
Chapter 1. General Information

CALIFORNIA
Assessment of Student Performance and Progress

CAASPP Post-Test Guide
Technical Information for Student Score Reports of
the Summative Assessments
for CAASPP LEA and Test Site
Coordinators and Research Specialists

2016-17 Administration

Smarter Balanced for English Language Arts/
Literacy and Mathematics Summative Assessments
California Alternate Assessments for English
Language Arts/Literacy and Mathematics
Standards-based Tests in Spanish for Reading/
Language Arts

  

Measuring the Power of Learning.™

Posted June 2017
Prepared by Educational Testing Service ©

New in 2016–17

Table I.1 What’s New in 2016–17

Feature	Change
California Assessment of Student Performance and Progress (CAASPP) System of Assessments	
California Science Test (CAST) and CAA for Science	During the 2016–17 test administration, students in grades five and eight, and ten, eleven, or twelve (depending on the high school grade level assigned to each high school) participated in pilot tests in science. The pilot test is meant to evaluate test questions, as well as help students and schools become familiar with new standards and content. There are no student scores for the science pilot.
Policy	
Administration of the Classroom Activity	Performance tasks for the Smarter Balanced Summative Assessments in California did not involve a Classroom Activity. All performance tasks include sufficient built-in supports to ensure that students are able to demonstrate what they know and can do based on the resources provided in the performance task itself.
CAASPP Student Score Reports	
CAASPP Smarter Balanced Student Score Reports	Student Score Reports for the Smarter Balanced Summative Assessments in grades four through eight include the student’s score history with up to two previous years’ results shown alongside the current year’s results.
CAA Student Score Reports for grades four through eight	Student Score Reports for the CAAs in grades four through eight include the student’s score history with the previous year’s results shown alongside the current year’s results.
Science	While student scores were not calculated for the California Science Test (CAST) and CAA for Science pilot tests, a section describing this testing milestone is included on the CAST and CAA Student Score Reports for grades five, eight, and, if assigned, grade eleven. Separate information sheets on either the CAST or CAA were printed and shipped for students in grades ten and twelve who were assigned to take the pilot.

Introduction

Additional Resources:

- California Department of Education (CDE) Data Privacy Web page—
<http://www.cde.ca.gov/ds/dp/>
- CDE California Assessment of Student Performance and Progress (CAASPP) System Web page—<http://www.cde.ca.gov/ta/tg/ca/>
- CDE Early Assessment Program (EAP) Web page—
<http://www.cde.ca.gov/ci/gs/hs/eapindex.asp>
- CDE Matrix One Web page—<http://www.cde.ca.gov/ta/tg/ai/caasppmatrix1.asp>
- CAASPP Results (also called the CAASPP Public Reporting Web site)—
<http://caaspp.cde.ca.gov/>
- CAASPP Post-Test Workshops: Connecting Assessments to Instruction—
linked at <http://www.caaspp.org/training/caaspp/index.html>
- *Online Reporting System User Guide for California*—
<http://www.caaspp.org/rsc/pdfs/CAASPP.ORS-guide.2016-17.pdf>
- Smarter Balanced Assessment Consortium Reporting Scores Web page—
<http://www.smarterbalanced.org/assessments/scores/>
- Smarter Balanced Assessment Consortium (in collaboration with the CDE)
California-specific Test Score Guide Web page—
<http://testscoreguide.org/ca/>
- California State University Success Web site—<http://CSUSuccess.org/>

Purpose of the Reports and Using the Results

In 2016–17, the California Assessment of Student Performance and Progress (CAASPP) summative online tests were administered for English language arts/literacy (ELA) and mathematics to California students in grades three through eight and grade eleven as part of California’s membership in the Smarter Balanced Assessment Consortium. These tests were also available in paper-pencil testing (PPT) versions to students in local educational agencies (LEAs) that could not offer these assessments electronically.

Students whose individualized education program (IEP) teams designated the use of an alternate assessment on statewide assessments and who have a cognitive disability that prevents him or her from taking the online CAASPP Smarter Balanced Summative Assessments were assigned to take the California Alternate Assessments (CAAs).

In addition, Spanish-speaking English learners (ELs) in grades two through eleven took an optional PPT in reading/language arts (RLA).

Results for tests within the CAASPP System are used for two primary purposes:

1. Communicating students' progress in achieving the state's academic standards to students, parents/ guardians, and teachers.
2. Informing decisions, along with local assessment data, that teachers and administrators make about improving the educational program.

Overview of Online Smarter Balanced Assessments

The Smarter Balanced Online Summative Assessments for ELA and mathematics are available to students in grades three through eight and grade eleven. These assessments are aligned to the Common Core State Standards in their respective content areas and are intended to measure student progress toward college and career readiness. Student test results are reported in the following overall achievement levels:

- Level 4—Standard Exceeded
- Level 3—Standard Met
- Level 2—Standard Nearly Met
- Level 1—Standard Not Met

These achievement levels were determined by a standard-setting process. Information on the process can be found on the [Reporting Scores](#) Web page of the Smarter Balanced Assessment Consortium Web site.

Each content area of the online assessments consists of a computer adaptive test (CAT) as well as a performance task (PT). Summary results are available online, in the secure Online Reporting System (ORS), first as preliminary results and then, when scores have been received and/or updated for all students, as final results.

Most student responses are machine-scored, while other responses to questions are hand-scored. A student's results from the scores from the CAT and PT are combined to determine an overall scale score for that student. Scores of the test are based on the specific test questions with different difficulty levels that students responded to, instead of the sum of the number correctly answered. See the subsection [Scale Scores for the CAASPP System](#) for more information about the process used to determine a scale score.

In addition to achievement levels for the total test, claim achievement categories are also reported, as above standard, near standard, and below standard. These categories were identified after the standard setting for the total-test achievement levels by using the distance a student's performance on the claim is from the Level 3 "Standard Met" achievement level criterion.

Claim achievement categories are based on a smaller collection of items. This makes it more difficult to provide information about a student's claim achievement level without increasing

Privacy of Student Records Collected and Maintained by the CDE for CAASPP

To meet its statutory responsibilities, the California Department of Education (CDE) collects and maintains personally identifiable information from the education records of California students. Additional information on the CDE data privacy policies may be found on the [Data Privacy](#) Web page.

The CDE and its CAASPP test administration contractor use student information for the purposes of meeting the CDE's statutory responsibilities. Neither the CDE nor its CAASPP test administration contractor sells student data or uses student data for any other purposes.

the amount of classification error—more claim achievement levels, coupled with very few items within a claim, will result in more students being misclassified as belonging to one achievement level when they actually belong to another. This classification error is lessened by reducing the number of claim achievement levels, to three.

The claim achievement category indicates that the score on a claim is one of the following:

- If the scale score of a claim is above the “Standard Met” achievement level on the total content-area test, the achievement category for the claim is “Above Standard.”
- If the scale score of a claim is at or near the “Standard Met” achievement level on the total content-area test, the achievement category for the claim is “Near Standard.”
- If the scale score of a claim is below the “Standard Met” achievement level on the total content-area test, the achievement category for the claim is “Below Standard.”

When presented in the Student Score Reports for parents/guardians, the scale scores and achievement levels of both the current year and for available previous years are included. For example, for students in grades five, six, seven, and eight, three years’ scores, including the scores from 2014–15, are presented.

For students in grade eleven, the following achievement statuses are associated with a level of readiness for college-level coursework:

- The “Standard Exceeded” achievement level suggests that the student is ready for college-level coursework in ELA and/or mathematics.
- The “Standard Met” achievement level suggests that the students is conditionally ready for college-level college courses in ELA and/or mathematics. However, he or she must take an approved English and/or mathematics course in grade twelve and receive a grade of C or better.

Details regarding the EAP can be found on the CDE [Early Assessment Program](#) Web page. Additionally, the [California State University Success](#) Web site has more information about the steps students can take in grade twelve to be ready for college.

Computer Adaptive Test (CAT)

A **CAT** is designed to adjust the level of item difficulty, based on the responses provided, to match the ability of a student. By adapting to the student’s ability as the assessment is being taken, the CAT presents an individually tailored set of questions that is appropriate to each student and provides more accurate scores for all students across the full range of the achievement continuum. A CAT requires fewer questions as compared to a fixed-form assessment—that is, a test where students are given the same questions regardless of the student’s responses or ability—to obtain an equally precise estimate of a student’s ability.

During the test, if a student gives a wrong answer, the computer will follow up with an easier question; while if the student answers correctly, the next question will be slightly more difficult. Since the answers of items used to estimate the student’s ability are machine-scored, the correctness of the student’s response can be known immediately, and the successive items are selected to adapt to the current ability of the student. This process continues until the test content outlined in the test’s blueprint is covered.

The CAT requires a large pool of test questions statistically calibrated on a common scale to cover the ability range.

Performance Task (PT)

A **PT** is a nonadaptive form designed to provide students with an opportunity to demonstrate their ability to apply their knowledge and higher-order thinking skills to explore and analyze a complex, real-world scenario. It is a required portion of the test. PTs are not targeted to students' specific ability levels.

Overview of the California Alternate Assessments

The California Alternate Assessments (CAAs) for ELA and mathematics are online, summative, grade-level assessments for students whose IEP teams designate the use of an alternate assessment. The CAAs give students the opportunity to demonstrate their achievement of the Core Content Connectors (Connectors), which are derived from the Common Core State Standards (CCSS), by taking a test commensurate with their abilities. The Connectors are the alternate achievement standards assessed on the CAAs. The Connectors take the main achievement standards from the CCSS and make them more accessible for students with significant cognitive disabilities. Student test results are reported in the following overall achievement levels:

- Level 3—Alternate
- Level 2—Alternate
- Level 1—Alternate

These achievement levels were determined by a standard-setting process.

The CAAs are administered one on one by a test examiner reading scripted instructions to a student. At the start of testing, a test examiner administers a Student Response Check (SRC) using the first one to four items in the test to identify whether the student has a consistent and observable way of indicating responses to test items. For students who do not orient or provide an observable, consistent response, test examiners were directed to end the assessment.

Most student responses are machine-scored, while a few constructed response questions are scored by the test examiner at the time of testing. A student's results from the machine-scored and examiner-scored items are combined to determine an overall scale score for that student.

In addition to taking the CAAs for ELA and mathematics, students enrolled in or assigned to grades five and eight and a selected high school grade also take the CAA for Science pilot, for which results are not reported.

Overview of the Standards-based Tests in Spanish

These CAASPP tests are fixed-form, paper-pencil tests. Optionally, EL students in grades two through eleven take the Standards-based Tests in Spanish (STS) for RLA. Student test results are reported in the following performance levels:

- Advanced
- Proficient
- Basic
- Below basic
- Far below basic

Types of CAASPP Reports

Results for the CAASPP Summative Assessments are delivered in four ways, as follows:

<p>1. Reports in the ORS These reports are described in the Online Reporting System User Guide for CAASPP.</p>	<ul style="list-style-type: none"> • Home Page Dashboard • Subject Detail • Claim-level Detail • Assessment Target Reports • Listing (Group, Roster, Student) • Student Detail
<p>2. Student Results File in TOMS</p>	<ul style="list-style-type: none"> • Student Score Data Extract (Final data files will include participation data for the CAST and CAA for Science.)
<p>3. Student Score Reports These reports are printed and available as downloadable PDFs from the Test Operations Management System (TOMS).</p>	<ul style="list-style-type: none"> • Student Score Report for Smarter Balanced Summative Assessments for ELA and mathematics—Grades three, four, six, seven, and eleven (The Student Score Reports for students in grade eleven also includes information—but no score results—about the CAST for students assigned to take this test.) • Student Score Report for Smarter Balanced Summative Assessments for ELA and mathematics, and CAST (assignment and information only, no score results)—Grades five and eight • Student Score Report for CAAs for ELA and mathematics—Grades three, four, six, seven, and eleven (The Student Score Reports for students in grade eleven also includes information—but no score results—about the CAA for Science for students assigned to take this test.) • Student Score Report for CAAs for ELA and mathematics, and science (assignment and information only, no score results)—Grades five and eight • Student Score Report for CAST (information only, no score results)—Grades ten and twelve, if assigned • Student Score Report for CAA for Science (information only, no results)—Grades ten and twelve, if assigned • Student Score Report for STS for RLA—Grades two through eleven
<p>4. Aggregated Internet Reports (Internet reporting) These reports are available at the CAASPP Public Reporting Web site.</p>	<ul style="list-style-type: none"> • Smarter Balanced ELA scores • Smarter Balanced Mathematics scores • CAA ELA scores • CAA Mathematics scores • STS for RLA scores

Grades and Subjects Reported

CAASPP results are reported for the tests students took. Students who took the grade-level STS for RLA in addition to the required Smarter Balanced assessment(s) and/or the CAST will receive at least two reports, for example: one for the Smarter Balanced Summative Assessments and/or the CAST taken and another for the STS. The matrix in Table I.2 shows, for each grade, the test results that will appear on a report. Note that the optional STS was administered only to Spanish-speaking English learners.

Please note that the Student Score Report for Smarter Balanced Summative Assessments in grade eleven includes a section that associates the student’s overall score and achievement level for ELA and mathematics with the student’s Early Assessment Program (EAP) status, which is an early indicator of the student’s conditional readiness for college-level coursework. Note that the student’s results will not be sent to the California State University (CSU) and participating California Community Colleges (CCCs) unless the student opted to do so after completing the ELA assessment (for ELA results) and mathematics assessment (for mathematics results). Students who did not send their results to the CSU and participating CCCs at time of testing may provide those results upon request at a later date.

Table I.2 Reporting Matrix

Test Name	Grade Enrolled											
	2	3	4	5	6	7	8	9	10	11	12	
Smarter Balanced Online Summative Assessments												
English language arts/literacy	–	✓	✓	✓	✓	✓	✓	–	–	✓	–	
Mathematics	–	✓	✓	✓	✓	✓	✓	–	–	✓	–	
California Alternate Assessments												
English language arts/literacy	–	✓	✓	✓	✓	✓	✓	–	–	✓	–	
Mathematics	–	✓	✓	✓	✓	✓	✓	–	–	✓	–	
Science	–	–	–	◇	–	–	◇	–	⊗	◇	⊗	
Science												
CAST	–	–	–	◇	–	–	◇	–	⊗	◇	⊗	
Standards-based Tests in Spanish												
STS for RLA (grade level)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	

Legend

–	Assessment not available in this grade
✓	Assessment given and scores reported in this grade
◇	Assessment given but no scores reported in this grade; information about the pilot assessment provided on the Student Score Report (For grade eleven, assessment given if grade eleven students were assigned in that high school.)
⊗	Assessment given but no scores reported in this grade; information about the pilot assessment provided in an information sheet

A Note about Universal Tools, Designated Supports, and Accommodations

The “Universal Tools, Designated Supports, and Accommodations for the California Assessment of Student Performance and Progress” (Matrix One) Web document that displays the list of available CAASPP accessibility supports and the assessment(s) for which their use is approved is linked on the CDE [Matrix One: CAASPP](#) Web page. The four parts of Matrix One listing the universal tools, designated supports, and accommodations available in CAASPP testing are as follows:

1. Part 1—Embedded supports available only on computer-based tests
2. Part 2—Non-embedded supports available for use with the online tests as well as on all paper-pencil assessments (Smarter Balanced for ELA and mathematics and/or the STS for RLA)
3. Part 3—Instructional supports and resources available for the CAAs
4. Part 4—Approved unlisted resources

Universal tools are available to all students per student preference and selection; designated supports are available to students by teacher recommendation. Accommodations are available to students with documented need in a student’s IEP or Section 504 plan.

The use of universal tools, designated supports, or accommodations does not change the way scores are reported, although the score for a student who used an unlisted resource that changes the construct being measured might be noted (Student Score Report) or not counted because the student is listed as having a valid score (aggregated reporting).

Test administration for the CAAs allows for the examiner to provide any instructional support, identified in students’ IEP, needed by the students to access the test questions and tasks.

Embedded Universal Tools, Designated Supports, and Accommodations (Part 1 of Matrix One)

Embedded universal tools, designated supports, and accommodations are digital accessibility tools that are available in the online summative assessments and do not change the construct being measured. For example, the use of a digital notepad during any CAASPP online assessment does not change what is being measured.

Non-embedded Universal Tools, Designated Supports, and Accommodations (Part 2 of Matrix One)

Non-embedded universal tools, designated supports, and accommodations do not change the construct being measured. For example, the use of scratch paper during any CAASPP test does not change what is being measured.

Instructional Supports and Resources on the CAAs (Part 3 of Matrix One)

Students taking the CAAs, which are administered one on one by a test examiner, are provided with additional instructional and physical supports when testing. These may be in addition to the resources documented in the student's IEP or Section 504 plan. Examples of these include alternate text to describe illustrations and allowing the student test-taker to direct another person, such as an aide or the test examiner, to respond to assessment items.

Unlisted Resources (Part 4 of Matrix One)

“Unlisted resources” are non-embedded accessibility supports that are either listed in Part 4 of Matrix One or are not identified in Parts 1 or 2; they may have been previously identified as modifications or may be unlisted supports. Unlisted resources may fundamentally change what is being measured. All unlisted resources must be listed in the student's IEP or Section 504 plan; an online form in TOMS requesting their use should have been submitted to the CDE before a student was tested.

If the CDE determines the unlisted resource changes the construct being measured, the unlisted resource may be approved and used by the student and the student will receive a Student Score Report, but the student will not be counted as participating in statewide testing, which will impact the accountability participation rate indicator for the LEA (*California Code of Regulations*, Title 5, Section 854.9 [d] [1]).

Interpreting Results

Additional Resources:

- California Department of Education (CDE) California Assessment of Student Performance and Progress (CAASPP) Technical Reports and Studies Web page—<http://www.cde.ca.gov/ta/tg/ca/caaspprptstudies.asp>
- CDE Technical Reports and Studies for California Standardized Testing Program and the California Alternate Performance Assessment Web page—<http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp>
- Smarter Balanced Assessment Consortium Development and Design Web page—<http://www.smarterbalanced.org/assessments/development/>



Scale Scores for the CAASPP System

Note: After estimating the student’s overall ability, the score for a Smarter Balanced Summative Assessment is mapped onto the reporting scale through a linear transformation:



$$\begin{aligned} & \textit{Mathematics} \\ \text{scale score} &= 2514.9 + 79.3 * (\text{estimated ability}) \\ & \textit{English Language Arts/Literacy (ELA)} \\ \text{scale score} &= 2508.2 + 85.8 * (\text{estimated ability}) \end{aligned}$$

Scale scores are important measures for the CAASPP System. Student achievement or performance levels are assigned on the basis of scale scores for all tests.

The advantage of the scale score metric is that it allows a particular score (for example, 2533 on the Smarter Balanced Summative Assessment for mathematics assessment) to mean the same thing regardless of what items students took for a grade-level, content-area test. Scale scores provide a common reference over the years.

Each program/grade level/content area of the Smarter Balanced assessments and California Alternate Assessments (CAA) has its own scale score range.

Teachers and administrators should not use CAASPP results in isolation to make inferences about instructional needs. Anyone using CAASPP results to identify strengths and weaknesses in instructional programs should be familiar with the cautions and procedures described in the next chapter, “Comparing Results.”

Equating and Scaling

When tests are constructed for each grade, every effort is made to make the tests parallel and of the same level of difficulty from one year to another. However, even with those efforts, small differences in test difficulty still exist between test forms. A psychometric procedure called equating makes adjustments for test difficulty so that students in one year are held to the same standards as students in another year.

Details about equating and scaling for the CAASPP System tests are described in each of the following technical reports:

- Smarter Balanced Summative Assessments for English language arts/literacy (ELA) and mathematics—*CAASPP Smarter Balanced Technical Report*
- California Science Test (CAST)—*CAST Pilot Summary Report*
- CAAs for ELA and Mathematics—*CAA Technical Report*
- CAA for Science—*CAA for Science Pilot 1 Summary Report*
- STS—*Standards-based Tests in Spanish Technical Report*

The technical reports for the CAAs for ELA and mathematics and the STS also include raw-score-to-scale-score conversions for the testing year.

The technical report for the CAASPP Smarter Balanced Summative Assessments and CAAs are linked on the CDE [CAASPP Technical Reports and Studies](#) Web page.

The technical report for the STS is linked on the CDE [Technical Reports and Studies for California Standardized Testing Program and the California Alternate Performance Assessment](#) Web page.

Smarter Balanced Summative Assessments for ELA and Mathematics (Online and PPTs)

Final scores represent the ability estimates for students. Once the responses from the performance task (PT) and computer adaptive test (CAT) portions are merged for final scoring, the resulting ability estimates are based on the responses to the specific test questions that a student answered, not the total number of questions answered correctly. Higher ability estimates are associated with students who correctly answer more difficult and more discriminating questions; lower ability estimates are associated with students who correctly answer easier and less discriminating questions. Two students can arrive at the same scale score by very different paths. This type of scoring is called “item pattern scoring.”

The PPT versions of the Smarter Balanced Summative Assessments use the same scale as the online assessments to report student results.

Scale Score Ranges

Online assessments were scaled vertically, which means that scores for certain questions that were common between adjacent grades were linked. This will make it possible to monitor students’ year-to-year progress in assimilating the Common Core State Standards (CCSS) and to describe student progress over time across grade levels.

Scale scores offer a more precise way to determine students’ performance on the online assessments than achievement levels (which are described in the next subsection) because each level is based on a range of numbers, rather than an individual number like a scale score. Scale score ranges for the Smarter Balanced assessments, which vary from test to test and range from 2114–2795 in ELA and 2189–2862 in mathematics, are listed in [Appendix A](#).

Scale scores for the online summative assessments, in particular, were built on a common vertical scale for each content area, which allows meaningful comparisons between individual students and group comparisons between schools and local educational agencies (LEAs)

across grades within the same content area. Student achievement levels are assigned based on scale scores for all tests, which are described in the next subsection.

Achievement Levels

Smarter Balanced overall achievement levels are categorical labels given to particular scale score ranges. The achievement levels are Standard Exceeded, Standard Met, Standard Nearly Met, and Standard Not Met. The minimum and maximum scale scores for each achievement level vary for grade and content area. Achievement levels were set during a process called *achievement level setting*, which established the association between scores and their category of achievement. Achievement level setting also ensures that the achievement levels align to the CCSS.

CAAs for ELA and Mathematics

CAA scores reflect estimates of student ability that are based on which items a student correctly answers in a multistage adaptive test setting. A two-stage testing approach adapts the difficulty of a test to each student's ability in order to achieve more precise measurement. The first stage consists of a routing test that provides an initial student ability estimate. The second stage consists of a test that varies in difficulty depending on that initial ability estimate. A student whose initial ability estimate is high will respond to a second stage module consisting of difficult items that will help to determine just how high his or her ability is. A student whose initial ability estimate is low will respond to a second stage module consisting of less difficult items, and a student whose initial ability estimate is intermediate will respond to a second stage module consisting of items that are intermediate in difficulty. A student correctly responding to 15 difficult items will earn a higher CAA scale score than a student correctly responding to 15 less challenging items.

Scale Score Ranges

Scale scores are used in the evaluation of overall student achievement in the CAA because psychometric analyses underlying these scores account for the variations in difficulty for the questions that students are administered. If equivalent students were administered forms varying in difficulty, student scale scores would still be comparable.

Scale scores are associated with achievement levels that describe the underlying student achievement. The ranges of scale scores that are associated with each achievement level are held constant from year to year for each grade level and content area, while the number- or percent-correct score (i.e., the raw score) associated with each scale score may change.

Scale score ranges are listed in [Appendix A](#).

Achievement Levels

CAA overall achievement levels are categorical labels given to particular scale score ranges. The achievement levels from lowest to highest performance are Level 1—Alternate, Level 2—Alternate, and Level 3—Alternate. Regardless of the grade level—which is indicated by the first digit of the scale score—the minimum and maximum scale scores for each performance level are the same within each content area. Achievement levels were set during a process called standard setting, which established the association between students' scores and achievement category. Standard setting also ensures that the performance levels align to the CCSS PLDs.

STS Paper-Pencil Tests

Scale scores are used in the evaluation of overall student performance. Unlike raw scores (i.e., number-correct scores or percent-correct scores) that allow only comparisons between students under the same test setting, scale scores provide a common reference statewide, making interpretation easier. The scale score performance-level cut points are held constant from year to year for each grade level and content area, while the number- or percent-correct score (i.e., the raw score) associated with each scale score may change.

Because percent-correct scores are defined in terms of the number of questions answered correctly (the raw score metric) they are, by definition, associated with the specific form of the test taken, unadjusted for difficulty—that is, they are dependent on the difficulty of the test *questions* and the ability level of those who are taking the test.

Scale Score Ranges

The scale score ranges for the performance levels are found in [Appendix A](#).

The range of possible scale scores for the STS is from 150 to 600 for each grade and subject. The scale of 150–600 was selected before the first tests were scaled. When the tests were administered and scored for the first time after the performance standards were set, the number-correct scores were associated with scale scores.

Performance Levels

Performance levels for the STS are advanced, proficient, basic, below basic, and far below basic. The goal in California is to have all students perform at the proficient or advanced level.

The minimum scale score for the proficient level in all grades is set at 350. The basic level is set at a minimum scale score of 300. The minimum scale scores for below basic and advanced differ by content area and grade.

Reporting Clusters

Reporting clusters are groups of questions related to the same standard on a test. Reporting cluster scores are not reported in 2016–17 for the STS for RLA.

Smarter Balanced Claims and Assessment Targets

The Smarter Balanced content areas of ELA and mathematics are broken down into claims and assessment targets.

Claims are broken down into content categories, which contain a varying number of assessment targets. An assessment target defines the grade-specific knowledge, skill, or ability that students should know or be able to demonstrate within the domain.¹ For example, the overall claim “Reading” has a content category called “Literary” that contains an assessment target called “Reasoning and Evaluation.”

¹ California Department of Education. (2016.) “Assessment Target Reports Frequently Asked Questions.” Retrieved from <http://www.caaspp.org/rsc/pdfs/CAASPP.target-report-FAQs.2016.pdf>

Claims and their assessment targets are listed in [Appendix B](#). Please note that not all assessment targets are tested for all students given the adaptive nature of the CAT portion of the test.

Claims

Assessment claims are evidence-based statements about what students know and can do as demonstrated by their achievement on the summative assessments. They are defined in the item specifications for ELA and mathematics available on the Smarter Balanced Assessment Consortium [Development and Design](#) Web page. There are no assessment claims or claim scores for the CAAs.

There are four claims (but three reporting categories) per mathematics assessment and four claims per ELA assessment, each with a varying number of content categories (subcategories that may apply to some specific claims) and assessment targets.

Results for claims are presented for individual students on the Student Score Reports and in the Online Reporting System for schools, LEAs, and the state. Achievement on claims is reported as one of three levels:

- Above Standard
- Near Standard
- Below Standard

Achievement levels for claims are very similar to subscores. They provide supplemental information regarding a student's strengths or weaknesses. Only three achievement levels for claims were developed since there are fewer items within each claim.

A student's ability, along with the corresponding standard error, are estimated for each claim. Achievement levels for claims are based on the distance a student's performance on the claim is from the Level 3 Standard Met achievement level. Using the standard error, an interval estimate corresponding to the student's true performance on the claim is constructed and an interval defined. If the interval does not contain the Level 3 Standard Met criterion value for a particular claim, it would indicate a strength or weakness.

No achievement level-setting occurred for claims.

Assessment Targets

While the claims do not vary among grades, assessment targets for ELA Claims 1–4 and mathematics Claim 1 are unique to each grade. Note that assessment targets are reported for mathematics Claim 1 only, because “For mathematics Claims 2, 3, and 4, items are intended to emphasize the mathematical practices, and therefore, items may align to the content included in several mathematics assessment targets. The best common descriptors of the items included in these claims are the claim labels themselves.”¹

Assessment targets describe what is to be assessed within a claim and are used to develop test questions. Assessment targets are reported at the group level in the Online Reporting System (ORS) and provide information regarding a group's strengths and weaknesses relative to its achievement on the assessment as a whole. Assessment target reports show how a group of students performed on an assessment target compared to their overall achievement on the content-area assessment.

Strength and weakness indicators on assessment target reporting are as follows:

- 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

For example, while a group of students might have achieved the overall Level 3 Standard Met achievement level on the ELA assessment, their performance on the “Reasoning and Evaluation” assessment target might be “Worse than performance on the test as a whole.” This is not to say that the students lack reasoning and evaluation skills, only that their achievement on that portion of the assessment is not as high, statistically, as their achievement on the entire assessment.

Note, that like all results for the 2016–17 CAASPP administration in the ORS, assessment target report results are partial and may change as additional data are received.

Comparing Results

Additional Resources:



- California Assessment of Student Performance and Progress (CAASPP) Results (also called the CAASPP Public Reporting Web site)—
<http://caaspp.cde.ca.gov/>

Standard 12.10 of the *Standards for Educational and Psychological Testing* (2014) states, “In educational settings, a decision or characterization that will have major impact on a student should take into consideration not just scores from a single test but other relevant information.”²

Test results should be interpreted as a student’s achievement on a single assessment. They are meant to represent approximations of students’ mastery of content areas.

Any comparison of groups should not be used for diagnostic, placement, or promotion or retention purposes. Decisions about promotion, retention, placement, or eligibility for special programs may use or include CAASPP System results only in conjunction with multiple other measures including, but not limited to, locally administered tests, teacher recommendations, and grades.

Using the Conditional Standard Error of Measurement (CSEM) to Compare Scale Scores and Achievement Levels for the Online Summative Assessments for Individual Students

In any test, one can assume that scores for an individual would vary if it were somehow possible to give the same test over and over again. For example, students may vary in their performance because of the way they are feeling on the day of the test or they may be especially lucky or unlucky when they guess at questions they do not know. This random variation in individual scores is quantified through the use of a statistic of measurement precision called the conditional standard error of measurement (CSEM). CSEMs are available in the Online Reporting System (ORS) and the student data files.

Given a single score for a student, it can be assumed that if the student were to take the test over and over again, the student would score within plus or minus one CSEM of the observed score about 68 percent of the time. In the ORS, this idea is expressed as follows:

“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. For example, 2300 (+/-10) indicates a score range between 2290 and 2310.”

For the online assessments, an error band is a useful tool that describes the amount of precision associated with a reported scale score. CSEM is calculated for each student who takes the online assessments. In the 2016–17 reports, the averaged CSEM at each scale score point was used. Error bands are used to construct an interval estimate corresponding

² 2014, American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.

to a student's true ability/proficiency for a particular content area with a certain level of confidence.

Comparing Results for the Smarter Balanced Online Assessments

Because of the vertical scaling of the Smarter Balanced assessments, scale scores for a test may be compared to scale scores for the same student or groups of students in different years for the **same content area**, as well as for between specific grade levels and content areas. This allows users to say that achievement for a given content area and grade was higher or lower one year as compared with another. Scale scores for the Smarter Balanced assessments may be compared across grades since the scales are vertically aligned across grades.

Scores for the paper-pencil versions of the Smarter Balanced Summative Assessments are linear forms but have the same scale as the online tests.

In addition to the 2016–17 scores, results for the 2015–16 and the 2014–15 Smarter Balanced Summative Assessments are available. Aggregate results can be downloaded from the public [CAASPP Results](#) Web site as well as from the secure ORS (for properly credentialed users).

Results for any year are not included in aggregate reporting if the student met one of the following conditions during CAASPP Smarter Balanced testing:

- Not tested by parent/guardian request
- Not tested due to significant medical emergency
- Absent (note that this is not a condition used for 2015–16 or 2016–17 testing)
- Exempt from taking the English language arts/literacy (ELA) assessment
- Student completed only one part of the Smarter Balanced content area test
- Not tested

Scores are included on the Student Score Report with an asterisk indicator if the test results were invalidated for any of the following reasons:

- Invalidated appeal for online tests
- Student observed cheating (paper-pencil tests only)
- Student earned the lowest obtainable scale score (Smarter Balanced, CAAs) or lowest obtainable scale score +1 (CAAs only)
- Student has an approved unlisted resource request that changed the construct of the test

Comparing Achievement Results

When comparing results for the Smarter Balanced Summative Assessments, compare results only within the same content area or cohort; that is, compare grade five ELA in 2015–16 to grade five ELA in 2016–17 or grade seven mathematics in 2015–16 to grade eight mathematics in 2016–17.

Two types of comparisons are possible:

1. Comparing the average scale score; or
2. Comparing the percent of students scoring at each achievement level.

When making comparisons across years within a given grade and content area, it is important to understand that even when the number of students is the same, the group's composition from year to year may be quite different if student mobility (transiency) is high.

When comparisons are made across years, they are actually a comparison of different groups of students with different traits taking different tests. Generally, there will be more variance in scores from year to year when small numbers of students are tested.

Comparing ELA and Mathematics Scale Scores and Achievement Levels for Groups



Note: A scale score is derived from a statistical process. It is *not* possible to calculate a scale score by multiplying a student's percent correct in a content area with another number, such as 2795 for ELA or 2862 for mathematics.

An example of how *group-level* scale scores for 2016–17 may be compared to the 2015–16 scale scores for the same content area and grade is shown in Table I.3, below. In this table, hypothetical average scale scores (SS) for ELA are compared between 2015–16 and 2016–17 for the students in a particular school. In addition to comparisons for all students, similar grade-by-grade comparisons of scale scores may be made for different subgroups of interest.

Table I.3 Hypothetical Example of Using the Smarter Balanced for ELA to Measure Progress by Comparing Average Scale Scores

Grade	2015–16 Smarter Balanced for ELA		2016–17 Smarter Balanced for ELA		Difference
	No. of Students	Mean SS	No. of Students	Mean SS	
Grade 5	120	2440.0	111	2451.3	11.3
Grade 6	100	2510.0	124	2510.3	0.3
Grade 7	90	2590.0	102	2593.2	3.2

Table I.4 provides a second hypothetical example of how group-level Smarter Balanced results may be compared. In this example, the percent of students scoring at “Standard Met or Standard Exceeded” in mathematics are compared between 2015–16 and 2016–17 across grades for the same school. Comparisons between 2015–16 and 2016–17 in Table I.4 indicate the same trends as indicated by Table I.3: a slightly higher percentage of students in grades five and six scored at Standard Met or Standard Exceeded and the same percentage of grade seven students scored at Standard Met or Standard Exceeded. Note that Table I.4 also provides a comparison of overall results for the entire school. Because Standard Met or Standard Exceeded in mathematics is a standards-based classification, 2015–16 and 2016–17 results for the entire school may be calculated by averaging across grades. The resulting school-level averages may be compared from year to year. However, for each year, these school-level averages should be weighted to reflect the number of students in each grade. For example, the results for grade five carry more weight in the calculations for

2015–16 than the other two grades, but grade six carries more weight in the calculations for 2016–17 than the other two grades.

While these examples have made comparisons across only one year, it is important for program evaluation that results be compared across a number of years to verify that the trend is stable. The same sort of table could be used to compare year-to-year results for any test group.

Table I.4 Hypothetical Example of Using the Smarter Balanced for Mathematics to Measure Progress by Comparing Percentages of Students at Standard Met or Standard Exceeded

Grade	2015–16 Smarter Balanced for Mathematics		2016–17 Smarter Balanced for Mathematics		Difference
	No. of Students	% Standard Met or Standard Exceeded	No. of Students	% Standard Met or Standard Exceeded	
Grade 5	120	31%	111	35%	4%
Grade 6	100	33%	124	33%	0%
Grade 7	90	29%	102	31%	2%
All Grades	310	31%	337	33%	2%

Using Assessment Target Reports for Information about Test Achievement as a Whole

Achievement on a particular assessment target is provided in an Assessment Target Report in the ORS. Assessment Target Reports are produced for all claims in ELA but only for Claim 1 for Mathematics. Note, however, that the results in the Assessment Target Reports do not imply proficiency or that a particular content standard has been met. Assessment Target Reports are available for assessment targets with at least 10 unique items. Assessment target reports are one of many sources of information that should be used to evaluate student achievement

Assessment Target Reports are provided at the aggregate (group) level, for example, for a classroom, local educational agency (LEA), roster, etc., and provide information regarding a group’s strengths and weaknesses relative to the test achievement as a whole. Each group has its own unique overall total test achievement level. That is, different student groups will have different overall test achievement. Therefore, comparisons across groups might not be appropriate unless they have similar overall achievement levels.

Assessment Target Reports are not appropriate at the individual student level since individual students may receive only two or three items per assessment target. Assessment Target Reports for larger group sizes provide more reliable information in addition to contributing more unique items to the overall assessment target summary—assessment target scores based on fewer than 50 students may be less reliable and will have fewer unique items contributing to the overall assessment target summary.

For example, a group of students in grade eight might have exceeded expectations in mathematics, but their achievement on “investigating patterns of association in bivariate data”

might be relatively lower than their overall achievement; an educator might make this an area of focus for these students.

Assessment target scores are a starting point in an overall investigation of students' strengths and weaknesses and constitute only one of many sources of evidence that should be used in evaluating student performance.

Comparing Results for the California Alternate Assessments

Comparisons of CAA results should only be made within the same content area and grade; that is, compare grade four ELA in 2015–16 to grade four ELA in 2016–17 or grade eight mathematics in 2015–16 to grade eight mathematics in 2016–17. No direct comparisons should be made between grades and between subjects; for example, results for the CAA for Mathematics (Grade 7) should not be compared with results for the CAA for Mathematics (Grade 8), and results for the CAA for Mathematics (Grade 11) should not be compared with the result of the CAA for ELA (Grade 11).

Comparing Results for the Standards-based Tests in Spanish

Scale scores for the Standards-based Tests in Spanish (STS) for Reading/Language Arts (RLA) in 2016–17 may be compared to scale scores for a prior year for the **same content area, grade level, and testing program**. This allows users to say that performance for a given content area and grade was higher or lower in 2016–17 compared with 2015–16, for instance. However, scale scores for the same content area for these paper-pencil tests (PPTs) may not be compared *across* grades because scale scores are not vertically scaled, or scaled across grades. Scale scores may not be compared across tests, because the scale scores for the STS do not mean the same thing as the scale scores for another assessment.

In addition, comparing results for the STS administered to the target population to results of the STS for Non-English Learner Students in Dual-immersion Programs should be made with caution as the scale scores and performance standards were set for all STS content areas based on the STS's target population.

Comparing Performance Results

When comparing results for the STS PPTs, compare results only within the same content area and grade; that is, compare grade five RLA in 2015–16 to grade five RLA in 2016–17 or grade eight RLA in 2015–16 to grade eight science in 2016–17. No direct comparisons should be made between grades; for example, results for the STS for RLA (Grade 5) cannot be compared with results for the STS for RLA (Grade 8). In addition, comparisons should be made only within the same testing program. Results for the STS should always only be compared with other results for the STS. The matrix in Table I.5, below, shows which STS administration results may be reasonably compared with this year's results.

Table I.5 STS Years Available for Comparison to 2016–17 Results Matrix

STS Grade	STS Years Available for Comparison				
	2009	2010	2011	2012	2013–present
RLA; grades two through four	✓	✓	✓	✓	✓
RLA; grades five through seven		✓	✓	✓	✓
RLA; grades eight through eleven					✓

Two types of comparisons are possible:

1. Comparing the average scale score; or
2. Comparing the percent of students scoring at each performance level.

When making comparisons across years within a given grade and content area, it is important to understand that even when the number of students is the same, the group’s composition from year to year may be quite different if student mobility (transiency) is high.

When comparisons are made across years, they are actually a comparison of different groups of students with different traits taking different tests. Generally, there will be more variance in scores from year to year when small numbers of students are tested.

While there may be a valid comparison to be made between students within a grade and content area, it is *not* valid to subtract a student’s or class’s scale score received one year in a given content area from the scale score received the previous year in the same content area in order to show growth. While the scale scores may look the same, they are independently scaled so that differences for the same students across years cannot be calculated using basic subtraction.



Note: Any comparison of the results for the STS administered to the target population to results of the optional STS for Non–EL Students in Dual-immersion Programs (which are scored locally and not reported) should be made with caution as the scale scores and performance standards were set for all STS content areas based on the STS’s target population.

Comparing STS Scale Scores and Performance Levels for Groups



Note: A scale score is derived from a statistical process. It is *not* possible to calculate a scale score by multiplying a student’s percent correct in a content area by 600.

An example of how *group-level* scale scores for 2016–17 may be compared to the 2016–17 scale scores for the same content area and grade is shown in Table I.6, below. In this table, hypothetical average STS scale scores (SS) at three grade levels are compared between 2015–16 and 2016–17 for the students in a particular school. Compared with average scale scores in 2015–16, these data indicate higher scores in 2016–17 for grades five and ten and a virtually identical score for grade eight. In addition to comparisons for all students, similar grade-by-grade comparisons of scale scores may be made for different subgroups of interest. However, because the test scales are independent for each grade, it is not appropriate to calculate and compare average scale scores for the entire school or across grades. The same sort of table could be used to compare year-to-year results for any test group.

Table I.6 Hypothetical Example of Using the STS to Measure Progress by Comparing Average Scale Scores

Grade	2015–16 STS for RLA		2016–17 STS for RLA		Difference
	No. of Students	Mean SS	No. of Students	Mean SS	
Grade 5	120	322.2	111	333.5	11.3
Grade 8	100	331.4	124	331.7	0.3
Grade 10	90	319.9	102	323.1	3.2

Table I.7 provides a second hypothetical example of how group-level STS results may be compared. In this example, the percent of students scoring at or above proficient in science are compared between 2015–16 and 2016–17 across grades for the same school. Comparisons between 2015–16 and 2016–17 in Table I.7 indicate the same trends as indicated by Table I.6: a slightly higher percentage of students in grades five and ten scored at proficient or above and the same percentage of grade eight students scored at proficient or above. Note that Table I.7 also provides a comparison of overall results for the entire school. Because “proficient or above” in the STS is a standards-based classification, 2015–16 and 2016–17 results for the entire school may be calculated by averaging across grades. The resulting school-level averages may be compared from year to year. However, for each year, these school-level averages should be weighted to reflect the number of students in each grade. For example, the results for grade five carry more weight in the calculations for 2015–16 than the other two grades, but grade eight carries more weight in the calculations for 2016–17 than the other two grades.

While these examples have made comparisons across only one year, it is important for program evaluation that results be compared across a number of years to verify that the trend is stable. The same sort of table could be used to compare year-to-year results for any test group.

Table I.7 Hypothetical Example of Using the STS to Measure Progress by Comparing Percentages of Students at Proficient or Above

Grade	2015–16 STS for RLA		2016–17 STS for RLA		Difference
	No. of Students	% Prof or Above	No. of Students	% Prof or Above	
Grade 5	120	31%	111	35%	4%
Grade 8	100	33%	124	33%	0%
Grade 10	90	29%	102	31%	2%
All Grades	310	31%	337	33%	2%

This page is left blank intentionally.
