demonstrations, exhibitions, and portfolios.

Anchor Papers: Examples of student performance that serve as a standard against which other papers or performances may be judged are called anchor papers. They are often used as examples of performances at different points on a scoring rubric for a particular grade level. In math problem solving, for example, anchor papers are selected from actual student work that are considered to exemplify the quality of a performance level of “1”, “2”, “3”, and so forth. If used with analytical scoring, there may be anchor papers or benchmark performances for each trait being assessed.

Assessment: Assessment is the process of gathering information to make decisions. In an educational context, assessment is the process of observing learning: describing, collecting, recording, scoring, and interpreting information about a student’s or one’s own learning.

Assessment Literacy: Assessment literacy means having knowledge about the basic principles of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment.

Authentic Assessment: Assessments that involve engaging tasks built around important questions/issues reflecting meaningful contexts found in the particular field of study or in adult life. The tasks involved focus on non-routine, multi-stage (i.e., “real”) problems, generally requiring students to produce some kind of quality product and/or performance. Authentic assessments are usually accompanied by explicitly-defined standards and criteria shared with students (see also performance assessment).

Benchmark: A benchmark is point in time (e.g., 4th grade) which may be used to measure student progress. Benchmarks are designed to help educators organize and make sense of a complex process of interaction between the student, the teacher, and the learning process.

Checklists: Checklists are lists of characteristics or behaviors. Checklists are used to guide evaluation of student performances by noting the presence or absence of any given characteristic or behavior.

Classroom-based Assessment/Evidence: An assessment developed, administered, and scored by a teacher or set of teachers with the purpose of evaluating individual or classroom student performance on a topic is known as classroom-based assessment. Ideally, the results of a classroom assessment are used to inform instruction so that students reach high standards. (In Washington State, a classroom-based component is a major part of the state assessment system.)

Competence: The individual’s demonstrated capacity to perform, i.e., the possession of knowledge, skills and personal characteristics needed to satisfy the special demands or requirements of a particular situation is referred to as competence.

Content Standard: See standard.
ASSESSMENT GLOSSARY

Essay Test: A test that requires students to answer questions in writing is an essay test. Responses can be brief or extensive. The essay test usually measures knowledge as well as the ability to apply knowledge of a subject to questions about the subject.

Evaluation: Evaluation is the process of making judgments based on criteria and evidence.

Exemplar: Models of excellence are known as exemplars.

Holistic Scoring: In assessment, holistic scoring means 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. Holistic scoring criteria might combine a number of elements on a single scale.

Outcome: An operationally defined educational goal, usually a culminating activity, product or performance that can be measured is referred to as an outcome.

Performance-based Assessment: Performance-based assessment refers to systematic observation and rating of student performance of an educational objective. Such assessment is often an ongoing observation over a period of time, and typically requires the student to finish 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 scoring guide or rubric.

Performance Criteria: A description of the characteristics that will be considered when a performance task is judged are called performance criteria. Performance criteria are often defined in a rubric or scoring guide are referred to as holistic, or analytical trait; general, or specific. Anchor papers or benchmark performances may be used to identify each level of competency in the rubric or scoring guide.

Performance Standards: Performance standards provide clear statements of the kinds of performances that constitute evidence that students had met the content standards. They answer the question, how well must a student perform? For example, we can set up assessments that correspond to our curriculum standards (say, mathematical problem solving), and then specify the level and kind of responses students need to give in order to be considered skilled (performance standards). (Also see Benchmark and Anchor Performances).

Performance Task: A performance task gives the student the opportunity to illustrate, perform, or demonstrate what they know and can do.

Portfolio: A purposeful collection of student work that tells the story or his/her growth as a learner is called a portfolio. There are many kinds of portfolios. The usefulness (for assessment and instruction) of any portfolio is enhanced by performance criteria, student involvement, and student self-reflection.

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.

Rating Scales: A scale based on descriptive words or phrases that indicate performance levels is called a rating scale. Qualities of a performance are described (e.g., advanced, intermediate, novice) in order to
ASSESSMENT GLOSSARY

designate a level of achievement. The scale may be used with rubrics or descriptions of each level of performance.

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

Rubric: A rubric is an established set of criteria for scoring or rating students’ performance on tests, portfolios, writing samples, or other performance tasks.

Scoring Criteria: Scoring criteria are rules for assigning a score or rating a student’s performance on tests, portfolios, writing samples, or other performance tasks. Scoring criteria may include rating scales, checklists, answer keys, and other scoring tools.

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

Selected Response Items: Selected response items are those that give the student choice and the student must select a response. These include multiple-choice, true-false, and matching items.

Self-Assessment: Self-assessment is the process of doing a systematic review of one’s own performance, usually for the purpose of improving future performance. Such assessment may involve comparison with a standard, established criteria. Self-assessment may involve critiquing one’s own work or may be a simple description of one’s performance.

Standard: A description of the outcomes or expectations of achievement which serves as a basis for defining more specific performance criteria and making assessment judgments (sometimes referred to as content standards as opposed to performance standards). (Note: while meeting a specific standard may be selected as a goal; not all goal statements are written specifically enough to be used as a standard).

Standardized Test: A test that is given and scored in a uniform manner is known as a standardized test. These 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. Standardized tests may produce norm-referenced or criterion-referenced information.

Task: A task is anything from a discrete multiple-choice or short-answer item to a complex project requiring students to use many different types of learning to solve a problem, investigate a situation, write a story, or do any other real-world task. The task is a whole. Within a task there may be several dependent items.

Thinking Skills: Thinking skills include thinking analytically, logically and creatively to form reasoned judgments and solve problems.

Validity: Validity refers to the extent to which the assessment measures the desired performance and appropriate inferences can be drawn from the results. A valid assessment accurately reflects the learning it was designed to measure.

Adapted from WA Commission on Student Learning
June 4, 1997 (Rev. 09/24/02)