

## GLOSSARY

Term	Definition	Source
<b>Ability (cognitive, developing, learned)</b>	Refers to a complex cognitive construct, such as reading comprehension, writing, or mathematical problem-solving.	Messick, S. (1989). Validity. In R. Linn (Ed.). <i>Educational measurement</i> . Mahwah, NJ: Lawrence Erlbaum Associates.
<b>Accommodation</b>	Changes in the administration of an assessment (such as setting, scheduling, timing, presentation format, response mode, or others), including any combination of these that does not change the construct intended to be measured by the assessment or the meaning of the resulting scores. Accommodations are used for equity, not advantage, and serve to level the playing field. To be appropriate, assessment accommodations must be identified in the student's Individualized Education Program (IEP) or Section 504 plan and used regularly during instruction and classroom assessment.	<i>Assessing Special Education Students</i> SCASS, Council of Chief State School Officers. (2003).
<b>Achievement levels/proficiency levels</b>	Descriptions of a test taker's competence in a particular area of knowledge or skill, usually defined as ordered categories on a continuum, often labeled from "basic" to "advanced," or "novice" to "expert," that constitute broad ranges for classifying performance. See <i>Cut score</i> .	Ibid., p. 171.
<b>Achievement test</b>	A test to evaluate the extent of knowledge and skills attained by a test taker in a content domain in which he or she has received instruction.	Ibid.

Term	Definition	Source
<b>Alternate assessment</b>	<p>An assessment designed for the small number of students with disabilities who are unable to participate in the regular state assessment, even with appropriate accommodations.</p> <p>An alternate assessment is not one particular format of assessment. Rather, an alternate assessment might include materials collected under several circumstances, including, but not limited to, (a) teacher observation of the student, (b) samples of student work produced during regular classroom instruction that demonstrate mastery of specific instructional strategies in place of performance on a computer-scored multiple choice test covering the same content and skills, or (c) standardized performance tasks produced in an “on demand” setting, such as completion of an assigned task on the test day.</p>	U. S. Department of Education. (2003, Dec. 9). Title I—Improving the academic achievement of the disadvantaged; Final rule, 68 Fed. Reg. 236.
<b>Argument for validation</b>	In the process of validation, the test developer formulates a rationale for why a test score might be validly used for a specific kind of interpretation or use.	Kane, M. T. (2002). Validating high-stakes testing programs. <i>Educational Measurement: Issues and Practice</i> , 21(1), 31–41.
<b>Assessment</b>	Any systematic method of obtaining information from tests and other sources; used to draw inferences about characteristics of people, objects, or programs.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 172.
<b>Assessment approach system</b>	A system for assessing students that uses some combination of selected or constructed responses and includes one of four dominant methods: (a) ratings based on teacher reflections or observations; (b) portfolios that include collections of evidence; (c) performance tasks comprising constructed responses; or (d) performance events that provide more extended complex constructed responses.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).
<b>Assistive technology</b>	A device or service that is used to increase, maintain, or improve the functional capabilities of a student with a disability (see 34 C.F.R. 300.5 and 300.6).	Assessing Special Education Students SCASS, Council of Chief State School Officers. (2003).

Term	Definition	Source
<b>Bias</b>	In a statistical context, a systematic error in a test score. In discussing test fairness, bias may refer to construct underrepresentation or construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 172.
<b>Breadth</b>	The comprehensiveness of the content and skills embodied in the standards, curriculum, or assessments.	Assessing Special Education Students SCASS, Council of Chief State School Officers. (2003).
<b>Complexity</b>	The confluence of breadth and depth with requisite skills implied in either the standards or the assessments (e.g., more complex standards and assessments require more requisite skills that are more advanced).	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Concept</b>	Concepts comprise three components: (a) a label, (b) attributes that define essential features specifying examples versus nonexamples, and (c) illustrative examples. Various categories exist for organizing concepts (e.g., concrete or abstract).	Tennyson, R. D. & Park, O. (1980). The teaching of concepts: A review of instructional design research literature. <i>Review of Educational Research</i> , 50(1), 55–70; and Martorella, P. H. (1972). <i>Concept learning: Designs for instruction</i> . San Francisco: Intext Educational Publishers.
<b>Constrained inferences</b>	The inferences about achievement on grade-level content standards is limited in some aspect due to the changes that have been made in the supports used to give or take the large-scale assessment or due to changes in breadth, depth, or complexity. Nevertheless, meaningful inferences can be made about student achievement and are important for tracking student's progress toward grade-level content standards.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Construct</b>	The concept or the characteristic that a test is designed to measure.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 173.
<b>Construct domain</b>	The set of interrelated attributes (e.g., behaviors, attitudes, and values) that are included under a construct's label. A test typically samples from this	Ibid.

Term	Definition	Source
	construct domain.	
<b>Construct equivalence</b>	(a) The extent to which the construct measured by one test is essentially the same as that measured by another test. (b) The degree to which a construct measured by a test in one cultural or linguistic group is comparable to the construct measured by the same test in a different cultural or linguistic group.	Ibid.
<b>Construct irrelevance</b>	The extent to which test scores are influenced by factors that are irrelevant to the construct that the test is intended to measure. Such extraneous factors distort the meaning of test scores from what is implied in the proposed interpretation.	Ibid., pp. 173–174.
<b>Construct under-representation</b>	The extent to which a test fails to capture important aspects of the construct that the test is intended to measure. In this situation, the meaning of test scores is narrower than the proposed interpretation implies.	Ibid., p. 174.
<b>Construct validity</b>	A term used to indicate that the test scores are to be interpreted as indicating the test taker's standing on the psychological construct measured by the test. A construct is a theoretical variable inferred from multiple types of evidence, which may include the interrelations of the test scores with other variables, internal test structure, or observations of response processes, as well as the content of the test. In the current standards, all test scores are viewed as measures of some construct; construct validity is synonymous with validity. The validity argument establishes the claim for the construct validity of a test score interpretation or use.	Ibid.

Term	Definition	Source
<b>Constructed-response item format</b>	An exercise for which test takers must create their own responses or products rather than choose a response from an enumerated set. Short-answer items require a few words or a number as an answer whereas extended-response items require a few sentences or more.	Ibid.
<b>Content domain</b>	The set of behaviors, knowledge, skills, abilities, attitudes, or other characteristics to be measured by a test, represented in a detailed specification, and often organized into categories by which items are classified.	Ibid.
<b>Content standard</b>	A statement of a broad goal describing expectations for students in a subject matter at a particular grade or at the completion of schooling.	Ibid.
<b>Content validity</b>	In the current standards, this type of evidence is characterized as evidence based on test content.	Ibid.
<b>Context</b>	The setting in which students receive their primary instructional program.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Cue</b>	Assistance given to a student to increase the likelihood that he or she will give the desired response. See <i>Prompt</i> .	Assessing Special Education Students SCASS, Council of Chief State School Officers. (2003).
<b>Cut score</b>	A specified point on a score scale such that scores at or above that point are interpreted or acted upon differently from scores below that point. See <i>Performance standard</i> .	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 175.
<b>Depth</b>	The level of cognitive processing, i.e., recognition, recall, problem solving, analysis, synthesis, and evaluation, required for success relative to the performance standards.	Assessing Special Education Students SCASS, Council of Chief State School Officers. (2003).

Term	Definition	Source
<b>Disability</b>	<p>A physical, mental, or developmental impairment that substantially limits one or more of a student's major life activities.</p> <p>Alternative definition: A deviation in cognitive, motor, or sensory functioning that results in difficulty responding to environmental demands.</p>	<p>American Educational Research Association, American Psychological Association, &amp; National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i>. Washington, DC: American Educational Research Association, p. 101.</p> <p>Technical Working Group on Large-Scale Assessments for Special Education. (2005).*</p>
<b>Discriminant evidence</b>	Evidence based on the relationship between test scores and measures of different constructs.	Ibid., p. 175.
<b>Documentation</b>	The body of literature (e.g., test manuals, manual supplements, research reports, publications, technical reports, users' guides, etc.) made available by publishers and test authors to support test use.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Domain sampling</b>	The process of selecting test items to represent a specified universe of performance.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 175.
<b>Flag</b>	Where evidence about the validity of a test accommodation or modification is lacking, a flag notes this fact to help others interpret the test result.	Ibid., standards 9.5 and 10.11.
<b>Functional equivalence</b>	In evaluating test translations, the degree to which similar activities or behaviors have the same functions in different cultural or linguistic groups.	Ibid., p. 176.
<b>Grade equivalent</b>	The school grade level for a given population for which a given score is the median in that population. Can also be applied to age equivalent.	Ibid.
<b>High-stakes test</b>	A test used to provide results that have important, direct consequences for examinees, programs, or institutions involved in the testing.	Ibid., p. 177.

Term	Definition	Source
<b>Item</b>	A statement, question, exercise, or task on a test for which the test taker is to select or construct a response or perform a task. See <i>Item prompt</i> .	Ibid.
<b>Item prompt</b>	The question, stimulus, or instructions that direct the efforts of examinees in formulating their responses to a constructed-response exercise.	Ibid.
<b>Knowledge form</b>	Classification of information types consisting of facts, concepts, principles, skills, or procedures.	Tindal, G. A., & Marston, D. B. (1990). <i>Classroom-based assessment: Evaluating instructional outcomes</i> . Columbus, OH: Charles Merrill Publishing.
<b>Performance assessments</b>	Product and behavior measurements based on settings designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 179.
<b>Performance standard</b>	(a) An objective definition of a certain level of performance in some domain in terms of a cut score or a range of scores on the score scale of a test measuring proficiency in that domain. (b) A statement or description of a set of operational tasks exemplifying a level of performance associated with a more general content standard. The statement may be used to guide judgments about the location of a cut score on a score scale. The term often implies a desired level of performance.	Ibid.
<b>Portfolio</b>	In assessment, a systematic collection of educational or work products that have been compiled or accumulated over time according to a specific set of principles.	Ibid.
<b>Proficient</b>	Status of being at a satisfactory level of performance on an achievement standard.	Cizak, G. J. (2004). <i>Standard setting</i> . Mahwah, NJ: Lawrence Erlbaum Associates.

Term	Definition	Source
<b>Prompt</b>	Any form of verbal, nonverbal, or physical cue to structure, pace, or signal a response to be made by the student. Examples include verbalisms like “continue,” “next,” “now what,” or reminders of each step. Physical guidance is another example of a prompt. See <i>Cue</i> .	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Random error</b>	An unsystematic error; a quantity (often indirectly observed) that appears to have no relationship with any other variable. Related to reliability and the standard error of measurement.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 180.
<b>Reliability</b>	The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group. See <i>Generalizability theory</i> .	Ibid.
<b>Scaffold</b>	Any type of structural assistance (from peers or teachers and consisting of either verbal or physical prompts) introduced to organize information or guide responses embedded in the presentation of the item or task. Examples include the addition of highlights, underlines, outlines, crib sheets, or other information to “essentialize” the task or response.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Score</b>	Any specific number resulting from the assessment of an individual; a generic term applied for convenience to such diverse measures as test scores, estimates of latent variables, production counts, absence records, course grades, ratings, and so forth.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 181.
<b>Selected-response item (multiple choice)</b>	A test exercise where a test taker chooses an answer from an enumerated list of choices.	Haladyna, T. (2004). <i>Constructing multiple-choice tests</i> . Mahwah, NJ: Lawrence Erlbaum Associates.



Term	Definition	Source
<b>Standards-based assessment</b>	Assessments intended to represent systematically described content and achievement standards	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 183.
<b>Stipulated</b>	Limited in making any inferences to only the assessment approach and breadth, depth, and complexity reflected in the assessment.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Systematic error</b>	A consistent score component (often observed indirectly) not related to the test performance. Related to construct-irrelevance and construct underrepresentation.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 183.
<b>Technical report</b>	A document that summarizes validity evidence over a time period (usually a year) for a specific test score interpretation or use.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Test manual</b>	A publication prepared by test developers and publishers to provide information on test administration, scoring, and interpretation and to provide technical data on test characteristics.	Technical Working Group on Large-Scale Assessments for Special Education. (2005).*
<b>Test specifications</b>	A detailed description for a test (often called a test blueprint) that specifies the number or proportion of items that assess each content and process or skill area; the format of items, responses, and scoring rubrics and procedures; and the desired psychometric properties of the items and test, such as the distribution of item difficulty and discrimination indices.	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). <i>Standards for educational and psychological testing</i> . Washington, DC: American Educational Research Association, p. 183.
<b>Validation</b>	The process through which the validity of the proposed interpretation of test scores is investigated.	Ibid., p. 184.

Term	Definition	Source
<b>Validity</b>	The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test.	Ibid.

- 
- Members of the Technical Working Group on Large-Scale Assessments for Special Education (2005) included: Patricia Almond, Assessment Consultant; Diane Browder, University of North Carolina at Charlotte; Malinda Crawford, University of Colorado at Colorado Springs; Steve Ferrara, American Institutes for Research; Tom Haladyna, Arizona State University; Huynh Huynh, University of South Carolina; Gerald Tindal, University of Oregon; and Naomi Zigmond, University of Pittsburgh.

The U.S. Department of Education is reviewing public comments received on the notice of proposed rulemaking regarding modified achievement standards. As this analysis is not completed, the content of this document may not necessarily reflect the final views or policies of the Department concerning modified achievement standards.

This document was produced under U.S. Department of Education Contract No. EDO4CO0025/0002 with the American Institutes for Research. Renee Bradley served as the contracting officer's representative. No official endorsement by the U.S. Department of Education of any product, commodity, service or enterprise mentioned in this report or on Web sites referred to in this report is intended or should be inferred.