

2014-2015 Smarter Balanced Scoring and Reporting Factsheet

1. How are partially complete tests handled in participation reports and score reports?

Below are the Smarter Balanced rules for calculating participation and performance. States may elect to implement the Consortium rules regarding calculating participation and achievement rates, but are not required to do so:

- **Participation Reports:** Students are reported as having “participated” in the test if they logged in to both the Performance Task and Non-Performance Task parts of the test, even if they did not answer any items. For computer-based Summative tests, the Non-Performance Task is a Computer Adaptive Test (CAT).
- **Score Reports:** For a student’s score to be reported, the student must have answered at least one Non-Performance Task (i.e. CAT) item **and** one Performance Task item.

2. What is a scale score?

- Scale scores are the basic units of reporting. These scores, which fall along a continuous vertical scale (from approximately 2000 to 3000) that increases across grade levels, can be used to illustrate students’ current level of achievement and their growth over time in a relatively fine-grained fashion.
- When aggregated, scale scores, unlike raw scores, can also describe school- or district-level changes in performance on the tests and can measure gaps in achievement among different groups of students.

3. What is the Standard Error of Measurement (SEM)?

The purpose of the Standard Error of Measurement (SEM) is to show the score range that a student would likely fall within if they took the same Smarter Balanced ELA or math test multiple times with exactly the same level of knowledge and preparation. For example, as seen in the figure below, a scale score of 2535 ± 22 (circled in red) indicates that if the student could take the same test multiple times, they would likely score between 2513 and 2557. Scale scores will vary based on the test and on the student.

Figure 1: Example of a Student Listing in the Online Reporting System that displays both scale scores and SEM.

Scale Score and Achievement Level Smarter Summative ELA/Literacy Grade 3 Test for Students in demo roster			
Name	EDUID	Scale Score	Achievement Level
Ia, Fi M.	999990001	2535 ± 22	Level 4
La, Fi M.	999990003	2480 ± 9	Level 4

4. Why is Standard Error of Measurement important?

The Standard Error of Measurement (SEM) is important because a student's score is best interpreted when recognizing that the student's knowledge and skills fall within a score range and not just a precise number. All test results, including scores on tests and quizzes designed by classroom teachers, are subject to Standard Error of Measurement.

5. What do achievement levels represent and why are they useful?

- Achievement levels are categories that describe what a student knows and is able to do, based on the student's scale score.
- A higher score on the test reflects a greater accumulation of knowledge, skills, and processes. A high score will place a student in a high achievement level.
- The achievement levels on the Smarter Balanced Summative and ICA tests are Level 1, Level 2, Level 3, and Level 4. To meet state standards, a student must be in either the Level 3 or Level 4 achievement level.

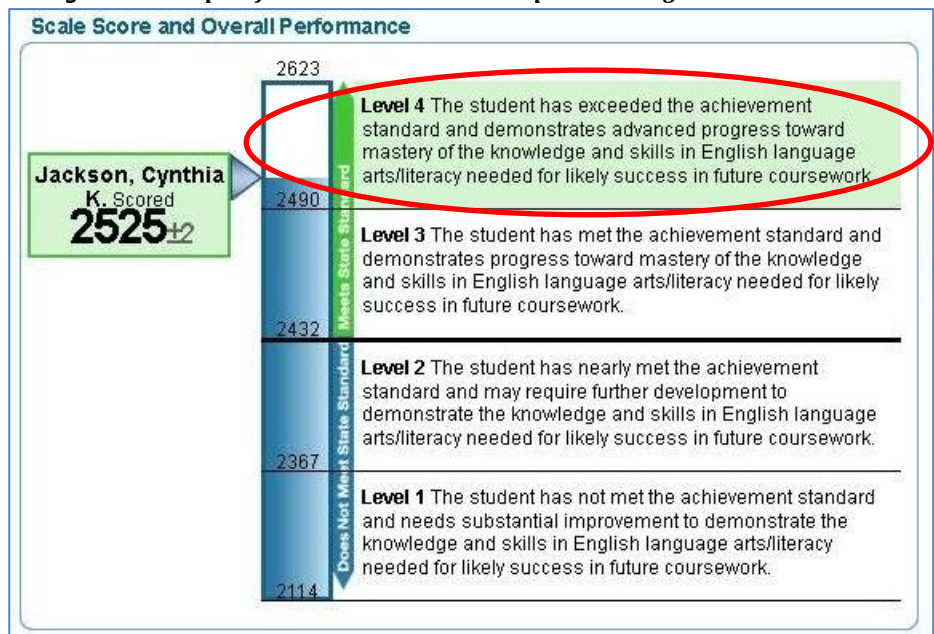
6. What are achievement level descriptors?

- Each achievement level is described by an Achievement Level Description, or ALD that describes a student's overall performance.
- For example, if a student earns a scale score of 2525 on a grade 10 math test, her score is higher than the threshold for Level 4, which places her in the Level 4 achievement level (see Figure 2 below). Level 4 is the highest achievement level of the Smarter Balanced Summative test, indicating that the student has "exceeded the achievement standard and demonstrates advanced progress." The full achievement level text is circled in the example image for grade 10 math.

- ALDs communicate the meaning of test scores by specifying, in content terms, the knowledge and skills that students display at four levels of achievement.
- Achievement Level Descriptors are cumulative, where the knowledge, skills, and processes of lower level ALDs are assumed by the higher level ALDs. For instance, the Level 4 student in the above example is assumed to be able to possess the knowledge, skills, and processes described in Levels 1, 2 and 3.

- Note from this example that the achievement levels are not always spaced evenly apart.

Figure 2: Example of an Individual Student Report showing achievement levels.



7. Who determines where one achievement level ends and the next begins?

The scores that separate achievement levels from one another are called threshold scores. Threshold scores and achievement levels are developed by thousands of K-12 educators, higher education faculty, experts, parents and other stakeholders in a process called standard setting.

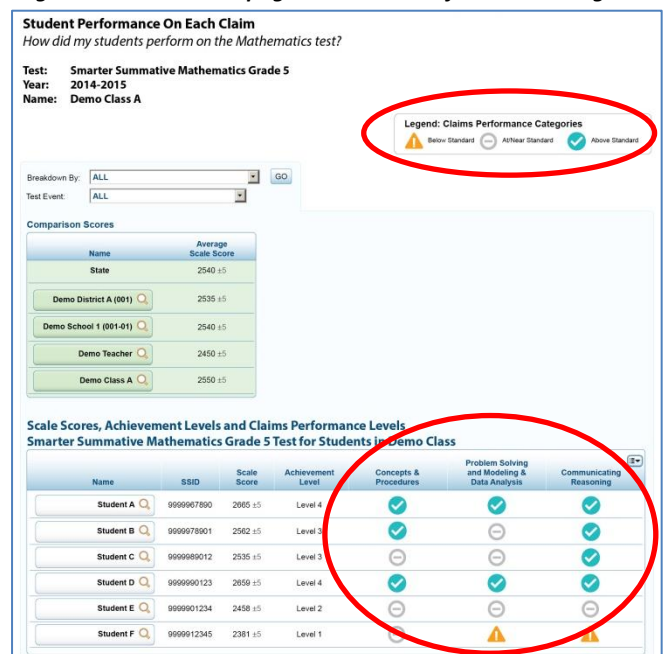
8. What are assessment claims?

- Assessment Claims are broad evidence-based statements about what students know and can do as demonstrated by their performance on the assessments.
- For example, there are three assessment claims on the Smarter Balanced Summative Grade 5 Math test: Concepts and Procedures, Problem Solving and Modeling & Data Analysis, and Communicating Reasoning.
- In addition to receiving a scale score and achievement level for the whole test, students are also placed into performance categories for specific claims.

9. What are the claim performance categories and how are they derived?

- For each claim, students are assigned to one of three claim performance categories: Below Standard, At/Near Standard, or Above Standard.
- In ELA, claim performance categories will be reported for each claim. In math, claim performance categories will be reported for Claim 1, Claim 2 and 4 combined, and Claim 3 (see Figure 3 at right for an example of a Claims Detail page).
- A student's claim performance category for each claim is derived based on their performance on the items contained in that claim.
- For example, a student may receive a scale score on the Smarter Balanced Summative Grade 3 ELA test that places him or her in the Level 3 performance level (meeting state standards). The student may have performed "Above Standard" on the Reading and Writing claims, "At/Near Standard" on the Listening claim, and "Below Standard" on the Research/Analysis claim. These claim performance categories contain useful information about math and ELA strengths and weaknesses for the student's teachers and parents.

Figure 3: Claims Detail page with Claim Performance Categories



11. How do claims scores relate to overall scores?

A student's scale score and achievement level are based on the student's performance on the test as a whole, while a student's claim performance category is based only on the student's performance on the test items contained in that particular claim (please see page 6 for a chart showing the relationship between the test subject, claims, and targets).

12. What are assessment targets?

- Targets are narrowly defined skills that are tied directly to the Common Core standards. Because of the breadth in coverage of the individual claims, the targets within them are needed to define more specific performance expectations within claim statements.

- For example, the “Reading” claim on the Smarter Balanced Summative ELA test includes targets such as “(Informational Text) KEY DETAILS: Use explicit details and implicit information from the text to support answers or inferences about information presented.” Targets are reported only at the aggregate level (district, school, roster) and not for individual students.

13. What are strength and weakness indicators?

- Targets are assigned “strength and weakness indicators” instead of achievement levels or performance levels.
- Unlike achievement levels and claim performance categories, which are assigned to students **based on state standards**, strength and weakness indicators display a group of students’ performance on a target **relative to the group’s performance on the test as a whole** (see the table above for a description of all the indicators).
- Unlike achievement levels based on overall test performance, these strengths and weaknesses do not imply proficiency. Instead, they show how a group of students’ performance is distributed across the target content relative to their overall subject performance on a test.

Icon	Target Level	Description
	Better than performance on the test as a whole	This target is a relative strength. The group of students performed better on items from this target than they did on the rest of the test as a whole.
	Similar to performance on the test as a whole	This target is neither a relative strength nor a relative weakness. The group of students performed about as well on items from this target as they did on the rest of the test as a whole.
	Worse than performance on the test as a whole	This target is a relative weakness. The group of students did not perform as well on items from this target as they did on the rest of the test as a whole.
	Insufficient Information	Not enough information is available to determine whether this target is a relative strength or weakness.

14. How are strengths and weaknesses for targets determined?

- In math, target scores will be computed for the targets in Claim 1 only. In ELA, target scores will be computed for the targets within each Claim.
- Strengths and weaknesses are reported for groups of students based on whether there is a statistically significant difference between that group’s performance on each target compared to their performance on the test as a whole (see Figure 4 at right for an example of a Target Detail page).
- For example, a group of students may have performed very well in a subject, but performed slightly lower in several targets. Thus, the minus sign for a target does not imply a lack of proficiency. Instead, it simply communicates that these students’ performance on that target was statistically lower than their performance across all other targets put together. Although the students are doing well, an educator may want to focus instruction on these areas.

Figure 4: Target Detail page showing strength and weakness indicators.

Legend: Strength and Weakness Indicator

Better than performance on the test as a whole Similar to performance on the test as a whole
 Worse than performance on the test as a whole Insufficient Information

Performance on Each Target
Smarter Summative ELA/Literacy Grade 3 Test for Student (School)

Target	Performance Level
Reading	
(Informational Text) KEY DETAILS: Use explicit details and implicit information from the text to support answers or inferences about information presented.	
(Informational Text) CENTRAL IDEAS: Identify or summarize central ideas/ key events, or procedures and details that support them.	
(Informational Text) WORD MEANINGS: Determine intended meanings of words, including domain-specific (tier 3) words and academic (tier 2) words with multiple meanings, based on context, word relationships, word structure (e.g., common roots, affixes), or use of resources (e.g., beginning dictionary, glossary)	
(Informational Text) REASONING & EVIDENCE: Use supporting evidence to interpret and explain how information is presented or connected within or across texts (author’s point of view, ideas and supporting details, relationships)	

15. Are summative and interim tests reported any differently?

Yes, there are differences in how interim and summative tests are reported. There are two types of Smarter Balanced interim assessments: Interim Comprehensive Assessments (ICA) and Interim Assessment Blocks (IAB). Below are the reporting features that are unique to the Smarter Balanced Summative, ICA, and IAB tests:

- **Summative tests** allow one opportunity per student. Summative reports include scale scores, achievement levels, claim performance categories, and target strength and weakness indicators (see Figure 5a below for an example of a Smarter Balanced Summative Individual Student Report).
- **ICA tests** use the same blueprint as summative tests and are reported in nearly the same way, except that scores are shown for multiple student opportunities if the student has taken the test multiple times, and target-level information is not reported (see Figure 5b below for an example of a Smarter Balanced ICA Individual Student Report).
- **IAB tests** consist of 5-17 individually scored blocks per subject per grade. Students may be administered as many or as few blocks as necessary, and they may have multiple opportunities. IAB reports look different from Summative/ICA reports, because instead of a scale score and achievement level, students receive a Block Performance Level for each block tested (see Figure 5c below for an example of a Smarter Balanced IAB Individual Student Report).

Figure 5a: Individual Student Report for a Summative test.

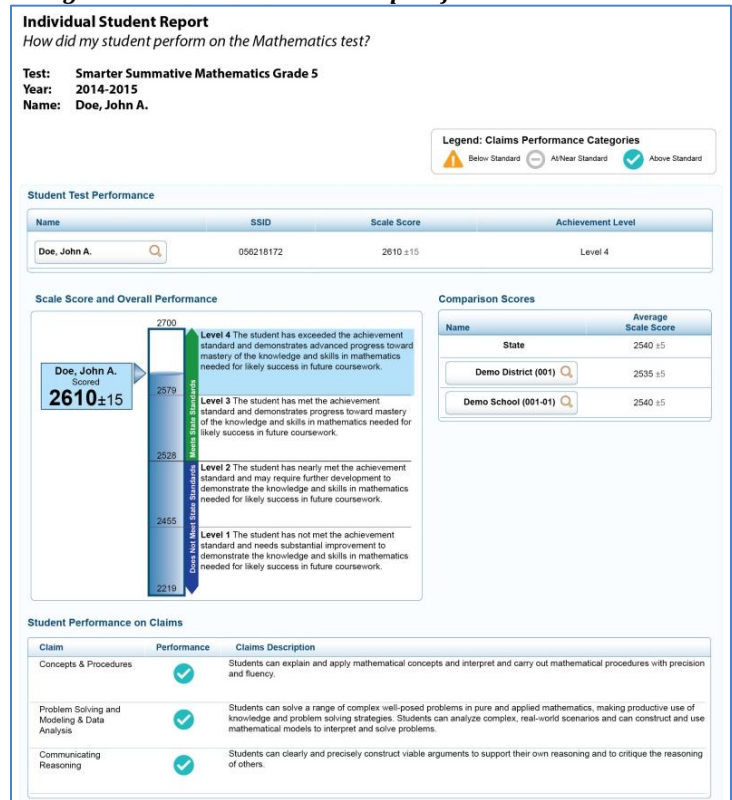


Figure 5b: Individual Student Report for an ICA test.

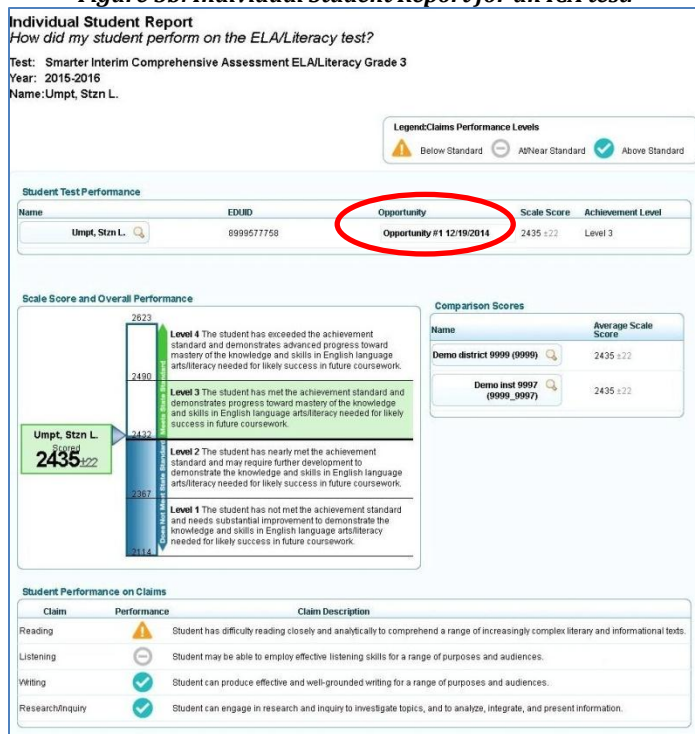


Figure 5c: Individual Student Report for an IAB test.

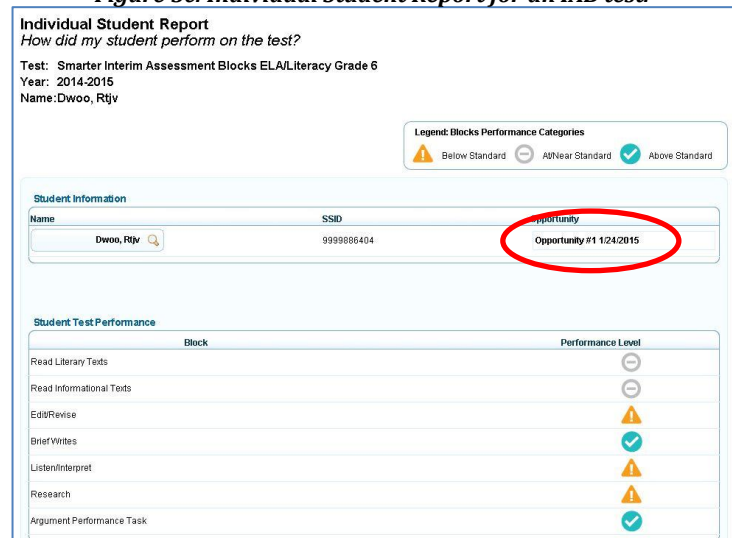


Figure 6: Illustration of the relationship between test subject, claims, and targets.

