



Glossary

Assessment Literacy Standards

Terminology	Definition
Accountability	Holding organizations, educators, or others responsible for the performance of students, educators, or school programs.
Achievement Level	The standard of performance typically 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 or groups of students in a specific disciplinary or content domain.
Alignment	<p>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 an assessment instrument measures.</p> <p>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.</p>
Alternate Assessment	An assessment for students with the most significant cognitive disabilities that measures grade-level academic content knowledge and skills or English language proficiency. Alternate assessments are reduced in depth, breadth, and complexity from the general assessments, as appropriate for these students.



Terminology	Definition
Alternative Assessment – Idaho Reading Indicator specific	An early literacy assessment for students with the most significant cognitive disabilities who cannot access the current Idaho Reading Indicator, that measures the same skills or standards, yet in a different modality. Students with the most significant cognitive disabilities may take an alternative to the Idaho Reading Indicator in lieu of the current Idaho Reading Indicator as outlined in a student’s IEP.
Aptitude	A term to describe the ability of an individual to carry out a task or activity.
Aptitude Test	A test used to determine the ability of an individual to carry out a task or activity. Also predicts the extent to which an individual may be successful in a future activity.
Assessment	<p>An evidence-based tool or procedure designed to determine the status of an individual student or group of students on an attribute of interest. This includes a variety of types of methods such as tests, observational protocols, interviews, performance measures, and so forth. A test is only one type of assessment.</p> <p>There are two broad types of assessment:</p> <p>Assessment for Learning — Assessment that takes place while instruction is occurring. The information is used to enhance instruction and student learning.</p> <p>Assessments of Learning — Summative measures of student achievement used at the conclusion of instruction, when little or no new instruction is anticipated on the content assessed.</p>
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.



Terminology	Definition
Assessment Methods	<p><i>Selected-response Item</i>—Students select a correct answer from among several answer choices. This item type includes multiple-choice, true-false, and matching items. The multiple-choice item format is the selected- response format most used in a large-scale assessment program.</p> <p><i>Constructed-response Item</i>—This item type requires the individual to create their own response(s) rather than select from predetermined options. There are usually several ways in which these items can be answered correctly. These items may be scored using a standardized scoring rubric that is objective and clearly defined.</p> <p><i>Performance Assessment</i>—Requires the student to perform some activity. There are two types—performance task and performance event, distinguished by their complexity and the length of time students have to respond to them.</p> <p><i>Performance Task</i>—In this type of 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 multiple parts.</p> <p>Embedded in the task may be written response items, presentations, papers, student self-reflections, and so forth.</p> <p><i>Performance Event</i>—An on-demand performance assessment in which students are given little or no time to rehearse their performance and limited opportunities to improve their initial performance. Such assessments may take a class period or less to administer.</p> <p><i>Personal Communication</i>—An assessment conducted one-on-one between an adult and a student—sometimes an observation or interview.</p>



Terminology	Definition
Assessment Purposes	<p>Assessments are built and used for different purposes. Among these are:</p> <p><i>Student Improvement</i>—Using test results to review past instruction or to alter future instruction provided to the student, due to performance on the test.</p> <p><i>Accountability</i>—Using test results to hold educators or others responsible for the performance of students, educators, or school programs.</p> <p><i>Program Evaluation</i>—Using test results to determine the success of a program and perhaps to suggest improvements in it.</p> <p><i>Prediction</i>—Using test results to determine the likelihood of success of an individual in some future activity.</p>
Balance of Representation	<p>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 of the assessment match that of the content standards?”</p>
Balanced Assessment System	<p>The use of different types of assessment for different purposes by different users. Can also mean the use of assessments for learning (to guide instruction as it is occurring) and of learning (to measure how much students have learned at the conclusion of instruction). All users have access to relevant assessment information.</p>
Bias	<p>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.</p>



Terminology	Definition
Causation	A demonstration that one variable has a direct and predictable impact on another variable. Note: a correlation between two variables (e.g., poverty and student achievement) does not mean that one variable (poverty) causes the other one (low achievement).
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 definitions of thinking skills.
Comprehensive Assessment System	A coordinated system of assessments aligned to standards and success criteria for the whole child. The assessments and assessment processes are used to inform instruction, measure progress, specify additional learner needs, guide in-depth supports, indicate growth toward competencies, and provide information regarding outcomes.
Computer Adaptive Assessment	An assessment administered online in which the questions asked of students are determined by their performance on previous questions, permitting a more precise determination of the level of student performance on the assessment.
Constructed- or Written-response Items	Test items that require students to write out their responses. Often, responses take the form of short or extended responses to one or more prompts. Items might include essays, student drawings, and online technology-enhanced assessments such as constructing a mathematics table, drawing a flow chart, and so forth. Constructed- or written-response items may use a checklist or rubric for scoring.
Content Standard	The knowledge, skills, and disposition expectations comprising entire disciplines (mathematics, science, etc.).



Terminology	Definition
Correlation	A demonstration that two variables move in the same or opposite manner. This does not mean that one variable causes the other. See Causation.
Criterion-referenced Score 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, analyzed, and used.
Depth of Knowledge (DoK)	The four levels of conceptual rigor developed by Norm Webb that can be used to classify the cognitive complexity of test items, content standards, and learning objectives. The DoK levels are 1-recall; 2-skill/concept; 3-strategic thinking; and 4-extended thinking.
Diagnostic Test	Measures specific skills, behaviors or characteristics that comprise more complex student outcomes, to identify learners who may need additional targeted interventions. Diagnostic tests can follow the initial use of screeners to more fully understand the achievement and learning needs of individual students.
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.



Terminology	Definition
English Language Development Assessment	Assessments in the areas of reading, writing, listening, and speaking (as well as comprehension) for students in grades K–12 whose home language is other than English. Participation qualifies students for English language development opportunities, and it continues until (and beyond) when students reach pre-defined levels of English language proficiency.
Essential Learning	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	Trying out of newly-created items in a formal manner on a representative sample of students.
Field Test	Trying out newly created items in a formal manner with a representative sample of students.
Formative Assessment	A planned, ongoing process used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become more self-directed learners (CCSSO FAST SCASS, 2018).
Formative Feedback	Information in relation to a specific learning target that is given to students to describe actionable information to enhance their learning.



Terminology	Definition
Grading	Rating an individual or program based on external standards. As student learning and achievement changes, grades change to reflect current levels of student learning and achievement.
High Quality Assessment	An assessment that provides reliable, valid, bias-free, and useful information.
Horizontally-Aligned Instruction	The alignment of instruction provided by multiple teachers teaching at the same grade level or in the same course.
Instructional Decisions	The choice 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.
Instructional Program Improvement	The use of test results to determine areas of the instructional program that need to be modified and/or improved in the future.
Instructionally-Embedded Assessment	Assessments or assessment-related 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.



Terminology	Definition
Interim/Benchmark Assessments	Assessments are administered periodically throughout the school year for one or more of the following purposes: predictive (identify learner readiness for success on a later summative assessment); evaluative (to appraise ongoing educational programs); and/or instructional (to supply teachers with individual learner performance data).
Interviews	In this type of assessment, a teacher typically asks an individual student a series of questions and records students' responses to the questions.
Item	An assessment question, problem, or exercise. The individual questions that are used in a test.
Learning Progressions	The sequence of learning topics that students typically go through to learn an important topic/skill.
Learning Targets	The individual learning outcomes based on content standards used for teaching and/or testing.
Levels of Proficiency	The different levels of performance on an assessment.
Measures of Central Tendency: Mean, Mode, and Median	<p><i>Mean</i>—The arithmetic average of a set of data, calculated by adding all the scores and dividing the total by the number of scores.</p> <p><i>Mode</i>—The most frequently occurring score in a set of scores.</p> <p><i>Median</i>—The score at the middle point in an ordered set of scores.</p>



Terminology	Definition
Measures of Variability: Variance and Standard Deviation	<p><i>Variance</i>—The sum of the squared deviations of scores in a set of scores from the mean score of the set, divided by one less than the total number of scores.</p> <p><i>Standard Deviation</i>—The square root of the variance.</p>
Multiple Measures	The use of different types of measures (e.g., assessment for learning and assessments of learning) to assess students or programs from different perspectives to obtain a broader picture of students or a program.
Norm-referenced Score 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, such as a percentile rank.
Observation	An assessment of one or more aspects of student performance by a trained observer, either in a natural setting or one that has been structured especially for the observation. A protocol or rubric may be used by the observer.
Performance Assessments	Assessments where students are asked to perform in some manner (e.g., completing an experiment, conducting an investigation, singing, acting in a theatrical production, or completing a painting). Performance assessments require a checklist or a rubric for scoring.
Performance Event	An on-demand performance assessment on which students are given little or no time to rehearse their performance and limited opportunities to improve their initial performance. Such assessments may take a class period or less to administer.



Terminology	Definition
Performance Task	On this type of assessment, students have days, weeks, or months to compose a response. A performance task may involve multiple responses of different types to multiple prompts. The resultant work may be lengthy and comprise multiple parts. Embedded in the task may be written-response items, presentations, papers, student self-reflections, and so forth.
Personal Communication	An assessment conducted one-on-one between an adult and a student— sometimes an interview.
Pilot Test	A preliminary use of assessment items to see if they yield useful information as anticipated. If they don't, they may be discarded or revised. If they do work, the next step is to field test them.
Placement/Selection Test	A test used to suggest the best educational program or service for an individual.
Prediction	The use of test results to determine the likelihood of success of an individual in some future activity.
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.
Priority Standards	A set of prioritized outcomes, derived from a larger set of content standards, that helps teachers focus on the most important aspects of the curriculum for instructional planning purposes.
Professional Learning	The learning programs and experiences in-service educators engage in to improve their knowledge and skills and, thus, their performance on the job.



Terminology	Definition
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 and other data to determine the success of a program and perhaps suggest improvements to it.
Progress Monitoring	Used to quantify a learner's rate of improvement or responsiveness to instruction, intervention, or supports and to guide future learning. Progress monitoring can be implemented with individual learners and small groups. The frequency of measures should match the level of intervention intensity. Progress monitoring can be used to inform whether individual student interventions are effective.
Protocols	Guidelines for students to use when completing an assessment. Protocols are also 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 statistical determination of the internal consistency, comparability, or stability of an assessment. A necessary, but not sufficient, condition for a summative assessment to be useful.
Reporting	Describing the performance of a student or a group of students on an assessment in written, graphical, or verbal terms.



Terminology	Definition
Rigor	The level of knowledge necessary to achieve a content standard or to correctly respond to an assessment item. Often 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.
Rubric	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 Checklist	Specifies the steps a student must carry out to complete an item. May also be a list 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 also 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, 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 criterion of the performance task, performance event, or constructed response item.
Screening Test	A relatively short test used to determine eligibility of an individual for a program or activity.



Terminology	Definition
Selected-response Item	A test item that requires students to pick a response from among two or more answer choices provided. Multiple-choice, true-false, and matching items are all examples 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.
Simulation	A performance assessment based on a real-life situation used to assess students' ability and readiness to carry out the activity in a real-life situation.
Standard	What a student must know and be able to do by the end of a course or grade level.
Student-friendly Language	Educational language written 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 students from a larger group, examined to ensure that all student groups in a school are achieving content standards equitably.



Terminology	Definition
Success Criteria	Lists, statements, models, or exemplars that tell or show students what they should know, understand, and be able to do at the end of instruction. These criteria identify elements of quality that should be present in student work. These criteria may be used by teachers to determine student proficiency.
Summative Assessment	An assessment that provides information regarding the level of learner, program, or school success at an endpoint in time. Administered at the conclusion of learning to 1) determine the effectiveness of a recently concluded program, 2) make inferences about a learner’s mastery of curricular aims, and/or 3) meet local, state and federal accountability requirements.
Summative feedback	Evaluative information or certification of how well a student has performed at the end of a learning cycle.
Technologically-Enhanced Assessments	Assessments administered online that use the capabilities of assessment software to permit students to respond to open-response types of assessment items, but yield machine-scorable responses.
Test Blueprint	A document that describes the key attributes of an assessment, such as standards to be assessed, the depth of knowledge of the items to be used, and the types and numbers of items to be included.
Types of Assessments	Different ways of assessing students or programs.
Universal Screeners	Designed to measure skills, behaviors, or characteristics that may be predictive of later outcomes to identify learners who might need additional diagnostic assessments and/or interventions. Screeners are efficient and provide comparative data for all learners within a defined group or setting.



Terminology	Definition
Unpacking Content Standards	Determining the key attributes and aspects of a content standard, such as depth of knowledge and learning targets.
Validity	The extent to which a test measures what it is intended to measure. A test itself is neither “valid” nor “invalid.” Validity is determined based on the adequacy of the information in support of the intended use of the assessment.
Vertically Aligned Instruction	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. Can also mean a school administrator who periodically observes teachers in their classrooms.

ACKNOWLEDGEMENT

We sincerely thank the Michigan Assessment Consortium for building the assessment literacy standards and their glossary. For more information, visit their website at: [Assessment Literacy Standards - Michigan Assessment Consortium](#).

For Questions Contact

Assessment & Accountability

Idaho Department of Education

650 W State Street, Boise, ID 83702

208 332 6800 | www.sde.idaho.gov