Title: Smarter Balanced Assessment: Determining Summative Claim-Level Performance Categories

Slide 1: Title Slide

Welcome to this presentation on the Smarter Balanced Assessment: Determining the Claim-Level Performance Categories.

Students who take a Smarter Balanced Summative Assessment in either English Language Arts or Mathematics receive an overall vertical scale score in that subject. They also receive one of three performance categories for each claim within each subject. The three categories are Above Standard, At or Near Standard, or Below Standard. But how are those claim-level performance categories determined? This presentation will answer that question.

Slide 2: Three Important Topics

To understand how the claim-level performance categories are determined, we first need to become familiar with three important topics the Smarter Balanced Vertical Scale, the Smarter Balanced Achievement Levels, and the standard error of measurement.

Slide 3: The Smarter Balanced Vertical Scale

Let's start by reviewing the Smarter Balanced Vertical Scale.

Each student who attempts the Smarter Balanced Summative Assessment in a subject receives an overall vertical scale score. The vertical scale score is the basic unit of reporting. It allows for fair comparisons at both the individual student level and the aggregate level. This scale ranges from approximately 2100 to 2800 for Grades 3 through 8. The scale is a vertical scale, and student performance in all grades is reported on the same scale. This allows us to compare a student's scale score from a test in one grade, to that student's scale score from a test in another grade.

The Smarter Balanced Summative Assessment is a computer adaptive test. The test adjusts the questions it presents to students based on a set of criteria on how students are performing on the test. Students who are performing well will tend to see harder questions, while students who are not doing so well will tend to see easier questions. Though different students are presented with different test questions, psychometric procedures are used to ensure that a given scale score represents the same level of performance. To further assist with the interpretation of results, the scale scores within a grade are divided into four

Achievement Levels: 1, 2, 3, and 4. Level 4 is the highest level. Levels 3 and 4 are the desired levels of performance.

Notice how the overall score range for each grade is steadily increasing. Notice also that the threshold scores between each level are increasing across the grades. Those are important aspects of a vertical scale.

Slide 4: Smarter Balanced Overall Scale Scores for ELA

Scale scores are the basic units of overall reporting for the Smarter Balanced Summative Assessments and Interim Comprehensive Assessments (ICAs). These scores fall along a continuous vertical scale (from approximately 2000 to 3000) that increases across grade levels. Scores are used to describe an individual student's current level of achievement. They can also be used to track growth over time. When aggregated, scale scores are used to describe achievement for different groups of students.

This table presents the scale scores for ELA for all grades tested in Connecticut. The threshold vertical scale scores that separate one level from the next are the lower numbers in each range for Levels 2, 3, and 4. For example, the threshold score for Level 3 in Grade 3 ELA is 2432.

These threshold, or cut scores, were developed based on feedback from thousands of K–12 educators, higher education faculty, psychometric experts, parents, and other stakeholders through a process called standard setting. All Smarter Balanced states, including Connecticut, have adopted these achievement levels and their corresponding threshold scores.

Slide 5: Smarter Balanced Overall Scale Scores for Math

For Math, this table presents the precise vertical scale score ranges for each achievement level in each grade.

The threshold vertical scale scores that separate one level from the next are the lower numbers in each range for Levels 2, 3, and 4. For example, the threshold score for Level 3 in Grade 3 Math is 2436.

Slide 6: The Standard Error of Measurement (SEM)

Let's review the last topic before we move into the claim-level performance categories: the standard error of measurement (SEM).

Psychometric theory tells us that a test score is an estimate of a student's achievement and comes with a certain amount of measurement error. This is why, in the Individual

Student Report shown here, we indicate that a student's test score can vary if the student were to take the test multiple times. Each time a student takes a Smarter Balanced test, psychometric procedures are used to calculate not only the vertical scale score, but also the standard error of measurement for each student. Since this measurement error is known, the Individual Student Report also provides the range of scores the student would earn if that student were to take the test multiple times. This range represents one standard error of measurement above and below the student's scale score. In this sample report, the student's overall vertical scale score in ELA is 2495, and the standard error of measurement is 10; so, the range is between 2485 and 2505.

Slide 7: Smarter Balanced Math Claims

Now let's look at the Smarter Balanced Mathematics Claims.

Claims are broad evidence-based statements about what students know and can do as demonstrated by their performance on the assessments. There are four claims in each subject but for reporting purposes in Connecticut, Claims 2 and 4 in both subject areas are combined into one reporting category. As a result, the three Mathematics Claims for reporting purposes are Concepts and Procedures, Problem Solving and Modeling and Data Analysis, and Communicating Reasoning.

Slide 8: Smarter Balanced ELA Claims

Now let's look at the Smarter Balanced ELA Claims. Similar to the Math Claims we just reviewed, there are four claims in ELA but for reporting purposes in Connecticut, Claims 2 and 4 are combined into one reporting category.

The three ELA Claims for reporting purposes are Reading, Listening, and Writing and Research/Inquiry.

Slide 9: ELA Claim-Level Performance Categories

For each claim, students are assigned to one of three performance categories: Below Standard, At or Near Standard, or Above Standard.

To determine these performance categories, the first step in the psychometric process is to identify the subset of test items that are related to the particular claim that were taken by the student. Using that subset of items, a student's claim-level scale score and claim-level standard error of measurement are calculated. These psychometric calculation procedures are identical to those used to calculate the overall vertical scale score, except in this instance, the calculation is based on the subset of items related to that particular

claim for that student. These claim-level scale scores are calculated and placed on the same scale as the overall score that ranges from approximately 2100 and 2800.

Students are then provided with a Claim-Level Performance Category for each of the three ELA Claims as seen above: Reading, Listening, and Writing and Research Inquiry.

Slide 10: Math Claim-Level Performance Categories

Students are then provided with a Claim-Level Performance Category for each of the three Math Claims as seen above: Concepts and Procedures, Problem Solving and Modeling and Data Analysis, and Communicating Reasoning.

Slide 11: Grade 3 ELA Claim-Level Categories

Now that we have the claim-level scale score and standard error of measurement for each student, let's look at an example to see how those scale scores and standard errors are used to determine the claim-level performance category. This figure represents scale scores for seven students in Grade 3 English Language Arts. The scale scores are represented by the blue dots. The vertical lines above and below the blue dots represent the confidence interval for the scale score for each student. Because the SEM represents the extent of uncertainty in a student's scale score, this confidence interval is established as 1 times the standard error of measurement for each student, both above and below the overall scale score for that student. The black horizontal line represents the Grade 3 English Language Arts overall threshold score that divides Achievement Level 2 from Level 3. This score is 2432; it is the minimum overall vertical scale score needed on the Grade 3 ELA test for a student to be classified in Achievement Level 3.

If a student's scale score and confidence interval are entirely below the horizontal line, then that student is said to be performing "Below Standard" on that Claim. See Student G.

If a student's scale score and confidence interval touch the horizontal line, then that student is said to be performing "At/Near Standard" on that Claim. See Students A, B and C.

If a student's scale score and confidence interval are entirely above the horizontal line, then that student is said to be performing "Above Standard" on that Claim. See Students D, E and F.

Slide 12: Thank you!

I hope this presentation has provided you with an understanding of how the claim-level performance categories for the Smarter Balanced Summative Assessments are

determined. For additional information, please refer to the Smarter Balanced Interpretive Guide or contact our office. Thank you!