

Assessment Literacy Glossary

This list of terms relates to the Local Alternative Assessment work and is not intended to be an exhaustive list of assessment terms.

Accountability systems – The mechanisms used (generally by states) to evaluate the performance of their education systems. In recent years, accountability systems have increasingly used the school as the unit for monitoring and intervention, based largely on the scores of each school’s students on a set of standardized tests.

Alternative assessment – Alternative assessments are used primarily to determine what students can and cannot do, in contrast to what they do or do not know. In other words, an alternative assessment measures applied proficiency more than it measures knowledge. There are multiple types of alternative assessments, of which performance assessment is one.

Assessment—Any systematic basis for gathering data or information and making inferences about characteristics, proficiencies, or abilities of people, usually based on various sources of evidence; the global process of synthesizing information about individuals in order to understand and describe them better.

Authentic assessment – An alternative assessment in which students perform a real-world task. A student will typically have to employ critical thinking and problem-solving skills to successfully address the challenge presented. The more authentic an assessment task is, the more closely it approximates the way a similar task would be done in a setting outside the classroom (a workplace or community, for example). Student performance on a task is typically scored on the basis of a list of desired outcomes or criteria (known as a rubric). Authentic assessment is a type of performance assessment, sometimes referred to as an “authentic performance assessment.”

Balanced assessment system – An assessment system that employs multiple types of assessments so that: (1) achievement and growth are taken into account; (2) assessments are matched to learning goals (both core content mastery and skills for success in the modern world); and (3) the need for accountability measures is met, but not at the expense of meaningful information that informs classroom instruction.

Formative Assessment – The overall goal for formative assessment is to intentionally collect information about the nature and/or degree of student learning that can be used to provide feedback to students and for teachers to make instructional decisions that progress student learning during the learning process. Formative assessment is used to “inform” the learning process so that the teacher can make in-process adjustments and learning modifications based on the data collected from students. Formative assessments are generally referred to as assessments “for” learning and may be techniques or instruments.

Higher-Order Thinking Skills (HOTS) – A category of thinking skills that increases the cognitive load, requiring students to go beyond recalling facts, understanding content, or replicating rote procedures. Students employing higher-order thinking may make connections, solve problems different from those given in classroom examples, and use content to reach and justify conclusions. Examples of activities that require higher-order thinking are (1) analyzing the usefulness of information, (2) providing evidence to support conclusions, (3) creative thinking and design, and (4) determining implications and

consequences. Complementary to HOTS are lower-order thinking skills (LOTS). LOTS are a category of thinking skills that are characterized by knowledge, understanding, and application of procedural skills. A deep and rich use of HOTS is often dependent upon LOTS, while engaging in HOTS can often lead to a deepening of one's knowledge and understanding.

Integrated Performance Assessment (IPA) or Interdisciplinary Assessment – An assessment that measures student performance on content and/or skills across two or more subject areas.

Inter-rater Reliability – The degree of agreement among raters who are tasked with scoring a performance task or product.

Intra-rater Reliability – The degree of consistency with which a single rater scores a set of students' work on a performance-based assessment.

Local Alternative Assessment (LAA)—Assessments that are administered in the place of eliminated SOL tests. Currently that refers to the 5 eliminated SOL assessments in grades 3-7 and will include other subject areas where SOL assessments are removed in the future.

Performance Assessment or Performance-Based Assessment (PBA) – A type of alternative assessment in which students demonstrate the use of their acquired knowledge and skill. A performance assessment may include a written component, but generally focuses primarily on the student's demonstration of a specified task and/or the creation of a product. Performance assessments are typically scored using rubrics (see Rubrics), which explicitly describe levels of performance and designate which levels meet standards.

Performance Task – A performance task is any learning activity that asks students to *perform* to demonstrate their knowledge, understanding and proficiency, yield a tangible product and/or performance that serve as evidence of learning, or presents a situation that calls for learners to apply their learning in context.

Portfolio Assessments – A type of assessment that is a systematic collection of student work and artifacts that demonstrate mastery of course and/or content knowledge and skills over an identified period of time.

Project-Based Learning (PBL) – A teaching method or approach that engages students in sustained, collaborative, real-world investigations. Projects are organized around a driving question, and students participate in a variety of hands-on tasks that seek to meaningfully address this question (Buck Institute). Performance assessment is typically a component of a PBL approach to teaching and learning.

Reliability – The consistency or stability of test performance. Tests must be constructed and administered so that measurement error from factors such as ambiguous scoring, unclear questions/directions, bias, cheating, or environmental factors is minimized.

Rubric – A description of the criteria for success and levels of achievement for a task or product. Rubrics are used during instruction to help students maximize and improve the quality of their work and as scoring tools for multiple types of alternative assessments (see Performance Assessment).

Summative Assessment – Assessments that are used to evaluate student learning, skill acquisition, and academic achievement at the conclusion of a defined instructional period. Summative assessments

typically are administered at the end of a unit, project, course, semester, program, or school year. Summative assessments are typically comprehensive of some representative set of knowledge and skills, and such assessments are typically associated with a high-stakes decision (e.g., a grade in a course, promotion to another level, verification of a course credit). These are frequently described as assessments “of” learning.

Validity – The degree to which an assessment actually measures the learning it is intended to measure. In order to strengthen and account for the validity (and reliability) of an assessment, assessment designers use a combination of procedures and tools in the development of, the administration of, and the post-administration analysis of assessments.