

GLOSSARY

Accountability

Holding educators or others responsible for the performance of students, educators, or school programs.

Achievement Level

The standard of performance set through a standard setting procedure. Also called a “performance standard.”

Defines how well students need to do on an assessment to meet or exceed predefined targets for achievement, such as “proficient.”

Active Partner

An individual who takes a substantial and ongoing role in an activity.

Alignment

The match or agreement between different components of an assessment system: the content standards, the assessments, the scoring rubrics, and the performance standards.

Aptitude

A term to describe the ability of an individual to carry out a task or activity. Also indicates the extent to which an individual will be successful in a future activity.

Assessment: Development, Methods, Purposes, and Use of Tools

[See Appendix, page 18–19 of the ALS Brief.](#)

Assessment Administration Procedures

The set of policies, guidelines, and/or procedures in place to help ensure that the administration of an assessment provides valid results consistent with the designed purpose of the assessment.

Balance of Representation

The match between the relative emphasis of concept areas in a set of content standards and the assessment that measures those standards. The key question is does the balance of representation in the assessment match that of the content standards?

Balanced Assessment

The use of different types of assessment for different purposes. Can also mean the use of assessments *for* learning (to guide it as it is occurring) and *of* learning (to measure how much students have learned at the end of instruction).

Bias/Sensitivity/Distortion

Bias: The manner in which a test question is posed that disadvantages some students (due to factors other than their knowledge of the topic being assessed).

Sensitivity: The use of a topic in an assessment item that some students may find troubling or offensive.

Distortion

A factor in the assessment process that does not permit the accurate determination of student performance or that of a school or district.

Causation

A demonstration that one variable has a direct and predictable impact on another variable.

Cognitive Complexity

The type(s) of mental processing (i.e., thinking skills) required by an item or set of items. This may refer to the Depth of Knowledge (Webb), Bloom’s Taxonomy, or other definition of thinking skills.

Constructed- or Written-response Items

Test items that require students to write out their responses. Often, responses take the form of short- or extended-response essays, although other items might ask students to draw a picture, construct a table, show a flow chart, and so forth. A traditional “fill in the blank” type question is also a written response item. Constructed- or written-response items typically require a checklist or rubric for scoring.

Correlation

A demonstration that two variables move in the same or opposite manner, without proof that one causes the other.

Criteria

A basis for making a judgment.

Criterion-referenced (and interpretation)

Relating a test score to a pre-established absolute standard.

Data Management System

A computer software system that is used to store educational data and to permit these data to be retrieved and analyzed.

Depth of Knowledge (DOK)

A taxonomy of four levels, developed by Norm Webb, that can be used to classify the cognitive complexity of test items, content standards, and learning objectives.

Dispositions

Attitudes or beliefs about something.

Essential Learnings

A set of prioritized outcomes, derived from state standards, that helps focus on the most needed aspects of the curriculum for instructional planning purposes.

Feedback

Information about performance provided by another person or an instrument.

Field Test

Trying out of newly-created items in a formal manner on a representative sample of students.

Formative

Information collected and used during instruction to improve learning as it is occurring.

Grading

Rating an individual or program on the basis of external standards.

High Quality Assessment

An assessment externally judged to be of superior quality.

Horizontally Aligned

The alignment of instruction provided by multiple teachers teaching the same content at the same grade or in the same course.

Instructional Decisions

The choices made by educators as they teach.

Instructional Objective

A statement that specifies what a learner will know and be able to do as a result of instruction. Most often found in curriculum framework documents.

Instructionally Embedded

Assessments or activities that occur while instruction is taking place.

Interim

An assessment program that is administered periodically to students, such as at the conclusion of each marking period.

Interviews

In this type of assessment, a teacher typically works with an individual student, asks a series of planned and/or unplanned questions, and records students' responses to the questions.

Item

An assessment question, problem, or exercise. The individual measures used in a test.

Learning Progressions

The sequence of learning topics that students may go through to learn an important topic.

Learning Targets

The individual learning skills for teaching and/or testing.

Levels of Proficiency

The different levels of performance on an assessment.

Measures of Central Tendency: Mean, Mode, and Median

Mean: The arithmetic average of a set of data, calculated by adding up all the scores and dividing by the number of scores.

Mode: The most frequently occurring score in a set of scores.

Median: The score at the middle point in a set of scores.

**Measures of Variability:
Variance and Standard
Deviation**

Variance: The deviation of each score in a set of scores from the mean score of the set, squared.

Standard Deviation: The square root of the variance of each score in a set of scores, divided by the number of scores.

Multiple Measures

The use of different types of measures to assess students or programs from somewhat different perspectives in order to obtain a broader picture of students or a program.

Norm-referenced (and interpretation)

The comparison of a student or school score to a representative sample of students or schools – the norm group. Scores are interpreted as above or below the average (mean score) of the norm group.

Performance

Assessments that require the student to perform some activity.

Performance Assessments

Assessments where students are asked to reform in some way, such as completing an experiment in science, conducting an investigation in science, singing, acting out a character in a theatrical production, or completing a painting in an arts class. The products of performance assessment can be many types. Performance assessments typically require a checklist or a rubric for scoring.

Pilot Testing

A preliminary use of assessment items to see if they work. If they don't, they may be discarded or revised. If they do work, the next step is to field test them.

Professional Development (Targeted and Differentiated)

The learning programs and experiences provided to in-service educators to improve their knowledge and skills, and thus, their performance on the job.

Professional Learning Communities

Small groups of educators who work on a common issue or program over a period of time for the purposes of increasing educator effectiveness and student results.

Program Evaluation

The use of test results to determine the success of a program and perhaps to suggest improvements to it.

Protocols

An agreed upon set of guidelines for conversation; a code of behavior for groups to use when exploring ideas.

Quality Assessment

A judgment that an assessment is of high quality.

Reflection

The process that occurs when students think about how their work meets established criteria; they analyze the effectiveness of their efforts, and plan for improvement.

Reliability

A determination of the internal consistency, comparability, or stability of an assessment. A necessary, but not sufficient, condition for an assessment to be useful.

Reporting

Describing the performance of a student on an assessment in written or verbal terms.

Rigor

The level of knowledge necessary to achieve a content standard or to correctly respond to an assessment item. Typically measured in the Depth of Knowledge category, one of four dimensions of the Webb Alignment Tool, developed by Norm Webb, Wisconsin Center for Education Research.

Scoring

The process of determining how well a student did on an assessment.

Scoring Tools

[See Appendix, page 19 of the ALS Brief.](#)

Selected-Response Items

A test item that requires students to pick a response from among two or more answer choices provided for each item. Multiple-choice, true-false, and matching items are all examples of selected response items. Multiple-choice items are the most frequently used type of selected-response items.

Standards

The larger expectations we express in association with knowledge, skills and dispositions, comprising entire disciplines (mathematics, science, etc.).

Student-friendly Language

Writing of some educational language in a jargon-free manner understandable to students.

Subgroup Performance

The performance of a subset of the students in a larger group, examined to assure that all groups of students in a school are doing well academically.

Success Criteria

Statements that tells students what they should know, understand and be able to do, at the end of a lesson. These criteria identify elements of quality that will be present in student work. These criteria become the measures teachers use to determine proficiency.

Summative Assessment

As assessment of performance, conducted at the conclusion of a course or program completion.

Test Blueprint

A document that describes the key attributes of a new assessment, such as standards to be assessed, the types and numbers of items to be written, and how the results of the assessment will be reported to different audiences.

Types of Assessments

Different ways of assessing students or programs.

Unpacking Standards

To determine the key attributes and aspects of a content standard.

Validity

The collection of evidence to support the intended uses of an assessment. Note: The test itself is not “valid” or “not valid.” It is the uses of the assessment that are or are not valid.

Vertically Aligned

The alignment of instruction provided by multiple teachers teaching in the same content area across two or more grades.

Walk-Through

A dry-run of a process or a procedure. Also can mean a school administrator who periodically observes teachers in their classrooms.