

## ASSESSMENT 101: SUPPORTING HIGH QUALITY ASSESSMENT SYSTEMS

# Glossary

**Alignment** — Alignment refers to the degree to which an assessment covers the *depth* and *breadth* of the learning standards. Breadth is the degree to which an assessment covers the full range of learning standards. Depth is the degree to which an assessment allows students to demonstrate the range of higher-order, analytical thinking skills associated with a learning standard.

**Assessment<sup>1</sup>** — The term assessment refers to the wide variety of methods or tools that educators use to evaluate, measure, and document the academic readiness, learning progress, skill acquisition, or educational needs of students.

**Content standard** — See Learning standard

**Content validity** — The degree to which the content of an assessment matches the learning standards.

**Criterion validity** — The degree to which an assessment is related to an outcome (criterion) such as another assessment which has been demonstrated to be valid or a level of performance (e.g., grades in a course).

**Formative assessment<sup>2</sup>** — These assessments are small-scale (a few seconds, a few minutes, certainly less than a class period) and short-cycle (they are often called “minute-by-minute” assessment or formative instruction) for the purpose of providing feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes.

**High stakes assessment<sup>3</sup>** — A high-stakes test [assessment] is any test used to make important decisions about students, educators, schools, or districts, most commonly for the purpose of accountability—i.e., the attempt by federal, state, or local government agencies and school administrators to ensure that

students are enrolled in effective schools and being taught by effective teachers. In general, “high stakes” means that test scores are used to determine punishments (such as sanctions, penalties, funding reductions, negative publicity), accolades (awards, public celebration, positive publicity), advancement (grade promotion or graduation for students), or compensation (salary increases or bonuses for administrators and teachers).

**Interim assessment<sup>4</sup>** — These assessments fall between formative and summative assessment; these assessments (1) evaluate students’ knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and (2) are designed to inform decisions at both the classroom and beyond the classroom level, such as the school or district level for instructional, evaluative, or predictive purposes.

Assessment vendors sometimes refer to interim assessments as defined above as formative assessments. Interim assessments also sometimes called benchmark, diagnostic, or predictive assessments.

**Item** — See Test item.

**Learning standard<sup>5</sup>** — Learning standards are concise, written descriptions of what students are expected to know and be able to do at a specific stage of their education. Learning standards describe educational objectives—i.e., what students should have learned by the end of a course, grade level, or grade span—but they do not describe any particular teaching practice, curriculum, or assessment method. Learning standards are

**Low stakes assessment<sup>6</sup>** — A “low-stakes test” would be used to measure academic achievement, identify learning problems, or inform instructional adjustments, among other purposes. What distinguishes a high-stakes test from a low-stakes test is

not its form (how the test is designed) but its function (how the results are used). For example, if test results are used to determine an important outcome, such as whether a student receives a high school diploma, the test would be considered a high-stakes test regardless of whether it's a multiple-choice exam, an oral exam, or an essay exam. Low-stakes tests generally carry no significant or public consequences—the results typically matter far more to an individual teacher or student than to anyone else—and scores are not used to burnish or tarnish the reputation of teachers or schools.

**Reliability** — Reliability is the consistency with which an assessment measures what is being measured (e.g., reading, math). An assessment would be considered to have reliability if 1) a student took the same assessment several times AND 2) the conditions in which the assessment was given (e.g., classroom, time of day, lack of noise, etc.), student knowledge, and student effort were consistent across each administration AND 3) the student received similar (not necessarily identical) scores on the assessment.

**Standardized assessment**<sup>7</sup> — Standardized tests are large-scale tests that are administered to students and scored in the same manner. Students take the same test in the same conditions and, if possible, at the same time so that results can be attributed to student performance and not to differences in the test or the way it is given. Because of this, the results of standardized tests can be compared across schools and districts.

**Standard** — See Learning standard

**Summative assessment**<sup>8</sup> — These assessments are generally given one time at the end of some unit of time (end of a course, end of school year) to evaluate students' performance against a defined set of learning standards.

**Test** — See Assessment.

**Testing accommodation** — An accommodation is a change in how a test is administered in order to remove obstacles to the test-taking process that are presented by a disability or English proficiency without changing the skills or knowledge being tested. Changes in the test administration may include the method of presentation (e.g., reading the directions, Braille format), the method of response (e.g., marking responses in a test book), timing or scheduling (e.g., additional time to take a test, scheduling the test over multiple sessions), or setting (e.g., testing in a separate location or small group).

**Test item** — A test item is a task to which a test taker responds. The task may be presented in a variety of formats such as multiple-choice, a short response (e.g., fill in the blank), a constructed response (e.g., showing the work to complete a math word problem), or open-ended response (e.g., response to an essay prompt) to name a few.

**Validity** — Validity refers to the degree to which an assessment's results are representative of the actual knowledge and/or skills it is designed to measure AND whether the assessment results can be used to make accurate conclusions about the knowledge and/or skills.

## ENDNOTES

<sup>1</sup> <http://edglossary.org/assessment/> Visit the web page for additional information.

<sup>2</sup> Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Aspen, CO: Aspen Institute.

<sup>3</sup> <http://edglossary.org/high-stakes-testing/> Visit the web page for additional information.

<sup>4</sup> Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Aspen, CO: Aspen Institute.

<sup>5</sup> <http://edglossary.org/learning-standards/> Visit the web page for additional information.

<sup>6</sup> <http://edglossary.org/high-stakes-testing/> Visit the web page for additional information.

<sup>7</sup> <http://www.centerforpubliceducation.org/Main-Menu/Instruction/High-stakes-testing-and-effects-on-instruction-At-a-glance/Standardized-tests-and-their-impact-on-schooling-QA.html>

<sup>8</sup> Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Aspen, CO: Aspen Institute.