

HB 2680 Work Group: Key Terms

Achievement Standard: State Board-adopted scores on the Smarter Balanced scale that delineate the four achievement levels for each subject and tested grade. Achievement standards are also known as cut scores.

Achievement Level Descriptors: describe the knowledge, skills, and abilities that students should demonstrate on a standardized test in terms of categories of performance.

Alignment: the degree to which the content of an assessment is aligned with the content of the standards you intend to measure.

Assessment Accessibility: tools and supports that help students access test questions. The purpose of accessibility supports is **not** to reduce the rigor of the standards being assessed, but rather to avoid the creation of barriers for students who may need to demonstrate their knowledge and skills in different ways.

Bias: “...construct underrepresentation or construct-irrelevant components of test scores that differently affect the performance of different groups of test takers and consequently the reliability/precision and validity of interpretations and uses of their test scores.”¹

Computer Adaptive Test (CAT): An assessment that adjusts to a student’s ability, basing the difficulty of future questions on previous answers. A CAT aims to provide more accurate measurement of student achievement, particularly for high and low-performing students.

CAT algorithm: programmatic logic that selects the items to be administered to students based on the designated specifications.

Claims: broad statements of the assessment system’s learning outcomes, each of which requires evidence that articulates the types of data/observations that will support interpretations of competence towards achievement of the claims; claims serve as the fundamental drivers for the design of Smarter Balanced’s English language arts (ELA)/literacy and mathematics summative assessments.

Cluster: a grouping of related mathematics standards within a given domain (e.g., grade 4, “Generalize and understanding of place value for multi-digit numbers”); clusters are an effective means of communicating the focus and coherence of the mathematics standards because they provide an appropriate gain size for following the contours of important progressions in the standards across grades.

Common Core State Standards (CCSS): Academic standards in mathematics and English language arts (ELA)/literacy adopted by Oregon in 2010. The CCSS identifies learning goals that students should know and be able to do at the end of each grade.

Content Specifications: “bridge documents” document developed by Smarter

¹ AERA, APA, & NCME, 2014.

Balanced that outlines the Consortium’s interpretation and priority of the knowledge and skills identified in the Common Core State Standards that are intended to be measured by the ELA/literacy and mathematics assessments.

Depth of Knowledge (DOK): four descriptive levels of cognitive complexity from the Cognitive Rigor Matrix². The Cognitive Rigor Matrix integrates Bloom’s (revised) Taxonomy of Educational Objectives and Webb’s Depth-of-Knowledge Levels. DOK and the Cognitive Rigor Matrix are tools to help clarify the concept of rigor.

Domain: a larger grouping of related mathematics standards (e.g., Operations and Algebraic Thinking, Number and Operations—Base Ten).

Hand Scored: student responses that are scored utilizing rubrics that require human judgement. Items such as constructed response and items that are dependent on previous responses are hand scored.

Interim Assessment: provides mid-point or period evidence of performance against a defined learning target. Results can be meaningfully aggregated and reported alongside summative.

Item Specifications: documents that provide guidance specific to writing Smarter Balanced items; separate documents exist for English language arts (ELA)/literacy and mathematics. Each document includes a table for each claim and target combination expected to be addressed by the summative assessment.

Evidence Centered Design: each item elicits evidence that can be used to support a claim about the students’ knowledge, skills, and/or abilities that are the targets of assessment. The assessment incorporates clear definitions of knowledge and skills, applies universal design principles that accurately measure the targeted knowledge, skills, and abilities of all students, and provides accessibility features to better meet specific needs.

Evidence Statements: description of the specific knowledge and skills that an item or task elicits from students.

Formative Assessment: process used by teachers and students during instruction to provide feedback to adjust ongoing teaching and learning to improve students’ achievements of learning targets.

Machine Scored: student responses that are scored utilizing rubrics programed into the testing delivery system. Items such as hotspot, drag-and-drop, text highlighting, equation editor, graphing, matching tables and selected response are machine scored.

Mathematical Practice: a balanced combination of procedure and understand that describes the ways in which developing student practitioners of mathematics increasingly ought to engage with the subject matter as they grow in mathematical maturity and expertise throughout the

² Hess, Carlock, Jones, & Walkup, 2009.

elementary, middle, and high school years (e.g., “Make sense of problems and persevere in solving them,” “Attend to precision”).

Performance Task: Smarter Balanced Assessment item type that involves significant interaction of students with stimulus materials and/or engagement in a problem solution, ultimately leading to an exhibition of the students’ application of knowledge and skills, often in writing or spoken language; stimuli include a variety of information forms (e.g., readings, video clips, data) as well as assignment or problem situation.

Range Achievement Level Descriptors: mathematics target level evidence that describe the content and rigor of an assessment target across the performance levels.

Reliability/Precision: “The degree to which test scores for a group of test takers are consistent over repeated applications of a measurements procedure and hence are inferred to be dependable and consistent for an individual test takers; the degree to which scores are free of random errors measurement for given group”³

Rubric: statements of evidence required of the student response to earn score point(s).

Scoring Guide: Smarter Balanced documents provided to calibrate raters and score practice tests

Stimulus: a reading selection, sound clip, or other type of visual or audio presentation that serves as the basis for a series of test items.

Strand: a larger grouping of related English language arts (ELA)/literacy standards. (e.g., Reading, Writing, Speaking and Listening, Language).

Summative Assessment: provides culminating evidence of performance against a defined learning target.

Target: detailed information about the knowledge, skills, and abilities to be assessed by the items and tasks within each claim as well as the depth of knowledge (cognitive demand) required; targets represent the prioritized content for summative assessment.

Test Blueprint: series of documents that together describe the content and structure of an assessment; these documents define the total number of items/tasks for any given assessment component, the standards measured, the item types, and the point values for each.

Test Item: a question, problem or task on a test. Items may take different forms such as constructed response, selected response (multiple choice single or multiple correct), matching tables, drag-and-drop, hot spot, graphing, and equation editor.

Universal Design: a research-based framework for designing curricula (educational goals, assessments, methods, and materials) that enable all individuals to gain knowledge, skills, and

³ AERA, APA, & NCME, 2014.

enthusiasm for learning. This is accomplished by simultaneously providing rich supports for leaning and reducing barriers to the curriculum, while maintaining high achievement standards for all students. In the context of assessment, this means, “An approach to assessment development that attempts to maximize the accessibility of a test for all of its intended test takers.”⁴

Validity: “The degree to which accumulated evidence and theory support a specific interpretation of test scores for a given use of a test. If multiple interpretations of a test score for different uses are intended, validity evidence for each interpretation is needed.”⁵

⁴ AERA, APA, & NCME, 2014.

⁵ AERA, APA, & NCME, 2014.