

# 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.”

## Achievement Test

A test used to determine the current level of knowledge and skills of an individual.

## Alignment

Refers to whether an assessment item measures any part (ideally, the most important part) of a content standard. Also refers to how much of a set of content standards that an assessment instrument measures.

*Two-Way Alignment* refers to how much of a set of content standards is measured by an assessment instrument, as well as, whether the assessment instrument covers most, if not all, of the set of content 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.

## Aptitude Test

A test used to determine 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 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.

## Assessment For and Of Learning\*

Assessments for learning guide learning as it is occurring, and assessments of learning measure how much students have learned at the end of instruction.

## Assessment Methods

*Selected-Response Item* In this type of item, students select a correct answer from among several answer choices. This item type includes multiple-choice, true/false items, and matching items. The multiple-choice item format is the selected-response format most used in a large-scale assessment program.

*Constructed-Response/Item* The item type requires the individual to create their own answer(s) rather than select from prewritten options. There are usually several ways in which these items can be answered correctly. These items are scored using a standardized scoring rubric that is objective and clearly defined.

*Performance Assessment* These are types of assessments that require the student to perform some activity. There are two types, distinguished by their complexity and the length of time students have to respond to them.

*Performance Task* In this assessment, students have days, weeks, or months to compose a response. Thus, these assessments may involve multiple responses of different types to multiple prompts. The resultant work may be lengthy and comprise of multiple parts. Embedded in the Task may be written response items, presentations, papers, student self-reflections, and so forth.

*Performance Event* This is an on-demand performance assessment on which students are given little or no time to rehearse their performance nor limited opportunities to improve their initial performance. Such assessments may take a class period or less to administer.

*Personal Communication* An assessment conducted one-on-one between an adult and a student- sometimes an observation or interview.

## Assessment Purposes

*Student Improvement* Using test results to review past instruction or to alter future instruction provided to the student, due to performance on the test.

*Accountability* Using test results to hold educators or others responsible for the performance of students, educators, or school programs.

*Program Evaluation* Using results to determine the success of a program and perhaps to suggest improvements.

*Prediction* Using test results to determine the likelihood of a success of an individual in some future activity.

## Balance of Representation (Sampling)

The match between what is taught or emphasized from the standards or learning targets to what is sampled or addressed in the assessments.

## Balanced Assessment\* (verb)

Balanced Assessment is the act of meeting the needs of all assessment users equally well.

## Balanced Assessment System

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

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.)

## Causation

This is 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

This is a demonstration that two variables move in the same or opposite manner, although there is no 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 of performance.

## 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.

**Diagnostic Test**

A test used to determine the areas of strength and weakness of an individual.

**Dispositions**

Attitudes or beliefs about something.

**Distortion**

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

**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.

**Exemplars\***

Examples that give students insight into what excellence looks like or doesn't look like when working toward mastery of a learning target. Exemplars must be based on identified success criteria.

**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 Assessment**

Information collected and used by teachers and students during instruction to improve teaching and learning as it is occurring.

**Formative Feedback\***

Information in relation to a specific learning target that is given to students to describe actionable information to enhance their learning.

**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.

**Inter-rater Reliability\***

The extent of agreement or concordance of scores between different raters when viewing the same sample of work or performance. If there is a low level of agreement, this may mean that the scoring guide, including the scoring rubric, is in need of more development, or additional training is needed for the raters. In large scale assessment programs, complex research designs are required to establish inter-rater reliability due to the large number of raters used.

**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**

A continuum or trajectory of continuous, coherent development that connects knowledge, concepts and skills within a domain.

**Learning Targets**

The knowledge and skills students must acquire to master the standards.

**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**

Assessments where students are asked to perform 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.

**Performance Event**

This is an on-demand performance assessment on which students are given little or no time to rehearse their performance nor limited opportunities to improve their initial performance. Such assessments may take a class period or less to administer. One of two types of assessments that require the student to perform some activity. These two types are distinguished by their complexity and the length of time students have to respond to them.

**Performance Task**

On this assessment, students have days, weeks, or months to compose a response. Thus, these assessments may involve multiple responses of different types to multiple prompts. The resultant work may be lengthy and comprised of multiple parts. Embedded in the Task may be written response items, presentations, papers, student self-reflections, and so forth. One of two types of assessments that require the student to perform some activity. These two types are distinguished by their complexity and the length of time students have to respond to them.

**Personal Communication**

An assessment conducted one-on-one between an adult and a student—sometimes an observation or interview.

**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.

**Placement Test**

A test used to determine the best program or treatment for an individual.

**Prediction**

The use of test results to determine the likelihood of success of an individual in some future activity.



# Glossary

## 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.

## Progress Monitoring Test

A test used to gauge the improvement in performance of an individual or a program.

## Protocols

Protocols are 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.

## 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.

## Rubrics\*

A rubric identifies a coherent set of criteria for student work that includes expectations for performance at varying levels of quality.

## Scoring

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

## Scoring Checklists

This might be a series of steps used to remind students about a complete performance or used to score the responses of students.

## Scoring Guide

A scoring guide is composed of a rationale for the correct or preferred responses to the assessment. A guide includes one or more scoring rubrics, examples of student responses for each score level of each rubric, and sets of pre-scored student papers used to train, certify, and monitor the scorers.

## Scoring Rubrics

Often used to score constructed response items, and performance tasks and performance events. A rubric establishes the expectations for performance and delineates what a response must include. Performance levels are described for each dimension or criteria of the performance task, performance event, or constructed response item. Sample student work drawn from actual responses used to illustrate performance levels for each dimension/criteria are sometimes attached to a rubric.

## Screening Test

A test used to determine eligibility of an individual for a program or activity.

## 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.

## Selection Test

A test used to determine which individuals will most likely be successful in a program.

## Sensitivity

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

## Standard

What a student must know and be able to do by the end of a course or grade level.

## Student-friendly Language

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

## Student Improvement

The use of test results to review past instruction or to alter future instruction provided to the student, due to performance on the test.

## 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

These criteria identify elements of quality that will be present in student work at the end of a lesson. They become the measures teachers use to determine proficiency or performance.

## Summative Assessment

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

## Summative Feedback\*

Evaluative information or certification of how well a student has performed at the end of a learning cycle.

## 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.