Smarter Balanced Assessment Consortium:

2016-17 Technical Report

- Validity
- Reliability, Precision and Errors of Measurement
- Test Fairness
- Test Design
- Scores, Scales, and Norms and
- Administration
- Reporting and Interpretation

Table of Contents

Introduction and Overview

Technical Report Approach

The intent of this report is to provide comprehensive and detailed evidence in support of the validity and reliability of Smarter Balanced assessment program. This report focuses on the summative assessment, which consists of a performance component and a computer adaptive component. Information about the overall system is included as well to provide context. At the outset, it should be recognized that demonstration of validity and reliability is an ongoing process. Validity and reliability evidence provided here from the initial pilot and the field test phases as well as evidence from more recent operational assessments.

Because the consortium is comprised of members who contract separately for test delivery and scoring and have varied practices for test administration, some evidence of validity comes from individual members, not from the Consortium. This will be noted throughout this report. In some cases (e.g., the *Online Test Administration Manual*), the consortium provides a customizable template or a guidance document, that allows for members to document their test administration practices.

To inform the Consortium, the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), hereafter referred to as the *Standards*, was used as the foundation for developing the sufficient validity and reliability evidence. Also referenced is the U.S. Department of Education (U.S. DOE) *Peer Review of State Assessment Systems Non-Regulatory Guidance for States for Meeting Requirements of the Elementary and Secondary Education Act of 1965* (2015), which stipulates the requirements for assessment programs to receive federal approval under current ESEA legislation. With respect to Smarter Balanced, this information is necessary for understanding the degree to which the Consortium is meeting its goals, and in some cases, what further tasks remain to improve the system as it evolves operationally.

Peer Review Guidelines and Established Standards

Among the principles underlying the Smarter Balanced theory of action is adherence "to established professional standards" (Smarter Balanced, 2010, p. 33). In addition to adhering to the AERA et al. (2014) *Standards*, the Consortium will also meet selected requirements of the U.S. DOE peer review process for ESEA assessments. There is a great deal of overlap between the AERA et al. (2014) *Standards* and the U.S. DOE *Peer Review Guidance*. However, the *Guidance* stipulates many important requirements. In particular, to meet these requirements the validity and reliability evidence and the ongoing research agenda should include

- evidence concerning the purpose of an assessment system and studies that support the validity of using results from the assessment system based on their stated purpose and use,

- strong correlations of test and item scores, with relevant measures of academic achievement and weak correlations with irrelevant characteristics, such as demographics (i.e., convergent and discriminant validity),
- documentation of the definitions for cut scores and the rationale and procedures for establishing them,
- evidence concerning the precision of the cut scores and consistency of student classification,
- evidence of sufficient levels of reliability for the overall population and for each targeted subpopulation,
- evidence of content alignment over time through quality control reviews,
- evidence of comprehensive alignment and measurement of the full range of content standards, depth of knowledge, and cognitive complexity,
- evidence that the assessment plan and test specifications describe how all content standards are assessed and how the domain is sampled in a fashion that supports valid inferences about student performance on the standards, both individually and aggregated,
- scores that reflect the full range of achievement standards,
- documentation that describes a coherent system of assessment across grades and subjects including studies establishing vertical scales, and
- evidence of how assessments provide information on the progress of students.

These characteristics of high-quality assessment systems were given consideration in the development of the Smarter Balanced Assessment System to provide evidence that assessments meet these high standards. The theory of action and primary purposes and goals of Smarter Balanced are briefly described below.

Overview and Background of the Smarter Balanced Theory of Action

The Smarter Balanced Assessment Consortium supports the development and implementation of learning and assessment systems to reshape education in member states in order to improve student outcomes. Through expanded use of technology and targeted professional development, the consortium's theory of action calls for the integration of learning and assessment systems, leading to more informed decision-making and higher-quality instruction and ultimately increasing the number of students who are well prepared for college and careers.

The ultimate goal of Smarter Balanced is to ensure that all students leave high school are prepared for postsecondary success in college or a career through improved teaching and increased student learning. This approach suggests that enhanced learning will result from high-quality assessments that support ongoing improvements in instruction. A quality assessment system strategically "balances" summative, interim, and formative components (Darling-Hammond & Pecheone, 2010). An assessment system must provide valid measurement across the full range of performance on common academic content, including assessment of deep disciplinary understanding and higher-order thinking skills increasingly demanded by a knowledge-based economy.

Six Principles of Smarter Balanced Underlying the Theory of Action

The Smarter Balanced assessment is guided by a set of six principles shared by systems in high-achieving nations and some high-achieving states in the U.S.

1. Assessments are grounded in a thoughtful, standards-based curriculum and managed as part of an integrated system of standards, curriculum, assessment, instruction, and teacher development. Curriculum and assessments are organized around a well-defined set of learning progressions along multiple dimensions within subject areas. Formative assessment processes and tools and interim assessments are conceptualized in tandem with summative assessments; all of them are linked to the CCSS and supported by a unified technology platform.
2. Assessments produce evidence of student performance on challenging tasks that represent the CCSS. Instruction and assessments seek to teach and evaluate knowledge and skills that generalize and can transfer to higher education and multiple work domains. These assessments emphasize deep knowledge of core concepts and ideas within and across the disciplines—along with analysis, synthesis, problem solving, communication, and critical thinking—thereby requiring a focus on complex performances as well as on specific concepts, facts, and skills.
3. Teachers are integrally involved in the development and scoring of assessments. While many assessment components are efficiently scored with computer assistance, teachers must also be involved in the formative and summative assessment systems so that they understand and can teach in a manner that is consistent with the full intent of the standards while becoming more skilled in their own classroom assessment practices.
4. The development and implementation of the assessment system is a state-led effort with a transparent and inclusive governance structure. Assessments are structured to improve teaching and learning. Assessments as, of, and for learning are designed to develop understanding of learning standards, what constitutes high-quality work, to what degree is growth occurring, and what is needed for further student learning.
5. Assessment, reporting, and accountability systems provide useful information on multiple measures that is educative for all stakeholders. Reporting of assessment results is timely and meaningful—offering specific information about areas of performance so that teachers can follow up with targeted instruction, students can better target their own efforts, and administrators and policymakers can fully understand what students know and can do—in order to guide curriculum and professional development decisions.
6. Design and implementation strategies adhere to established professional standards. The development of an integrated, balanced assessment system is an enormous undertaking, requiring commitment to established quality standards in order for the system to be credible, fair, and technically sound. Smarter Balanced continues to be committed to developing an assessment system that meets critical elements required by U.S. DOE Peer Review, relying heavily on the Standards as its core resource for quality design.

The primary rationale of the Smarter Balanced assessments is that these six principles can interact to improve the intended student outcomes (i.e., college- and career-readiness).

Purposes for the Smarter Balanced Assessment System

The Smarter Balanced purpose statements are organized into three categories: (a) summative assessments, (b) interim assessments, and (c) formative assessment resources. This report provides technical information about the summative assessments. The purposes of interim assessments and formative resources are also stated in this section to provide context for summative assessments as a component of the assessment system.

The purposes of the Smarter Balanced summative assessments are to provide valid, reliable, and fair information about

- students' ELA/literacy and mathematics achievement with respect to the CCSS measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school;
- whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college readiness;
- whether grade 11 students have sufficient academic proficiency in ELA/literacy and mathematics to be ready to take credit-bearing, transferable college courses after completing their high school coursework;
- students' annual progress toward college- and career-readiness in ELA/literacy and mathematics;
- how instruction can be improved at the classroom, school, district, and state levels;
- students' ELA/literacy and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems; and
- students' achievement in ELA/literacy and mathematics that is equitable for all students and subgroups of students.

The purposes of the Smarter Balanced interim assessments are to provide valid, reliable, and fair information about

- student progress toward mastery of the skills in ELA/literacy and mathematics measured by the summative assessment;
- student performance at the claim or cluster of assessment targets so teachers and administrators can track student progress throughout the year and adjust instruction accordingly;
- individual and group (e.g., school, district) performance at the claim level in ELA/literacy and mathematics to determine whether teaching and learning are on target;
- teacher-moderated scoring of performance events as a professional development vehicle to enhance teacher capacity to evaluate student work aligned to the standards; and
- student progress toward the mastery of skills measured in ELA/literacy and mathematics across all students and subgroups.

The purposes of the Smarter Balanced formative assessment resources are to provide tools and resources to

- improve teaching and learning;
- to help teachers monitor their students' progress throughout the school year,
- illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning;
- help teachers and other educators align instruction, curricula, and assessments;
- assist teachers and other educators in using the summative and interim assessments to improve instruction at the individual and classroom levels; and
- offer professional development and resources for how to use assessment information to improve teacher decision-making in the classroom.

Overview of Report Chapters:

Chapters in the Technical Report follow elements in the 2014 Standards:

CH#	Chapter title
1	Validity
2	Reliability/Precision and Errors of Measurement
3	Test Fairness
4	Test Design
5	Scores, Scales, Norms
6	Test Administration
7	Reporting and Interpretation

Brief synopses of these chapters are given below in order to direct further review. At the suggestion of our members, we have written practical descriptions of the purpose of evidence in each chapter to provide context for teachers, parents and other stakeholders.

Chapter 1: Validity

In a sense, all of the information in this Technical Report provides validity evidence. This chapter provides information about test purposes and the overall approach to showing how scores are appropriate for those purposes. The information in this chapter answers questions such as:

- For what purposes was the summative assessment designed to be used?
- What evidence shows that test scores are appropriate for these uses?
- What are the intended test score interpretations for specific uses?

Chapter 2: Reliability/Precision and Errors of Measurement

The degree of accuracy and precision of scores contributes to evidence about appropriate test score interpretation and use. Decisions must be made with full knowledge of measurement error and reliability. Chapter 2 presents information about how the test performs in terms of measurement precision, reliability, classification consistency, and other technical criteria. The information is based on simulation studies and operational test data from the item pool and school year identified in the title of this report. Information presented in this chapter can answer questions such as:

- How do we know that scores are accurate?
- How do we know they are reliable and equally precise for all students?

Chapter 3: Test Fairness

Test fairness concerns whether score interpretations are valid for all relevant subgroups that minimizes construct irrelevant variance. The evidence for test fairness can be logical (e.g., bias review of items) or statistical in nature (e.g., differential item functioning) and includes availability of resources that increase participation and improve assessment of skills. Chapter 3 presents the Smarter Balanced Conceptual Framework for Usability, Accessibility, and Accommodations, bias and sensitivity reviews conducted during item and task development. Information is presented about the development and use of test accommodations and accessibility features. Statistical information pertaining to differences in item functioning (DIF) across demographic groups is presented. Information presented in this chapter can answer questions such as:

- How were test questions and tasks developed to ensure fairness to all students?
- How is the test administered so that each student can demonstrate their skills?
- How do we know that the test is fair to all students?

Chapter 4: Test Design

Test design is predominantly focused on the content validity of the test. Tasks and items must represent the domain of knowledge and skill as intended. For Smarter Balanced assessments, test design includes the relationship of claims and targets to the underlying CCSS and how adaptive and performance task components work together. The full test design also encompasses the delivery algorithm and the method of scoring the test. This chapter includes a description of item pool and task development supporting test design. Chapter 4 provides evidence that the items students receive are appropriate in terms of both content and difficulty. It also describes test structure (claims, targets) and its relationship to the CCSS, item and task development and alignment studies.

Chapter 4 also has information about the operational blueprints, adaptive algorithm, test scoring method and application and pool analysis. Information in Chapter 4 can answer questions such as:

- What's on the test? Is it consistent with stated test purposes?
- Does each student get a set of questions that fully represents the content domain?
- How does each student gets a test with an appropriate level of difficulty?

Chapter: 5 Scores, Scales and Norms

Chapter 5 summarizes how scales were established in pilot and field test stages. It describes how cut scores were developed from foundational achievement levels that delineated progress toward career and college readiness. It provides logit-to-scale transformations. Normative information, including means, percentiles and achievement level distribution is displayed. Information in Chapter 5 can answer questions such as:

- What do the test scores mean?
- How were they developed?
- What do achievement levels mean?
- How well did students perform this year compared to previous years?
- How did students in one demographic group perform compared to others.
- How do students in one grade perform compared to other grades?

Chapter 6: Test Administration

Part of test validity rests on the assumption that assessments are administered in a standard manner. Because Smarter Balanced tests are given on such a large scale, in different policy and operational contexts, the Consortium provides a common administration template that members customize for specific use. Chapter 6 describes the customizable Smarter Balanced Online Test Administration Manual. It presents operational item exposure rates and blueprint fidelity. Embedded field test results, including item scoring processes and inter-rater reliability of field tested items are shown. The information in Chapter 6 can answer questions such as:

- What are conditions for test administration to assure that every student was afforded the same chance for success?
- How was the test administered to allow for accessibility for all students?
- Was the test administration secure?
- Do test records show that the test was administered as intended?
- Were field tested items successful?

Chapter 7: Reporting and Interpretation

Reports based on test scores are among the most public-facing features of an assessment program. They must be useful as well as accurate – supporting the decisions and purposes for which the assessment was designed, while discouraging inappropriate conclusions and comparisons. Chapter 7 provides examples of the Smarter Balanced suite of reports and interpretive information, and discusses intended uses of report information. Information in Chapter 7 can answer questions such as:

- What information do Smarter Balance reports contain?
- What do scores mean?
- How can the reports best be used by teachers and parents?

Acknowledgments

Outside Groups and Organizations that Collaborated with the Smarter Balanced Assessment Consortium. Below is a partial list of individuals and groups that contributed time and expertise to the work of the Consortium.

2016-17 Technical Advisory Committee.

The Technical Advisory Committee (TAC) provides guidance on technical assessment matters pertaining to validity and reliability, accuracy, and fairness. Members of the TAC are highly regarded national experts who have been widely published in their fields. Areas of expertise include: assessment design; computer adaptive testing (CAT); assessment accommodations; uses of tests; mathematics, and English language arts/literacy. Following is a list of committee members and their affiliations.

- Randy Bennett, Ph.D. — *ETS*
- Derek C. Briggs, Ph.D. — *University of Colorado*
- Gregory J. Cizek, Ph.D. — University of North Carolina
- Shelbi Cole, Ph.D. — Student Achievement Partners
- David T. Conley, Ph.D. — *University of Oregon*
- Brian Gong, Ph.D. — The Center for Assessment
- Edward Haertel, Ph.D. — *Stanford University*
- *Gerunda Hughes, Ph.D.* — *Howard University*
- G. Gage Kingsbury, Ph.D. — *Psychometric Consultant*
- Joseph Martineau, Ph.D. — The Center for Assessment
- James W. Pellegrino, Ph.D. — University of Illinois, Chicago
- W. James Popham, Ph.D. — *UCLA, Emeritus*
- Joseph Ryan, Ph.D. — Arizona State University
- Guillermo Solano-Flores, Ph.D. — Stanford University
- Martha Thurlow, Ph.D. — University of Minnesota/NCEO
- Sheila Valencia, Ph.D. — University of Washington
- Joe Willhoft, Ph.D. — Consultant

English Language Learners Advisory Committee

The English Language Learners Advisory Committee is comprised of national experts in ELL assessment, bilingual education, and language acquisition. This committee will provide feedback to Smarter Balanced staff, work groups, and contractors to ensure that the assessments provide valid, reliable, and fair measures of achievement and growth for English learners. Following is a list of committee members.

- Stephanie Cawthon, Ph.D.
- Magda Chia, Ph.D.
- Gary Cook, Ph.D.
- Kathy Escamilla, Ph.D.
- James Green, Ph.D.
- Kenji Hakuta, Ph.D.
- Robert Linquanti
- Guillermo Solano-Flores, Ph.D.
- Guadalupe Valdés, Ph.D.

Students with Disabilities Advisory Committee

The Students with Disabilities Advisory Committee is comprised of national experts in learning disabilities, assistive technology, and accessibility and accommodations policy. This committee will provide feedback to Smarter Balanced staff, work groups, and contractors to ensure that the assessments provide valid, reliable, and fair measures of achievement and growth for students with disabilities. Following is a list of committee members.

- Donald D. Deshler, Ph.D.
- Barbara Ehren, Ed.D.
- Cheryl Kamei-Hannan, Ph.D.
- Jacqueline F. Kearns, Ed.D.
- Susan Rose, Ph.D.
- Jim Sandstrum
- Ann C. Schulte, Ph.D.
- Richard Simpson, Ed.D.
- Stephen W. Smith, Ph.D.
- Martha L. Thurlow, Ph.D.

Performance and Practice Committee

The Performance and Practice Committee is comprised of nearly 20 educators from around the nation who were nominated by state chiefs. This committee will assess the efficiency of Smarter Balanced assessments to meet their designed purpose and to deepen overall stakeholder investment. Following is a list of committee members and their member affiliation.

- Kandi Greaves (Vermont)
- Mary Jo Faust (Delaware)
- Shannon Mashinchi (Oregon)
- Susan Green (California)
- Steve Seal (California)
- Tanya Golden (California)
- Crista Anderson (Montana)
- Melissa Speetjens (Hawaii)
- Tiffany Seibel (Nevada)
- Mike Nelson (Idaho)
- Guyla Ness (South Dakota)
- Abby Olinger Quint (Connecticut)
- Michelle Center (California)
- Todd Bloomquist (Oregon)
- Jim O'Neill (Montana)
- Jen Paul (Michigan)
- Eva Payne (Oregon)
- Toni Wheeler (Washington)
- Joe Willhoft (Technical Advisory Committee)

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Darling-Hammond, L., & Pecheone, R. (2010). *Developing an Internationally Comparable Balanced Assessment System that Supports High-Quality Learning*. Retrieved from https://www.ets.org/Media/Research/pdf/Darling-HammondPechoneSystemModel.pdf .

Smarter Balanced Assessment Consortium (2010, June 23). *Race to the top assessment program application for new grants: Comprehensive assessment systems*. CFDA Number: 84.395B. OMB Control Number 1810-0699. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

U.S. Department of Education (2015, September). *Peer review of state assessment systems non-regulatory guidance for states for meeting requirements of the Elementary and Secondary Education Act of 1965*. Washington, DC: Author.

Sireci, S. G. (2012). Smarter Balanced Assessment Consortium: Comprehensive research agenda. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

Chapter 1: Validity

Introduction

Validity refers to the degree to which each interpretation or use of a test score is supported by the accumulated evidence (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; ETS, 2002). It constitutes the central notion underlying the development, administration, and scoring of a test and the uses and interpretations of test scores. Validation is the process of accumulating evidence to support each proposed score interpretation or use. The validation process does not rely on a single study or only one type of evidence. Rather, validation involves multiple investigations and different kinds of supporting evidence (AERA, APA, & NCME, 2014; Cronbach, 1971; ETS, 2002; Kane, 2006). It begins with test design and is implicit throughout the assessment process, which includes developing, field-testing and analyzing items; test scaling and linking; scoring; and reporting.

This technical report summarizes the test design, construction, implementation, scoring, reporting, and psychometric processes for the 2016-17 operational administration of the Smarter Balanced assessment. As such, it is an important source of evidence for the validity argument. This chapter provides a framework for the validation of the Smarter Balanced summative assessment (Sireci, 2013). Following this introductory section, we examine the validity argument, including intended purposes for the summative assessment, types of evidence collected, a high-level summary. The main portion of the section on the validity argument consists of an evidentiary framework supporting the validity argument and pointing the reader to supporting evidence in other parts of the technical report and in other studies for each of the intended uses. Evidence is organized around the principles in the AERA, APA, and NCME's *Standards for Educational and Psychological Testing* (2014), hereafter referred to as *the Standards*.

A Note on the Validity Evidence Presented in Technical Report

Validity is an ongoing process with continuous addition of evidence from a variety of contributors. This report summarizes development and performance of the instrument itself, addressing test content, response processes and internal structure. Other elements come from supplemental research projects or third-party studies.

As the *Standards* note, "validation is the joint responsibility of the test developer and the test user." (AERA, et. al. 2014, p, 13). The Consortium does not control aspects of test administration and use. Consortium members deliver the test, score operational items, provide reports, and incorporate test scores into their unique accountability models. This report documents guidelines for administration and use. For complete validity evidence, member documentation on specific test administration procedures, reporting, and use should be consulted.

This report also does not provide evidence related to the consequences of testing. Ultimate use of test scores is determined by consortium members. Each member decides the purpose and interpretation of scores and each has crafted its own system of reporting and accountability. The Consortium provides information about test content and technical quality, but does not interfere in member use of scores. The consortium does not endorse or critique member uses.

While it is beyond the scope and purpose of a technical report to evaluate evidence pertaining to consequences of testing, we believe that the breadth and depth of the supporting evidence demonstrates that the Smarter Balanced Assessment System adheres to guidelines for fair and high-quality assessment. The Smarter Balanced summative assessments have been thoroughly evaluated through the United States Department of Education's Peer Review process.

The Validity Argument

This section presents the intended purposes of the Smarter Balanced assessments, a brief discussion of the types of validity evidence collected to support those purposes, and a high-level overview of the validity argument. At the end of this section, we present an evidentiary framework where each intended purpose is listed along with available validity evidence.

Intended Purposes of the Smarter Balanced System for Summative Assessments

The validity argument begins with a statement of the intended purposes for the summative assessments. The purposes of the Smarter Balanced summative assessments are to provide valid, reliable, and fair information about:

1. Students' ELA/literacy and mathematics achievement with respect to those Common Core State Standards (CCSS) measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school.
2. Whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college readiness.
3. Whether grade 11 students have sufficient academic proficiency in ELA/literacy and mathematics to be ready to take credit-bearing, transferable college courses after completing their high school coursework.
4. Students' annual progress toward college and career readiness in ELA/literacy and mathematics.
5. How instruction can be improved at the classroom, school, district, and state levels.
6. Students' ELA/literacy and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems.
7. Students' achievement in ELA/literacy and mathematics that is equitable for all students and subgroups of students.

Types of Validity Evidence

The intended purposes must be supported by evidence. The *Standards* describe a process of validation, often characterized as a validity argument (Kane, 1992; Kane, 1996), that consists of developing a sufficiently convincing, empirically-based argument that the interpretations and actions based on test scores are sound.

> A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended

> interpretation of test scores for specific uses. Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system (AERA et al., 2014, p. 21-22).

The sources of validity evidence described in the *Standards* (AERA et al. 2014, pp. 26-31) include:

1. Evidence Based on Test Content
2. Evidence Based on Response Processes
3. Evidence Based on Internal Structure
4. Evidence Based on Relations to Other Variables
5. Evidence Based on Consequences of Testing[1].

Evidence related to test content, internal structure, response processes, and relations to other variables are presented throughout this technical report. The *Standards* also include evidence related to test consequences, which is beyond the scope of this technical report. These sources of validity evidence are intended to emphasize different aspects of validity; however, since validity is a unitary concept, they do not constitute distinct types of validity. We briefly describe each type of validity evidence (excluding test consequences) before examining the types of evidence available for each intended purpose.

Evidence Based on Test Content. Validity evidence based on *test content* refers to traditional forms of content validity evidence, such as the rating of test specifications and test items (Crocker, Miller, & Franks, 1989; Sireci, 1998), as well as "alignment" methods for educational tests that evaluate the interactions between curriculum frameworks, testing, and instruction (Rothman, Slattery, Vranek, & Resnick, 2002; Bhola, Impara & Buckendahl, 2003; Martone & Sireci, 2009). Administration and scoring can be considered aspects of content-based evidence. In the case of computer adaptive test administration, confirmation that each test "event" administered to students conforms to the test blueprint can provide content-based evidence.

Evidence Based on Response Process. Validity evidence based on *response process* refers to "evidence concerning the fit between the construct and the detailed nature of performance or responding actually engaged in by examinees" (AERA et al., 1999 p. 12). This type of evidence confirms that an assessment measures the intended cognitive skills, and that students are using these targeted skills to respond to the items.

Evidence Based on Internal Structure. Validity evidence based on *internal structure* refers to statistical analyses of item and score subdomains to investigate the primary and secondary (if any) dimensions measured by an assessment. Procedures for gathering such evidence include factor

[1] This report does not provide evidence related to the consequences of testing. Ultimate use of test scores is determined by consortium members. Each member decides the purpose and interpretation of scores and each has crafted its own system of reporting and accountability. The Consortium provides information about test content and technical quality but does not interfere in member use of scores. The consortium does not endorse or critique member uses.

analysis or multidimensional IRT scaling (both exploratory and confirmatory). For a test with a vertical scale, a consistent primary dimension or construct shift across the levels of the test should be maintained. Internal structure evidence also evaluates the "strength" or "salience" of the major dimensions underlying an assessment using indices of measurement precision such as test reliability, decision accuracy and consistency, generalizability coefficients, conditional and unconditional standard errors of measurement, and test information functions.

Evidence Based on Relations to Other Variables. Evidence based on *relations to other variables* refers to traditional forms of criterion-related validity evidence such as concurrent and predictive validity, as well as more comprehensive investigations of the relationships among test scores and other variables such as multitrait-multimethod studies (Campbell & Fiske, 1959). These external variables can be used to evaluate hypothesized relationships between test scores and other measures of student achievement (e.g., test scores and teacher-assigned grades); the degree to which different tests actually measure different skills; and, the utility of test scores for predicting specific criteria (e.g., college grades).

Overview of the Validity Argument

The crux of the validity argument presented here is that the technical quality of the summative assessments supports the intended purposes of the assessments. The CCSS, which have been adopted by Smarter Balanced members, are widely recognized content standards for college and career readiness in high school grades, and for being on track for college and career readiness in lower grades (Conley, et al. 2011). Content specifications and test blueprints show that the Smarter Balanced summative assessments essentially cover the breadth and depth of assessable standards. Content experts developed expanded item types that allow response processes that reveal skills and knowledge at various specified levels of depth. Most of each content area test is delivered adaptively so that blueprint requirements are met, scores are more accurate and reliable, and student experience is enhanced. Summative test scores are suitable for use in a variety of member accountability systems. Claim-level sub-score reports indicate directions for gaining further instructional information through the interim system or classroom observation.

The consortium chose its psychometric model after investigating a variety of models and establishing a clear structural relationship across grades. A vertical scale was constructed to provide measurement across grades, facilitating estimates of progress toward college readiness. The appropriateness of Smarter Balanced performance standards as predictors of college and career readiness in grade 11 and of being on-track for readiness in grades three through eight was established by an extended achievement-level-setting process. The process began with authoring achievement level policy definitions and continued through a rigorous process of setting achievement criteria. These processes involved participants from the post-secondary systems of member jurisdictions to ensure that readiness criteria represented skills needed for success in first-year credit-bearing college courses.

Evidentiary Framework

Sireci (2012) proposed a comprehensive validity framework for Smarter Balanced assessments in which the purposes of the Smarter Balanced assessments were cross-classified with the five sources

of validity evidence from the standards. Table 1-1 presents a similar cross-classification, but with the source "Consequences of Testing" omitted for reasons given above. For most cells in his table, Sireci described the kinds of validity studies that could be performed. Not all of the hypothetical research studies described in the Sireci paper have been performed and are publically available. The checks in Table 1-1 show the combinations of purpose and evidentiary source for which evidence is cited in this chapter. The supporting evidence is presented in tables consisting of two columns. One column points the reader to a one or more chapters in this technical report and describes the evidence these chapters contain. The other column lists studies and documents that are external to this report and which may be found elsewhere. As additional validity evidence becomes available, it will be similarly cited in future technical reports.

TABLE 1-1 ASSESSMENT PURPOSES CROSS-CLASSIFIED BY SOURCES OF VALIDITY EVIDENCE

Purpose	Source of Validity Evidence for Summative Assessments			
	Test Content	Internal Structure	Response Processes	Relations to
1. Report achievement with respect to the CCSS as measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school.	✓	✓	✓	✓
2. Assess whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college readiness.	✓	✓	✓	✓
3. Assess whether grade 11 students have sufficient academic proficiency in ELA/literacy and mathematics to be ready to take credit-bearing, transferable college courses after completing their high school coursework.	✓	✓	✓	✓
4. Measure students' annual progress toward college and career readiness in ELA/literacy and mathematics.	✓	✓	✓	✓
5. Inform how instruction can be improved at the classroom, school, district, and state levels.	✓	✓	✓	
6. Report students' ELA/literacy and mathematics proficiency for federal accountability purposes and potentially for state and local accountability systems.	✓	✓	✓	
7. Assess students' achievement in ELA/literacy and mathematics in a manner that is equitable for all students and subgroups of students.	✓	✓	✓	

Intended Purpose 1:

Report achievement with respect to the CCSS as measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school.

Intended Purpose 1 is supported by validity evidence related to content, internal structure, response processes, and relationship to other variables. Table 1-2 summarizes the sources of validity internal to this report and lists pieces of evidence found outside of this report for Intended Purpose 1.

TABLE 1-2. VALIDITY EVIDENCE THAT SUPPORTS INTENDED PURPOSE 1

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Test Content	
See Chapters 3, 4 • Bias is minimized through Universal Design and accessibility resources. • Test blueprint, content specifications, and item specifications are aligned to the full breadth and depth of grade level content, process skills, and associated cognitive complexity. • With very few exceptions, operational computer adaptive test events meet all blueprint constraints, both for the general student population and for students taking accommodated test forms. • Item pools for both the general assessment and accommodated assessments are sufficiently robust to support the computer adaptive delivery of tests that measure the full breadth and depth of the CCSS according to test blueprint requirements.	• *Evaluation of the Alignment Between the Common Core State Standards and the Smarter Balanced Assessment Consortium Summative Assessments for Grades 3, 6, and 7 in English Language Arts/Literacy and Mathematics – Final Report* (WestEd Standards, Assessment, and Accountability Services Program, November 2017) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016) • Smarter Balanced Blueprint Fidelity Studies and Item Pool Gap Analyses • *Evaluating the Content and Quality of Next Generation High School Assessments* (HumRRO, February 2016) • *Evaluating the Content and Quality of Next Generation Assessments* (Fordham, February 2016)
Evidence Related to Internal Structure	
See Chapter 2, 5 • Assessment blueprints and content specifications consistent with structure and content of CCSS. • The assessment supports precise measurement and consistent classification. • Achievement levels were set consistent with best practice. • Confirmatory Factor Analysis shows good model fit.	• *2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 6, 9, 10)* • *Development Process (NGA Center & CCSSO, 2016)* • *Dimensionality of the SBAC: An Argument for its validity* (CAASPP-CAHSEE Technical Advisory Group, October 2015) • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015)

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Response Processes	
See Chapters 3, 4, 5 • Bias is minimized through Universal Design and accessibility resources. • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. • Achievement levels were set consistent with best practice. • Cognitive Labs describe students' engagement with tasks and items and provide confirmation of content measurement.	• *Cognitive Laboratories Technical Report (AIR, 2013)* • *Development Process (NGA Center & CCSSO, 2016)* • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016) • *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines* (November 2014)
Evidence Related to Other Variables	
See Chapter 5 • Achievement levels are consistent with other measures	• *External Validity: Analysis of Existing External Measures* (UCLA/CRESST, May 2016) • *Linking Course Grades to Smarter Balanced Cut Scores* (OSPI, 2016) • *Linking Study Between Smarter Balanced Mathematics Field Test and CSU Entry Level Math Test* (ETS, 2015) • *Linking the Smarter Balanced English Language Arts/Literacy Summative Assessment with The Lexile Framework for Reading* (MetaMetrics, September 2016) • *Linking the Smarter Balanced Mathematics Summative Assessment with The Quantile Framework for Mathematics* (September 2016) • *Study of the Relationship Between the Early Assessment Program and the Smarter Balanced Field Tests* (ETS, 2015)

Intended Purposes 2 and 3:

Purpose 2: Whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college readiness.

Purpose 3: Whether grade 11 students have sufficient academic proficiency in ELA/literacy and mathematics to be ready to take credit-bearing, transferable college courses after completing their high school coursework.

Intended Purposes 2 and 3 are supported by validity evidence related to content, internal structure, response processes, and relationship to other variables. Table 1-3 summarizes the sources of validity internal to this report and lists pieces of evidence found outside of this report for Intended Purposes 2 and 3.

TABLE 1-3. VALIDITY EVIDENCE THAT SUPPORTS INTENDED PURPOSES 2 AND 3

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Test Content	
See Chapter 4 • CCSS are based on skills leading to CCR across grades. • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. • Achievement level descriptor writing and standard setting included broad stakeholder input and was based on skills outlined in the CCSS. • Computer adaptive test events meet blueprint constraints.	• *Development Process (NGA Center & CCSSO, 2016)* • *Evaluating the Content and Quality of Next Generation Assessments* (Doorey & Polikoff, 2016) • *Evaluating the Content and Quality of Next Generation High School Assessments* (Schultz, Michaels, Dvorak, & Wiley, 2016) • *Evaluation of the Alignment Between the Common Core State Standards and the Smarter Balanced Assessment Consortium Summative Assessments for Grades 3, 6, and 7 in English Language Arts/Literacy and Mathematics – Final Report* (WestEd Standards, Assessment, and Accountability Services Program, November 2017) • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016) • *Smarter Balanced Blueprint Fidelity Studies*
Evidence Related to Internal Structure	
See Chapter 5 • Scale is vertically articulated • Achievement levels are vertically articulated • Confirmatory Factor Analysis shows good model fit.	• *2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 6, 9, 10)* • *Dimensionality of the SBAC: An Argument for its validity* (CAASPP-CAHSEE Technical Advisory Group, October 2015) • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015)

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Response Processes	
See Chapters 4, 5 • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. • Achievement levels are vertically articulated • Cognitive Labs describe students' engagement with tasks and items and provide confirmation of content measurement.	• *Cognitive Laboratories Technical Report (AIR, 2013)* • *Development Process (NGA Center & CCSSO, 2016)* • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016)
Evidence Related to Other Variables	
See Chapter 5 • Achievement levels are consistent with other measures.	• Hawaii Smarter Balanced Technical Report, 2014-2015, pp. 48-50 • *Linking the Smarter Balanced English Language Arts/Literacy Summative Assessment with The Lexile Framework for Reading* (MetaMetrics, September 2016) • *Linking the Smarter Balanced Mathematics Summative Assessment with The Quantile Framework for Mathematics* (September 2016) • South Dakota Technical Report, 2014-2015, pp. 53-55 • *Study of the Relationship Between the Early Assessment Program and the Smarter Balanced Field Tests* (ETS, 2015)

Intended Purpose 4:

Measure students' annual progress toward college and career readiness in ELA/literacy and mathematics.

Intended Purpose 4 is supported by validity evidence related to content, internal structure, response processes, and relationship to other variables. Table 1-4 summarizes the sources of validity internal to this report and lists pieces of evidence found outside of this report for Intended Purpose 4.

TABLE 1-4. VALIDITY EVIDENCE THAT SUPPORTS INTENDED PURPOSE 4

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Test Content	
See Chapter 4 • CCSS are based on CCR and skills leading to CCR across grades. • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity.	• *Development Process (NGA Center & CCSSO, 2016)* • *Evaluating the Content and Quality of Next Generation Assessments* (Doorey & Polikoff, 2016) • *Evaluating the Content and Quality of Next Generation High School Assessments* (Schultz, Michaels, Dvorak, & Wiley, 2016) • *Evaluation of the Alignment Between the Common Core State Standards and the Smarter Balanced Assessment Consortium Summative Assessments for Grades 3, 6, and 7 in English Language Arts/Literacy and Mathematics – Final Report* (WestEd Standards, Assessment, and Accountability Services Program, November 2017) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016)
Evidence Related to Internal Structure	
See Chapters 2, 5 • The assessment supports precise measurement and consistent classification to support analysis and reporting of longitudinal data. • Scale is vertically articulated. • Achievement levels are vertically articulated.	• *2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 6, 9, 10)* • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015)
Evidence Related to Response Processes	
See Chapters 4, 5 • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. • Achievement levels are vertically articulated. • Cognitive Labs describe students' engagement with tasks and items and provide confirmation of content measurement.	• *Cognitive Laboratories Technical Report (AIR, 2013)* • *Development Process (NGA Center & CCSSO, 2016)* • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016)

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Other Variables	
See Chapter 5 • Will be addressed in future studies of annual observed growth.	• *Disaggregating Longitudinal Achievement Level Data with Student and Group Characteristics (Study Design)* (Cai, 2016) • *External Validity of Smarter Balanced assessments on placement and performance in entry-level, credit-bearing courses in colleges and universities (Study Design)*

Intended Purpose 5:

Inform how instruction can be improved at the classroom, school, district, and state levels.

Intended Purpose 5 is supported by validity evidence related to content, internal structure, and response processes. Table 1-5 summarizes the sources of validity internal to this report and lists pieces of evidence found outside of this report for Intended Purpose 5.

TABLE 1-5 VALIDITY EVIDENCE THAT SUPPORTS INTENDED PURPOSE 5

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Test Content	
See Chapters 4, 7 • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. • The blueprint was developed in consultation with educators. • Assessment Claims align with the structure of the CCSS to support the interpretation of the assessment results. • Assessments have been successfully linked to well-known scales such as Lexiles and Quantiles to inform classroom instruction, curriculum, and instructional materials at the individual student and aggregate levels.	• *End of Grant Report* (Smarter Balanced, 2015, p. 28) • *Evaluating the Content and Quality of Next Generation Assessments* (Doorey & Polikoff, 2016) • *Evaluating the Content and Quality of Next Generation High School Assessments* (Schultz, Michaels, Dvorak, & Wiley, 2016) • *Evaluation of the Alignment Between the Common Core State Standards and the Smarter Balanced Assessment Consortium Summative Assessments for Grades 3, 6, and 7 in English Language Arts/Literacy and Mathematics – Final Report* (WestEd Standards, Assessment, and Accountability Services Program, November 2017) • *Linking the Smarter Balanced English Language Arts/Literacy Summative Assessment with The Lexile Framework for Reading* (MetaMetrics, September 2016) • *Linking the Smarter Balanced Mathematics Summative Assessment with The Quantile Framework for Mathematics* (September 2016) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016)
Evidence Related to Internal Structure	
See Chapters 4, 5, 7 • Threshold, Range and Policy Achievement Levels were developed in consultation with educators, with the goal of providing information to educators. • Assessment Claims align with the structure of the CCSS to support the interpretation of the assessment results.	• *Development Process (NGA Center & CCSSO, 2016)* • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Technical Report Initial Achievement Level Descriptors* (April 2013)
Evidence Related to Response Processes	
See Chapters 4, 5 • Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. • Threshold, Range and Policy Achievement Levels were developed in consultation with educators, with the goal of providing information to educators.	• *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Alignment Study Report* (HumRRO, April 2016) • *Smarter Balanced Assessment Consortium: Technical Report Initial Achievement Level Descriptors* (April 2013)

Intended Purpose 6:

Report students' ELA/literacy and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems.

Intended Purpose 6 is supported by validity evidence related to content, internal structure, and response processes. Table 1-6 summarizes the sources of validity internal to this report and lists pieces of evidence found outside of this report for Intended Purpose 6.

TABLE 1-6. VALIDITY EVIDENCE THAT SUPPORTS INTENDED PURPOSE 6

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Test Content	
See Chapters 5, 6, 7 • Achievement levels were set for the explicit purpose of reporting student achievement as part of federal accountability. • Assessments are administered in a standardized manner sufficient to yield data that supports valid inferences.	• *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015) • *Smarter Balanced Assessment Consortium: Online, Summative, Test Administration Manual* • *Smarter Balanced Assessment Consortium: State Procedures Manual*
Evidence Related to Internal Structure	
See Chapters 2, 5, 7 • Achievement levels were set for the explicit purpose of reporting student achievement as part of federal accountability. • The assessment supports precise measurement and consistent classification to support analysis as part of state and local accountability systems.	• 2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 10) • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015)
Evidence Related to Response Processes	
See Chapters 5, 7 • Achievement levels were set for the explicit purpose of reporting student achievement as part of federal accountability. • Cognitive Labs describe students' engagement with tasks and items and provide confirmation of content measurement.	• *Cognitive Laboratories Technical Report (AIR, 2013)* • *Smarter Balanced Assessment Consortium:* Achievement Level Setting Final Report (January 2015)

Intended Purpose 7:

Assess students' achievement in ELA/literacy and mathematics in a manner that is equitable for all students and subgroups of students.

Intended Purpose 7 is supported by validity evidence related to content, internal structure, and response processes. Table 1-7 summarizes the sources of validity internal to this report and lists pieces of evidence found outside of this report for Intended Purpose 7.

TABLE 1-7 VALIDITY EVIDENCE THAT SUPPORTS INTENDED PURPOSE 7

Summary of Evidence in this Report	List of Other Evidence Sources
Evidence Related to Test Content	
See Chapters 3, 4, 6 • Bias is minimized through Universal Design and accessibility resources. • Assessments are administered in a standardized manner sufficient to yield data that supports valid inferences. • Computer adaptive assessments that meet blueprint constraints are consistently delivered to all students and subgroups of students.	• *Smarter Balanced Assessment Consortium: Accommodations for English Language Learners and Students with Disabilities: A Research-Based Decision Algorithm* (Abedi & Ewers, February 2013) • *Smarter Balanced Assessment Consortium: General Accessibility Guidelines* (April 2012) • Smarter Balanced: Online Test Administration Manual (September, 2017) • *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Implementation Guide* (November 2014) • Smarter Balanced Blueprint Fidelity Studies
Evidence Related to Internal Structure	
See Chapters 2, 3, 6 • The assessment supports precise measurement and consistent classification for all students. • Differential Item Functioning Analysis completed for all items across all required subgroups. • Multidisciplinary data review enacted to resolve each observed incident of DIF.	• 2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 10)
Evidence Related to Response Processes	
See Chapters 3, 4, 6 • Bias is minimized through Universal Design and accessibility resources. • Assessments are administered in a standardized manner sufficient to yield data that supports valid inferences. • Cognitive Labs describe students' engagement with tasks and items and provide confirmation of content measurement.	• *Cognitive Laboratories Technical Report (AIR, 2013)* • *Development Process (NGA Center & CCSSO, 2016)* • *Usability, Accessibility, and Accommodations Guidelines (Smarter Balanced, 2107)*

Conclusion for Summative Test Validity Results

Validation is a perpetual endeavor in which additional evidence can be provided but one can never absolutely "assert" an assessment is perfectly valid (Haertel, 1999). This is particularly true for the many purposes typically placed on tests. Program requirements are often subject to change and the populations assessed change over time. Nonetheless, at some point decisions must be made regarding whether sufficient evidence exists to justify the use of a test for a particular purpose. A review of the purpose statements and the available validity evidence determines the degree to which the principles outlined here have been realized. Most of this report focuses on describing some of the essential validity elements required for necessary evidence. The essential validity elements presented here constitute critical evidence "relevant to the technical quality of a testing system" (AERA et al., 2014, p. 22).

References

Abedi, J., & Ewers, N. (February, 2013). *Accommodations for English Language Learners and Students with Disabilities: A Research Based Decision Algorithm*. Smarter Balanced Assessment Consortium. http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/08/Accomodations-for-under-represented-students.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Institute for Research. (2013). *Cognitive laboratories technical report.* Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2016). *Development Process*. Washington, DC: Author. Retrieved from http://www.corestandards.org/about-the-standards/development-process/

Bhola, D., Impara, J.C., & Buckendahl, C. (2003). Aligning Tests with States' Content Standards: Methods and Issues. *Educational Measurement: Issues and Practice*. 22(3):21 - 29

Campbell, *D. T., &* Fiske, *D. W. (*1959*)*. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.

Conley, D. T., Drummond, K. V., de Gonzalez, A., Rooseboom, J., & Stout. O. (2011). *Reaching the goal: The applicability and importance of the Common Core State Standards to college and career readiness*. Eugene, OR: Educational Policy Improvement Center.

Crocker, L., Miller, M. D., & Franks, E. A. (1989). Quantitative methods for assessing the fit between test and curriculum. Applied Measurement in Education, 2, 179-194.

Cronbach, L. J. (1971).Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Doorey, N., & Polikoff, M. (2016). *Evaluating the content and quality of next generation assessments*. Washington DC: Thomas B. Fordham Institute.

Educational Testing Service (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.

Educational Testing Service (2015). *Study of the relationship between the early assessment program and the smarter balanced field tests*. Sacramento: California Department of Education.

Haertel, E.H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice*, 18(4), 5-9.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.

Kane, M. (2006). Validation. In R. L. Brennan (Ed). *Educational measurement* (4th ed., pp. 17-64). Washington, DC: American Council on Education/Praeger.

Linn, R. L. (2006). The standards for educational and psychological Testing: Guidance in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 27-38), Mahwah, NJ: Lawrence Erlbaum.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research, 79*, 1-76.

Rothman, R, Slattery, J. B., Vranek, J. L. & Resnick, L. B. (May, 2002). *Benchmarking and Alignment of Standards and Testing. CSE Technical Report*. National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA. 35 pp.

Schultz, S. R., Michaels, H. R., Dvorak, R. N. & Wiley, C. R. H. (2016). *Evaluating the Content and Quality of Next Generation High School Assessments. Final Report.* Human Resources Research Organization. Alexandria, VA.

Sireci, S. G. (2013). Agreeing on validity arguments. *Journal of Educational Measurement*. 50, 1, 99–104.

Sireci, S. G. (2012). *Smarter Balanced Assessment Consortium: Comprehensive research agenda*. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

Sireci, S.G. (1998a). Gathering and analyzing content validity data. Educational Assessment, 5, 299-321.

Smarter Balanced Assessment Consortium. (2015). *End of grant report*. Los Angeles, CA: Author. Retrieved from https://portal.smarterbalanced.org/library/en/v1.0/end-of-grant-report.pdf

Smarter Balanced Assessment Consortium. (September, 2017). *Smarter Balanced: Online Test Administration Manual. Available at https://portal.smarterbalanced.org/library/en/v2.0/online-test-administration-manual.docx*

Smarter Balanced Asssessment Consortium. (2016). 2013-2014 Technical Report. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/2013-14_Technical_Report.pdf

Smarter Balanced Assessment Consortium. (January, 2015). *Achievement level setting final report.* Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

Chapter 2: Reliability, Precision and Errors of Measurement

Introduction

This chapter addresses the technical quality of operational test functioning with regard to precision and reliability. Part of the test validity argument is that scores must be consistent and precise enough to be useful for intended purposes. If scores are to be meaningful, tests should deliver the same results under repeated administrations to the same student or for students of the same ability. In addition, the range of certainty around the score should be small enough to support educational decisions. The reliability and precision of a test are examined through analysis of measurement error and other test properties in simulated and operational conditions. For example, the reliability of a test may be assessed in part by verifying that different test forms follow the same blueprint. In computer adaptive testing (CAT), one cannot expect the same set of items to be administered to the same examinee more than once. Consequently, reliability is inferred from internal test properties, including test length and the information provided by item parameters. Items with difficulty parameters appropriate to examinee ability, and those with higher discrimination values provide more information. Longer tests give more information because they provide more certainty about student functioning. Smarter Balanced uses an adaptive model because adaptive tests are customized to each student, thereby yielding lower error and greater reliability than fixed form tests of the same length. Standard errors of measurement, the inverse of the square root of information, are related to reliability in that they represent the standard deviation of repeated test scores.

Simulation Studies for 2016-17 Operational Summative Tests

For Smarter Balanced tests with an adaptive component, test reliability is estimated through simulations conducted using the operational summative item pool. For fixed form tests, reliability and measurement error are calculated using the number of items and their psychometric properties relative to the population.

The National Center for Research on Evaluation, Standards, & Student Testing (CRESST) conducted simulation studies for the 2016-17 tests using packaged pools for this administration along with its own simulation engine (CRESST, February 2017). The results of the CRESST study serve as a baseline for service providers that deliver Smarter Balanced assessments. American Institutes for Research also conducted a simulation study of the CAT portion of the summative tests (AIR, October 2016).

Results from CRESST's simulation are presented here. For each grade and content area, true ability (theta) values for 1,000 simulated examinees were created using the consortium-wide, grade-specific population mean and standard deviations from the 2014-2015 summative administration as shown in Table 2-1.

TABLE 2-1 POPULATION PARAMETERS USED TO GENERATE ABILITY DISTRIBUTIONS FOR SIMULATED TEST ADMINISTRATIONS

Grade	ELA/Literary		Mathematics	
	Mean	SD	Mean	SD
3	-1.11	1.02	-1.16	1.00
4	-0.62	1.07	-0.66	1.02
5	-0.15	1.08	-0.31	1.13
6	0.11	1.06	-0.05	1.27
7	0.37	1.12	0.15	1.36
8	0.60	1.10	0.34	1.45
11	1.05	1.26	0.57	1.56

Using the CRESST adaptive algorithm with the operational pools, test events were created for the simulated examinees. Estimated ability ($\hat{\theta}$) was calculated from the simulated tests using maximum likelihood estimation (MLE) as described in the Smarter Balanced test scoring specifications (AIR, 2014a). In the 2016-17 administration, the test scoring algorithm resolved extreme scores by using the highest and lowest obtainable thetas (HOT and LOT) and corresponding scale scores (HOSS and LOSS). Theta-estimates less than LOT or higher than HOT were assigned LOT and HOT values respectively, which correspond to LOSS and HOSS after thetas are transformed to scale scores. These limits prevent undesirable extreme values from occurring in public reporting.

TABLE 2-2 HOT/LOT VALUES IN LOGIT UNITS AND PERCENTAGES OF AFFECTED SIMULATION RESULTS

Grade	Obtainable Score Range		Percentage of Affected Scores	
	LOT	HOT	LOT	HOT
English Language Arts/Literacy				
3	-4.59	1.34	0.4	0.0
4	-4.40	1.80	0.2	0.0
5	-3.58	2.25	0.0	0.0
6	-3.48	2.51	0.1	0.0
7	-2.91	2.75	0.3	0.0
8	-2.57	3.04	0.0	0.0
11	-2.44	3.34	0.0	0.0
Mathematics				
3	-4.11	1.33	0.0	0.0
4	-3.92	1.82	0.2	0.0
5	-3.73	2.33	0.7	0.0
6	-3.53	2.95	0.2	0.0
7	-3.34	3.32	1.0	0.0
8	-3.15	3.63	0.7	0.0
11	-2.96	4.38	0.2	0.0

Statistics computed from the simulations include the following:

- Bias: the statistical bias of the estimated theta parameter. This is a test of the assumption that error is randomly distributed around true ability. It is a measure of whether scores systematically underestimate or overestimate ability.
- Mean squared error (MSE): This is a measure of the magnitude of difference between true and estimated theta. The square root of MSE, denoted RMSE, is reported in tables of this chapter.
- Significance of the bias: indicator of the statistical significance of bias.
- Average standard error of the estimated theta: This is the average of the simulated standard error of measurement over all examinees. It is the marginal reliability for the simulated population.
- Standard error of estimates of theta at the 5th, 25th, 75th, and 95th percentiles.
- Percentage of students' estimated theta falling outside the 95% and 99% confidence intervals.

Computational details of each statistic are provided below. Bias is computed as

$$bias = N^{-1}\sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right), \quad (1)$$

and the error variance of the estimated bias is

$$var(bias) = \frac{1}{N(N^{-1})}\sum_{i=1}^{N}\left(\theta_i - \bar{\hat{\theta}}_i\right)^2, \quad (2)$$

Where $\bar{\hat{\theta}}_i$ is the average of the $\hat{\theta}_i$ and N denotes the number of simulees (N=1000 for all conditions). Statistical significance of the bias is tested using a z-test,

$$z = \frac{bias}{\sqrt{var(bias)}}, \quad (3)$$

for which the p-value for a two-tailed test is reported. The mean squared error (*MSE*) in the estimated scores is:

$$MSE = N^{-1}\sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right)^2, \quad (4)$$

and its square root is the root mean squared error (RMSE). Marginal reliability of the simulated tests is estimated as

$$\bar{\rho} = 1 - \frac{MSE}{var(\hat{\theta})}, \quad (5)$$

The average standard error of the score estimates is

$$mean(SE) = \sqrt{N^{-1}\sum_{i=1}^{N} SE(\hat{\theta}_i)^2}, \quad (6)$$

where $SE(\hat{\theta}_i)$ is the standard error of the estimated score for simulee *i*. Miss rates for the 95% and 99% confidence intervals are the percentage of cases for which the confidence intervals computed

from the score estimates and standard errors do not contain the true score. To this end, a t-statistic is computed for each case:

$$t = \frac{\theta_i - \hat{\theta}_i}{se(\hat{\theta}_i)}, \tag{7}$$

where $\hat{\theta}_i$ is the ability estimate for individual *i*, and θ_i is the true score for individual *i*. The percentage of times that a student's estimated theta falls outside the confidence interval is determined by comparing the absolute value of the *t*-statistic to a critical value of 1.96 for the 95% coverage and to 2.58 for the 99% coverage.

As shown in table 2-3 and 2-4, bias in overall scores is both small and insignificant for both ELA/literacy and mathematics. Claim scores do include some systematic bias. This is likely caused by application of HOT and LOT values. Bias at very high and very low levels of achievement due to the application of HOT and LOT values may have little or no impact since claim scores are reported in terms of 'above standard', 'near standard', and 'below standard' rather than scale values.

TABLE 2-3 BIAS OF THE ESTIMATED PROFICIENCIES: ENGLISH LANGUAGE ARTS/LITERACY

Grade	Bias	SE(bias)	*p value*	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall English Language Arts/Literacy						
3	0.02	0.03	0.58	0.09	4.60	0.80
4	0.01	0.03	0.70	0.10	4.10	0.90
5	0.00	0.03	0.98	0.09	4.90	0.80
6	0.01	0.03	0.80	0.10	4.50	0.90
7	0.02	0.03	0.48	0.19	4.50	0.90
8	0.01	0.03	0.88	0.14	4.50	0.60
11	-0.02	0.03	0.62	0.15	4.60	0.70
Claim 1: Reading						
3	0.04	0.03	0.20	0.28	4.70	1.40
4	0.07	0.03	0.05	0.40	5.30	1.90
5	0.05	0.03	0.19	0.32	4.20	1.60
6	0.08	0.03	0.01	0.45	4.50	1.10
7	0.05	0.03	0.13	0.49	5.70	1.70
8	0.03	0.03	0.37	0.44	6.20	1.80
11	0.06	0.04	0.13	0.40	5.00	1.60
Claim 2: Writing						
3	0.04	0.03	0.17	0.27	4.90	0.80
4	0.01	0.03	0.85	0.29	5.50	1.30
5	0.01	0.03	0.82	0.28	4.10	1.40
6	0.01	0.03	0.71	0.26	4.50	1.30
7	0.04	0.03	0.20	0.40	4.60	0.80
8	0.02	0.03	0.48	0.35	4.60	1.00
11	0.07	0.04	0.09	0.40	6.50	1.90
Claim 3: Speaking/Listening						
3	0.16	0.03	0.00	0.92	9.90	5.90
4	0.13	0.03	0.00	0.78	7.30	4.70
5	0.07	0.03	0.05	0.71	7.00	3.70
6	0.10	0.03	0.00	0.72	6.30	3.10
7	0.04	0.03	0.20	0.75	8.60	2.60
8	0.02	0.03	0.53	0.78	8.30	3.90
11	0.04	0.04	0.29	0.86	8.30	5.00
Claim 4: Research						
3	0.20	0.03	0.00	0.96	12.90	8.60
4	0.19	0.03	0.00	0.90	10.20	6.20
5	0.08	0.03	0.02	0.59	7.60	4.70
6	0.15	0.03	0.00	0.89	11.30	7.40
7	0.19	0.03	0.00	1.05	19.10	8.70
8	0.09	0.03	0.01	0.73	13.10	7.10
11	0.07	0.04	0.07	0.69	9.30	4.70

TABLE 2-4 BIAS OF THE ESTIMATED PROFICIENCIES: MATHEMATICS

Grade	Bias	SE(Bias)	*p value*	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall Mathematics						
3	0.00	0.03	0.89	0.07	4.60	1.20
4	0.01	0.03	0.86	0.07	4.70	0.70
5	0.03	0.04	0.41	0.12	5.20	0.90
6	0.01	0.04	0.80	0.12	4.00	0.80
7	0.00	0.04	0.99	0.15	4.20	1.10
8	-0.01	0.05	0.90	0.17	5.20	0.90
11	0.01	0.05	0.86	0.21	4.70	1.10
Claim 1: Concepts and Procedures						
3	0.00	0.03	0.96	0.12	5.00	0.90
4	0.03	0.03	0.37	0.15	4.90	0.90
5	0.05	0.04	0.15	0.24	4.70	1.40
6	0.01	0.04	0.80	0.22	5.30	0.90
7	0.05	0.04	0.24	0.34	6.00	1.70
8	0.02	0.05	0.68	0.29	4.60	1.00
11	0.04	0.05	0.41	0.35	4.80	1.20
Claim 2/4: Problem Solving/ Modeling and Data Analysis						
3	0.11	0.03	0.00	0.47	11.10	6.00
4	0.06	0.03	0.06	0.40	7.60	3.30
5	0.26	0.04	0.00	0.98	15.50	8.50
6	0.13	0.04	0.00	0.66	9.80	4.10
7	0.24	0.04	0.00	1.20	16.80	7.70
8	0.35	0.05	0.00	1.64	20.00	11.80
11	0.33	0.05	0.00	1.56	17.10	7.40
Claim 3: Communicating Reasoning						
3	0.16	0.03	0.00	0.57	12.10	7.10
4	0.10	0.03	0.00	0.44	8.00	4.40
5	0.16	0.04	0.00	0.60	8.00	3.80
6	0.20	0.04	0.00	0.79	8.30	4.90
7	0.21	0.04	0.00	1.03	8.80	4.40
8	0.13	0.05	0.01	0.83	7.30	3.00
11	0.16	0.05	0.00	1.18	9.30	4.50

Table 2-5 and table 2-6 show marginal reliability (mean ρ) and precision, for both the overall test and for each reported claim. As expected, overall estimated reliability coefficients are high and in the acceptable range for a large-scale, high-stakes test. Reliability estimates at the claim level are lower, and errors are higher. Claims with fewer items and fewer points from the adaptive section of the test exhibit the lowest reliability. (These are Claims 3 and 4 in English Language Arts/literacy (ELA/literacy) and Claims 2/4 and 3 in mathematics). This shows the importance of incorporating errors in claim-level reports.

Table 2-7 shows that error at the high end of the achievement distribution is smaller than the overall average error, while error in the lower deciles of student achievement is higher than the overall average error. This pattern is due to the fact that the item pools tend to be relatively difficult compared to the student population. The adaptive nature of the Smarter Balanced assessment mitigates, but does not entirely overcome the difference between pool difficulty and student achievement. One reason for this limitation is that the performance task (PT) segment of the assessment is not adaptive. Chapter 4 contains information about the difficulty and other attributes of the item pools.

TABLE 2-5 OVERALL SCORE AND CLAIM SCORE PRECISION/RELIABILITY: ENGLISH LANGUAGE ARTS/LITERACY

Grade	Mean Number of Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
Overall English Language Arts/Literacy					
3	45.42	1.08	0.28	0.30	0.9
4	45.53	1.13	0.31	0.31	0.9
5	45.61	1.12	0.30	0.30	0.9
6	43.29	1.12	0.31	0.32	0.9
7	43.13	1.08	0.42	0.43	0.8
8	43.38	1.11	0.38	0.37	0.8
11	45.33	1.16	0.39	0.39	0.8
Claim 1: Reading					
3	16.00	1.18	0.48	0.53	0.8
4	16.00	1.29	0.55	0.63	0.7
5	16.00	1.25	0.53	0.57	0.7
6	14.00	1.29	0.62	0.67	0.7
7	14.00	1.21	0.71	0.70	0.6
8	14.00	1.23	0.66	0.66	0.7
11	16.00	1.41	0.59	0.63	0.8
Claim 2: Writing					
3	12.00	1.19	0.50	0.52	0.8
4	12.00	1.24	0.52	0.54	0.8
5	12.00	1.22	0.51	0.53	0.8
6	12.00	1.22	0.49	0.51	0.8
7	12.00	1.17	0.66	0.64	0.7
8	12.00	1.21	0.59	0.59	0.7
11	12.00	1.39	0.61	0.64	0.7
Claim 3: Speaking/Listening					
3	8.98	1.45	0.80	0.96	0.5
4	8.99	1.43	0.80	0.88	0.6
5	8.95	1.36	0.83	0.84	0.6
6	8.96	1.40	0.82	0.85	0.6
7	8.98	1.22	0.88	0.86	0.5
8	8.97	1.33	0.87	0.88	0.5
11	8.97	1.50	0.89	0.93	0.6
Claim 4: Research					
3	8.44	1.50	0.66	0.98	0.5
4	8.54	1.48	0.73	0.95	0.5
5	8.66	1.37	0.63	0.77	0.6
6	8.34	1.48	0.70	0.94	0.6
7	8.15	1.34	0.82	1.03	0.4
8	8.41	1.30	0.75	0.85	0.5
11	8.34	1.49	0.76	0.83	0.6

TABLE 2-6 OVERALL SCORE AND CLAIM SCORE PRECISION/RELIABILITY: MATHEMATICS

Grade	Mean Number of Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
Overall Mathematics					
3	39.72	1.03	0.25	0.26	0.94
4	39.00	1.06	0.26	0.27	0.93
5	39.66	1.21	0.31	0.35	0.91
6	39.00	1.33	0.33	0.35	0.93
7	39.44	1.39	0.38	0.39	0.92
8	38.83	1.49	0.40	0.41	0.93
11	41.11	1.61	0.44	0.46	0.92
Claim 1: Concepts and Procedures					
3	20.00	1.06	0.34	0.35	0.89
4	20.00	1.12	0.36	0.39	0.88
5	20.00	1.28	0.43	0.49	0.85
6	19.00	1.38	0.44	0.47	0.88
7	20.00	1.47	0.51	0.58	0.85
8	20.00	1.53	0.54	0.54	0.88
11	22.00	1.66	0.57	0.60	0.87
Claim 2/4: Problem Solving/ Modeling and Data Analysis					
3	9.87	1.30	0.50	0.69	0.72
4	9.43	1.25	0.53	0.63	0.74
5	9.80	1.62	0.59	0.99	0.63
6	9.92	1.57	0.65	0.81	0.73
7	10.00	1.84	0.72	1.10	0.65
8	9.32	1.99	0.78	1.28	0.59
11	9.19	2.05	0.85	1.25	0.63
Claim 3: Communicating Reasoning					
3	9.85	1.30	0.56	0.76	0.66
4	9.57	1.28	0.54	0.66	0.73
5	9.86	1.42	0.60	0.78	0.70

Grade	Mean Number of Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
6	10.08	1.59	0.71	0.89	0.69
7	9.44	1.70	0.81	1.02	0.64
8	9.51	1.72	0.83	0.91	0.72
11	9.92	1.88	0.92	1.09	0.67

TABLE 2-7 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
English Language Arts/Literacy											
3	0.47	0.32	0.29	0.27	0.26	0.25	0.25	0.25	0.25	0.28	0.29
4	0.48	0.35	0.31	0.30	0.29	0.28	0.28	0.27	0.27	0.30	0.31
5	0.42	0.32	0.29	0.28	0.28	0.28	0.27	0.28	0.28	0.31	0.30
6	0.51	0.35	0.31	0.30	0.28	0.27	0.27	0.27	0.28	0.31	0.31
7	0.53	0.38	0.34	0.32	0.30	0.29	0.29	0.29	0.29	0.30	0.44
8	0.47	0.35	0.32	0.31	0.30	0.29	0.29	0.30	0.30	0.32	0.39
11	0.52	0.41	0.37	0.34	0.33	0.31	0.31	0.30	0.31	0.34	0.40
Mathematics											
3	0.38	0.29	0.25	0.24	0.23	0.22	0.22	0.21	0.21	0.24	0.25
4	0.42	0.31	0.27	0.24	0.24	0.22	0.22	0.22	0.21	0.24	0.27
5	0.58	0.41	0.33	0.30	0.27	0.24	0.23	0.21	0.21	0.23	0.33
6	0.56	0.41	0.35	0.32	0.30	0.28	0.26	0.25	0.24	0.26	0.35
7	0.71	0.54	0.46	0.41	0.36	0.32	0.29	0.26	0.23	0.24	0.41
8	0.70	0.54	0.48	0.43	0.40	0.37	0.33	0.29	0.26	0.26	0.43
11	0.81	0.63	0.52	0.46	0.41	0.36	0.33	0.30	0.27	0.26	0.48

Tests for Special Populations

The Consortium developed assessments in Braille for mathematics and ELA/literacy. Assessments in mathematics were also developed for translated glossaries, stacked Spanish translations and American Sign Language. American Sign Language pools were also developed for Claim 3 (Listening) in English Language Arts. The same set of items was used for all translated glossary pools in Arabic, Cantonese, Mandarin, Spanish, Tagalog, Korean, Punjabi, Russian, Ukrainian and Vietnamese. These tests followed the blueprints and were identical to the standard test except for the item pool. Students eligible for these test versions were given the appropriate pool.

Below, we provide information about the reliability of the accommodated summative assessment, based on simulated test administrations using the accommodated item pools. Specifically, we include tables depicting the bias of estimated proficiencies, overall score and claim score precision/ reliability, and average standard errors by grade and deciles of true proficiency scores. Table 2-8 and Table 2-9 show the numbers of items in the CAT pools for English Language Arts/Literacy and Mathematics, respectively. Further details of simulations for accommodated item pools can be found in the full simulation report (National Center for Research on Evaluation, Standards, & Student Testing [CRESST], February 2017).

TABLE 2-8 NUMBER OF ENGLISH LANGUAGE ARTS/LITERACY ITEMS BY GRADE ACROSS POOLS

Grade	General	ASL	Braille
3	879	879	292
4	835	835	279
5	828	828	299
6	769	769	270
7	699	699	245
8	751	751	282
11	2435	2435	528

TABLE 2-9 NUMBER OF MATHEMATICS ITEMS BY GRADE ACROSS POOLS

Grade	General	ASL	Braille	Translated Glossaries	Spanish
3	1272	909	419	252	409
4	1374	913	371	246	401
5	1295	892	392	254	417
6	1140	811	389	277	399
7	1030	757	371	255	357
8	901	670	298	236	320
11	2108	1772	418	298	538

Table 2-10 through Table 2-15 show the bias of the estimated proficiencies for the American Sign Language (ASL) and Braille pools for ELA/Literacy and ASL, Braille, Translated Glossary and Spanish accommodation pools for Mathematics. Overall bias ranged from -0.01 and 0.05 for the ASL and Braille pools and indicates very low evidence of bias in the overall scores. However, there is evidence of bias in the claim score estimates. This bias appears to be due to the assignment of the LOT and HOT values for examinees with extreme score estimates for a given claim—in particular, those examinees with an infinite ML score estimate due to a perfect score patterns (i.e., achieving either the minimum score for all items or the maximum for all items). Such score patterns are of course far more likely within a claim (based on a relatively small number of items) than for the full test. Bias at very high and very low levels of achievement due to the application of HOT and LOT values may have little or no impact since claim scores are reported in terms of 'above standard', 'near standard', and 'below standard' rather than scale values.

Confidence interval miss rates for overall scores are very close to their expected levels. The overall score miss rate for the 95% confidence interval—expected to be 5%—ranged from 3.6% to 5.9%, while the miss rate for the 99% confidence interval—expected to be 1%—ranges from 0.4% to 1.8%. Taken together with the results concerning average bias, these confidence interval miss rates suggest that the standard errors of measurement for the overall score estimates are well-calibrated (i.e., correctly reflecting the level of score uncertainty) across all pools for ELA/literacy and mathematics.

The confidence interval miss rates for the claim scores are less consistent and—for Claims 3 and 4, in particular—show evidence of poor calibration. This is not surprising, however, given the bias observed in these score estimates. It is likely that the deviations of the miss rates from their expected values are due to the assignment of the LOT and HOT for examinees with perfect item score patterns. Because such patterns are relatively common for the small number of items in a claim, the LOT or HOT is a poor estimate of the true score for many examinees. This makes it less likely that the confidence interval around the LOT/HOT will include the true score, increasing the miss rate.

TABLE 2-10 BIAS OF THE ESTIMATED PROFICIENCIES: ENGLISH LANGUAGE ARTS/LITERACY AMERICAN SIGN LANGUAGE (ASL) POOL (FROM SIMULATION STUDY)

Grade	Bias	SE(bias)	*p* value	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall English Language Arts/Literacy						
3	0.02	0.03	0.58	0.09	5.0	0.7
4	0.00	0.03	0.93	0.10	5.5	1.3
5	0.00	0.03	0.98	0.09	4.9	0.9
6	0.01	0.03	0.67	0.10	4.6	0.7
7	0.01	0.04	0.84	0.11	5.2	0.8
8	-0.01	0.04	0.78	0.11	4.3	0.6
11	0.02	0.04	0.64	0.13	5.4	1.3
Claim 1: Reading						
3	0.03	0.03	0.39	0.27	5.1	1.2
4	0.05	0.03	0.14	0.38	5.3	1.7
5	0.05	0.03	0.16	0.34	6.1	2.1
6	0.06	0.03	0.09	0.47	5.4	2.0
7	0.05	0.04	0.18	0.38	5.5	1.6
8	0.02	0.04	0.57	0.38	5.8	1.5
11	0.06	0.04	0.11	0.37	4.4	1.3
Claim 2: Writing						
3	0.06	0.03	0.08	0.30	5.0	1.9
4	0.01	0.03	0.67	0.32	5.1	1.1
5	-0.01	0.03	0.72	0.25	5.5	1.2
6	0.04	0.03	0.28	0.28	5.0	1.0
7	0.01	0.04	0.71	0.34	5.8	1.7
8	0.01	0.04	0.84	0.31	4.9	0.8
11	0.04	0.04	0.30	0.40	5.1	2.0
Claim 3: Speaking/Listening						
3	0.19	0.03	0.00	0.97	11.6	7.7
4	0.09	0.03	0.01	0.83	8.1	4.8
5	0.09	0.03	0.01	0.81	8.8	5.7
6	0.13	0.03	0.00	0.92	8.8	4.9
7	0.07	0.04	0.05	0.94	8.5	5.8
8	0.06	0.04	0.10	0.94	8.1	5.0
11	0.03	0.04	0.40	0.90	10.4	7.0
Claim 4: Research						
3	0.22	0.03	0.00	1.01	13.0	9.1
4	0.18	0.03	0.00	0.97	11.1	6.7
5	0.08	0.03	0.03	0.59	9.2	5.0
6	0.20	0.03	0.00	0.94	12.5	7.9
7	0.18	0.04	0.00	0.82	10.9	6.9
8	0.08	0.04	0.03	0.65	9.4	6.0
11	0.10	0.04	0.01	0.71	9.5	4.8

TABLE 2-11 BIAS OF THE ESTIMATED PROFICIENCIES: ENGLISH LANGUAGE ARTS/LITERACY BRAILLE POOL (FROM SIMULATION STUDY)

Grade	Bias	SE(bias)	*p* value	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall English Language Arts/Literacy						
3	0.00	0.03	0.94	0.08	3.8	1.0
4	0.02	0.03	0.65	0.10	4.8	0.8
5	0.00	0.03	0.91	0.10	5.1	1.0
6	0.02	0.03	0.56	0.11	5.2	1.1
7	0.00	0.04	0.98	0.11	4.9	0.4
8	0.01	0.04	0.78	0.11	3.8	0.5
11	0.02	0.04	0.56	0.14	5.7	1.8
Claim 1: Reading						
3	0.02	0.03	0.62	0.24	4.0	0.8
4	0.11	0.03	0.00	0.54	7.2	3.4
5	0.02	0.03	0.66	0.31	6.5	1.6
6	0.05	0.03	0.12	0.45	5.8	1.3
7	0.04	0.04	0.30	0.42	6.0	0.9
8	0.04	0.04	0.25	0.39	5.2	1.8
11	0.06	0.04	0.11	0.39	6.8	2.2
Claim 2: Writing						
3	0.01	0.03	0.67	0.28	4.4	1.3
4	0.01	0.03	0.77	0.29	5.0	1.1
5	-0.01	0.03	0.77	0.32	5.2	1.1
6	0.04	0.03	0.21	0.30	4.5	1.3
7	0.02	0.04	0.53	0.32	5.5	1.1
8	0.01	0.04	0.68	0.32	5.2	1.0
11	0.04	0.04	0.38	0.43	6.8	3.0
Claim 3: Speaking/Listening						
3	0.19	0.03	0.00	0.94	10.6	7.0
4	0.12	0.03	0.00	0.77	6.5	3.7
5	0.09	0.03	0.01	0.85	8.5	5.2
6	0.05	0.03	0.11	0.86	6.8	3.4
7	-0.03	0.04	0.39	0.67	6.2	3.5
8	0.01	0.04	0.69	0.85	8.2	5.1
11	0.07	0.04	0.08	1.02	9.8	6.1
Claim 4: Research						
3	0.13	0.03	0.00	0.83	11.9	7.2
4	0.16	0.03	0.00	0.97	10.9	6.5
5	0.08	0.03	0.02	0.59	9.3	5.3
6	0.20	0.03	0.00	0.88	11.4	6.9
7	0.21	0.04	0.00	0.90	12.8	7.7
8	0.12	0.04	0.00	0.79	8.5	5.3
11	0.17	0.04	0.00	0.87	11.0	6.2

TABLE 2-12 BIAS OF THE ESTIMATED PROFICIENCIES: MATHEMATICS AMERICAN SIGN LANGUAGE (ASL) POOL (FROM SIMULATION STUDY)

Grade	Bias	SE(Bias)	*p* value	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall Mathematics						
3	0.00	0.03	0.88	0.07	5.9	1.2
4	0.01	0.03	0.83	0.07	3.3	0.6
5	0.03	0.04	0.41	0.13	5.5	1.0
6	0.01	0.04	0.71	0.12	4.3	0.6
7	0.02	0.04	0.57	0.17	3.8	0.5
8	0.02	0.05	0.72	0.22	5.5	1.1
11	0.01	0.05	0.84	0.29	5.6	1.2
Claim 1: Concepts and Procedures						
3	-0.01	0.03	0.79	0.14	6.3	1.0
4	0.03	0.03	0.35	0.15	4.3	0.6
5	0.06	0.04	0.07	0.24	4.2	0.7
6	0.01	0.04	0.75	0.22	4.7	0.8
7	0.05	0.04	0.21	0.29	6.2	1.6
8	0.05	0.05	0.26	0.36	4.9	1.6
11	0.06	0.05	0.24	0.56	5.6	1.7
Claim 2/4: Problem Solving/ Modeling and Data Analysis						
3	0.06	0.03	0.06	0.36	8.2	3.9
4	0.11	0.03	0.00	0.61	9.5	5.2
5	0.25	0.04	0.00	0.96	15.1	9.3
6	0.26	0.04	0.00	1.02	15.7	8.0
7	0.37	0.04	0.00	1.58	18.2	9.9
8	0.49	0.05	0.00	2.12	26.5	14.0
11	0.37	0.05	0.00	1.65	18.1	8.5
Claim 3: Communicating Reasoning						
3	0.21	0.03	0.00	0.72	15.7	9.8
4	0.18	0.03	0.00	0.66	12.4	7.8
5	0.18	0.04	0.00	0.75	11.1	6.1
6	0.28	0.04	0.00	1.01	12.9	7.2
7	0.33	0.04	0.00	1.41	15.1	8.0
8	0.26	0.05	0.00	1.30	11.7	5.5
11	0.13	0.05	0.01	1.12	8.7	3.0

TABLE 2-13 BIAS OF THE ESTIMATED PROFICIENCIES: MATHEMATICS BRAILLE POOL (FROM SIMULATION STUDY)

Grade	Bias	SE(Bias)	*p* value	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall Mathematics						
3	0.00	0.03	0.92	0.07	5.2	1.0
4	0.01	0.03	0.84	0.08	4.5	1.2
5	0.02	0.04	0.57	0.12	4.3	1.2
6	0.01	0.04	0.76	0.11	5.1	0.6
7	0.01	0.04	0.78	0.16	5.2	1.1
8	0.01	0.05	0.87	0.19	4.4	1.3
11	0.03	0.05	0.49	0.25	4.8	0.9
Claim 1: Concepts and Procedures						
3	-0.02	0.03	0.62	0.13	5.1	0.8
4	0.02	0.03	0.48	0.15	4.6	0.9
5	0.05	0.04	0.21	0.22	4.0	0.9
6	0.01	0.04	0.83	0.21	5.4	1.0
7	0.05	0.04	0.29	0.29	5.4	1.6
8	0.03	0.05	0.49	0.30	4.7	0.6
11	0.07	0.05	0.18	0.46	4.7	0.9
Claim 2/4: Problem Solving/ Modeling and Data Analysis						
3	0.11	0.03	0.00	0.51	11.8	6.7
4	0.10	0.03	0.00	0.56	9.6	5.0
5	0.31	0.04	0.00	1.14	18.4	11.7
6	0.30	0.04	0.00	1.15	17.5	11.0
7	0.34	0.04	0.00	1.47	19.0	10.9
8	0.48	0.05	0.00	2.07	24.2	14.8
11	0.35	0.05	0.00	1.56	17.2	8.9
Claim 3: Communicating Reasoning						
3	0.13	0.03	0.00	0.52	11.5	6.2
4	0.14	0.03	0.00	0.60	9.4	6.4
5	0.16	0.04	0.00	0.66	9.8	5.2
6	0.15	0.04	0.00	0.70	8.0	3.7
7	0.16	0.04	0.00	0.86	7.8	3.9
8	0.14	0.05	0.00	1.04	8.5	5.0
11	0.11	0.05	0.03	0.89	5.7	2.8

TABLE 2-14 BIAS OF THE ESTIMATED PROFICIENCIES: MATHEMATICS TRANSLATED GLOSSARY POOL (FROM SIMULATION STUDY)

Grade	Bias	SE(bias)	*p* value	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall Mathematics						
3	-0.01	0.03	0.84	0.07	5.3	0.8
4	0.00	0.03	0.92	0.08	4.3	1.2
5	0.05	0.04	0.17	0.14	4.8	1.3
6	0.02	0.04	0.60	0.14	4.3	0.6
7	0.02	0.04	0.63	0.18	3.9	0.7
8	0.01	0.05	0.89	0.23	5.2	1.4
11	0.05	0.05	0.31	0.33	4.7	1.5
Claim 1: Concepts and Procedures						
3	-0.02	0.03	0.57	0.13	4.9	0.5
4	0.02	0.03	0.47	0.17	3.9	1.2
5	0.09	0.04	0.02	0.26	3.8	0.8
6	0.03	0.04	0.44	0.25	4.7	1.3
7	0.08	0.04	0.07	0.36	5.5	1.6
8	0.03	0.05	0.45	0.38	4.6	1.3
11	0.10	0.05	0.05	0.63	6.1	1.3
Claim 2/4: Problem Solving/Modeling and Data Analysis						
3	0.11	0.03	0.00	0.56	13.0	6.9
4	0.07	0.03	0.03	0.50	7.5	4.2
5	0.27	0.04	0.00	0.99	15.2	9.3
6	0.26	0.04	0.00	1.06	14.5	6.8
7	0.36	0.04	0.00	1.66	19.6	9.8
8	0.62	0.05	0.00	2.54	30.2	17.2
11	0.43	0.05	0.00	1.75	16.8	8.2
Claim 3: Communicating Reasoning						
3	0.28	0.03	0.00	0.84	17.0	11.6
4	0.13	0.03	0.00	0.54	9.4	5.6
5	0.28	0.04	0.00	1.00	14.7	8.6
6	0.29	0.04	0.00	1.14	14.5	7.8
7	0.21	0.04	0.00	1.08	9.8	5.1
8	0.22	0.05	0.00	1.30	9.9	5.1
11	0.31	0.05	0.00	1.65	12.4	5.2

TABLE 2-15 BIAS OF THE ESTIMATED PROFICIENCIES: MATHEMATICS SPANISH POOL (FROM SIMULATION STUDY)

Grade	Bias	SE(bias)	p value	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall Mathematics						
3	0.00	0.03	0.94	0.07	5.4	1.3
4	0.01	0.03	0.69	0.08	3.6	1.1
5	0.03	0.04	0.39	0.13	5.3	0.9
6	0.02	0.04	0.69	0.13	4.6	0.8
7	0.02	0.04	0.63	0.16	4.7	1.0
8	0.02	0.05	0.59	0.24	5.2	1.1
11	0.01	0.05	0.82	0.29	5.8	1.0
Claim 1: Concepts and Procedures						
3	-0.01	0.03	0.70	0.13	5.0	1.2
4	0.04	0.03	0.22	0.14	3.6	0.7
5	0.07	0.04	0.06	0.25	4.4	0.9
6	0.02	0.04	0.68	0.23	4.7	1.2
7	0.05	0.04	0.23	0.30	5.0	1.5
8	0.06	0.05	0.18	0.38	5.4	1.1
11	0.07	0.05	0.17	0.56	4.9	2.1
Claim 2/4: Problem Solving/Modeling and Data Analysis						
3	0.22	0.03	0.00	0.85	15.6	11.0
4	0.10	0.03	0.00	0.60	9.1	4.7
5	0.25	0.04	0.00	0.97	15.0	8.9
6	0.31	0.04	0.00	1.16	16.6	9.1
7	0.31	0.04	0.00	1.38	16.1	8.2
8	0.49	0.05	0.00	2.15	25.0	12.7
11	0.36	0.05	0.00	1.63	17.7	7.2
Claim 3: Communicating Reasoning						
3	0.28	0.03	0.00	0.91	17.4	12.7
4	0.15	0.03	0.00	0.60	10.8	6.6
5	0.21	0.04	0.00	0.77	11.6	6.2
6	0.27	0.04	0.00	1.06	12.6	6.6
7	0.32	0.04	0.00	1.35	14.1	7.6
8	0.32	0.05	0.00	1.47	13.6	6.7
11	0.12	0.05	0.02	1.17	8.2	3.5

Table 2-16 through Table 2-21 summarize the standard deviation in score estimates, average standard error, square root of the mean squared error (RMSE), and marginal reliability for the overall and claim score reliability/precision for the American Sign Language (ASL) and Braille pools for ELA/Literacy and ASL, Braille, Translated Glossary and Spanish accommodation pools for Mathematics. The results indicate that the standard errors for the overall score estimates across pools are well-calibrated. Average standard errors within each grade closely resemble the RMSE values. There are discrepancies between the average standard errors and the RMSE values for the claim scores, with the average standard error mostly smaller than the RMSE. This result is consistent with the earlier findings concerning average bias in the claim score estimates and the confidence interval miss rates.

Marginal reliability was computed from the RMSE and observed variance in the scale score estimates, as described earlier in this chapter. For the overall score, marginal reliability ranged from 0.88 to 0.94 across subjects and pools. Marginal reliability for the claim scores ranged from 0.72 to 0.82 for Claim 1 (Reading), 0.79 to 0.83 for Claim 2 (Writing), 0.55 to 0.65 for Claim 3 (Speaking/Listening), and 0.57 to 0.69 for Claim 4 (Research) for ELA/literacy. Marginal reliability was higher for mathematics and ranged from 0.79 to 0.89 for Claim 1 (Concepts and Procedures), 0.50 to 0.75 for Claim 2/4 (Problem Solving/Modeling and Data Analysis), and from 0.57 to 0.73 for Claim 3 (Communicating Reasoning) The lower levels of marginal reliability for Claims 3 and 4 are expected, given that these scores are based on fewer items than the scores for Claims 1 and 2.

TABLE 2-16 OVERALL AND CLAIM SCORE PRECISION/RELIABILITY: ENGLISH LANGUAGE ARTS/LITERACY AMERICAN SIGN LANGUAGE (ASL) POOL (FROM SIMULATION STUDY)

Grade	mean # Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\overline{\rho}$
Overall English Language Arts/Literacy					
3	45.4	1.08	0.29	0.30	0.92
4	45.5	1.12	0.31	0.31	0.92
5	45.7	1.12	0.30	0.30	0.93
6	43.3	1.11	0.31	0.32	0.92
7	43.2	1.17	0.33	0.34	0.92
8	43.4	1.15	0.32	0.32	0.92
11	45.3	1.27	0.34	0.37	0.92
Claim 1: Reading					
3	16.0	1.18	0.48	0.52	0.81
4	16.0	1.26	0.56	0.61	0.76
5	16.0	1.25	0.54	0.58	0.78
6	14.0	1.29	0.62	0.69	0.72
7	14.0	1.31	0.59	0.62	0.78
8	14.0	1.29	0.59	0.62	0.77
11	16.0	1.39	0.59	0.61	0.81
Claim 2: Writing					
3	12.0	1.20	0.50	0.55	0.79
4	12.0	1.25	0.52	0.56	0.80
5	12.0	1.21	0.51	0.50	0.83
6	12.0	1.23	0.49	0.53	0.82
7	12.0	1.29	0.54	0.58	0.80
8	12.0	1.25	0.55	0.56	0.80
11	12.0	1.39	0.61	0.63	0.80
Claim 3: Speaking/Listening					
3	9.0	1.47	0.81	0.99	0.55
4	9.0	1.41	0.78	0.91	0.58
5	9.0	1.44	0.83	0.90	0.61
6	9.0	1.45	0.86	0.96	0.56
7	9.0	1.45	0.90	0.97	0.55
8	9.0	1.48	0.92	0.97	0.57
11	9.0	1.51	0.90	0.95	0.60
Claim 4: Research					
3	8.4	1.52	0.66	1.00	0.57
4	8.5	1.51	0.73	0.99	0.57
5	8.7	1.38	0.63	0.77	0.69
6	8.3	1.51	0.71	0.97	0.59
7	8.2	1.46	0.73	0.91	0.62
8	8.4	1.40	0.68	0.81	0.66
11	8.3	1.48	0.75	0.84	0.67

TABLE 2-17 OVERALL AND CLAIM SCORE PRECISION/RELIABILITY: ENGLISH LANGUAGE ARTS/LITERACY BRAILLE POOL (FROM SIMULATION STUDY)

Grade	mean # Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
Overall English Language Arts/Literacy					
3	45.6	1.1	0.29	0.29	0.93
4	45.4	1.1	0.31	0.32	0.92
5	45.6	1.1	0.30	0.31	0.92
6	43.3	1.1	0.32	0.33	0.91
7	43.1	1.2	0.33	0.33	0.92
8	43.2	1.1	0.33	0.33	0.92
11	45.4	1.3	0.35	0.37	0.92
Claim 1: Reading					
3	16	1.2	0.49	0.49	0.82
4	16	1.4	0.59	0.73	0.71
5	16	1.2	0.51	0.56	0.79
6	14	1.3	0.63	0.67	0.72
7	14	1.3	0.61	0.64	0.76
8	14	1.2	0.59	0.62	0.74
11	16	1.4	0.60	0.63	0.80
Claim 2: Writing					
3	12	1.2	0.51	0.53	0.80
4	12	1.2	0.52	0.54	0.81
5	12	1.2	0.53	0.56	0.79
6	12	1.2	0.52	0.55	0.80
7	12	1.3	0.54	0.57	0.80
8	12	1.2	0.56	0.57	0.79
11	12	1.4	0.62	0.65	0.79
Claim 3: Speaking/Listening					
3	8.9	1.5	0.80	0.97	0.57
4	9	1.4	0.79	0.88	0.60
5	9	1.5	0.84	0.92	0.60
6	8.9	1.4	0.86	0.93	0.56
7	9	1.4	0.83	0.82	0.65
8	9	1.4	0.86	0.92	0.58
11	9	1.5	0.92	1.01	0.55
Claim 4: Research					
3	8.6	1.5	0.66	0.91	0.61
4	8.4	1.5	0.73	0.99	0.57
5	8.7	1.4	0.63	0.77	0.69
6	8.4	1.5	0.70	0.94	0.60
7	8.1	1.5	0.76	0.95	0.61
8	8.2	1.4	0.79	0.89	0.62
11	8.4	1.6	0.77	0.93	0.65

TABLE 2-18 OVERALL AND CLAIM SCORE PRECISION/RELIABILITY: MATHEMATICS AMERICAN SIGN LANGUAGE (ASL) POOL (FROM SIMULATION STUDY)

Grade	mean # Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
Overall Mathematics					
3	39.1	1.0	0.24	0.26	0.94
4	38.5	1.1	0.27	0.27	0.93
5	39.9	1.2	0.32	0.36	0.91
6	39.0	1.3	0.34	0.34	0.93
7	39.4	1.4	0.39	0.41	0.92
8	38.8	1.5	0.44	0.47	0.90
11	41.8	1.6	0.49	0.54	0.89
Claim 1: Concepts and Procedures					
3	20.0	1.1	0.34	0.37	0.88
4	20.0	1.1	0.36	0.39	0.88
5	20.0	1.3	0.45	0.49	0.85
6	19.0	1.4	0.44	0.47	0.89
7	20.0	1.5	0.51	0.54	0.86
8	20.0	1.5	0.57	0.60	0.85
11	22.0	1.7	0.69	0.75	0.80
Claim 2/4: Problem Solving/ Modeling and Data Analysis					
3	9.6	1.2	0.48	0.60	0.75
4	9.1	1.4	0.59	0.78	0.67
5	9.9	1.6	0.60	0.98	0.63
6	10.0	1.7	0.64	1.01	0.66
7	10.0	1.9	0.76	1.26	0.57
8	9.3	2.1	0.77	1.45	0.54
11	9.8	2.1	0.83	1.29	0.62
Claim 3: Communicating Reasoning					
3	9.5	1.4	0.55	0.85	0.62
4	9.4	1.4	0.55	0.81	0.66
5	10.0	1.5	0.61	0.87	0.65
6	10.0	1.7	0.76	1.00	0.64
7	9.4	1.8	0.83	1.19	0.57
8	9.5	1.8	0.91	1.14	0.62
11	10.0	1.9	0.97	1.06	0.68

TABLE 2-19 OVERALL AND CLAIM SCORE PRECISION/RELIABILITY: MATHEMATICS BRAILLE POOL (FROM SIMULATION STUDY)

Grade	mean # Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\overline{\rho}$
Overall Mathematics					
3	39.8	1.0	0.25	0.26	0.94
4	38.6	1.1	0.27	0.28	0.93
5	39.9	1.2	0.31	0.34	0.92
6	39.0	1.3	0.33	0.34	0.94
7	39.4	1.4	0.38	0.40	0.92
8	38.8	1.5	0.43	0.44	0.91
11	41.8	1.6	0.47	0.50	0.91
Claim 1: Concepts and Procedures					
3	20.0	1.1	0.34	0.36	0.89
4	20.0	1.1	0.35	0.38	0.88
5	20.0	1.3	0.44	0.47	0.86
6	19.0	1.4	0.44	0.46	0.89
7	20.0	1.5	0.50	0.54	0.87
8	20.0	1.5	0.54	0.55	0.87
1	22.0	1.7	0.63	0.68	0.84
Claim 2/4: Problem Solving/ Modeling and Data Analysis					
3	10.0	1.3	0.50	0.71	0.70
4	9.1	1.3	0.60	0.75	0.68
5	9.9	1.7	0.58	1.07	0.60
6	10.0	1.8	0.64	1.07	0.64
7	10.0	1.9	0.75	1.21	0.59
8	9.3	2.1	0.86	1.44	0.53
11	9.8	2.1	0.85	1.25	0.64
Claim 3: Communicating Reasoning					
3	9.8	1.3	0.56	0.72	0.68
4	9.4	1.3	0.59	0.77	0.67
5	10.0	1.4	0.62	0.81	0.68
6	10.0	1.5	0.71	0.84	0.70
7	9.4	1.6	0.81	0.93	0.68
8	9.5	1.7	0.87	1.02	0.65
11	10.0	1.8	0.92	0.94	0.73

TABLE 2-20 OVERALL AND CLAIM SCORE PRECISION/RELIABILITY: MATHEMATICS TRANSLATED GLOSSARIES POOL (FROM SIMULATION STUDY)

Grade	mean # Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
Overall Mathematics					
3	39.7	1.03	0.25	0.27	0.93
4	37.8	1.06	0.28	0.28	0.93
5	39.7	1.22	0.33	0.38	0.91
6	39.0	1.33	0.35	0.37	0.92
7	38.5	1.42	0.40	0.43	0.91
8	38.1	1.51	0.45	0.48	0.90
11	42.0	1.66	0.53	0.58	0.88
Claim 1: Concepts and Procedures					
3	20.0	1.07	0.34	0.35	0.89
4	20.0	1.12	0.38	0.41	0.87
5	20.0	1.29	0.47	0.51	0.84
6	19.0	1.40	0.46	0.50	0.87
7	20.0	1.52	0.53	0.60	0.84
8	20.0	1.55	0.58	0.61	0.84
11	22.0	1.75	0.71	0.79	0.79
Claim 2/4: Problem Solving/Modeling and Data Analysis					
3	9.7	1.31	0.52	0.75	0.67
4	8.6	1.29	0.61	0.71	0.70
5	9.7	1.62	0.58	0.99	0.62
6	9.8	1.75	0.68	1.03	0.66
7	9.4	1.92	0.77	1.29	0.55
8	9.0	2.26	0.81	1.59	0.50
11	10.0	2.11	0.86	1.32	0.61
Claim 3: Communicating Reasoning					
3	10.0	1.44	0.58	0.91	0.60
4	9.2	1.33	0.56	0.73	0.70
5	10.0	1.57	0.64	1.00	0.60
6	10.2	1.71	0.77	1.07	0.61
7	9.2	1.70	0.83	1.04	0.63
8	9.1	1.81	1.00	1.14	0.60
11	10.0	2.03	1.06	1.28	0.60

TABLE 2-21 OVERALL AND CLAIM SCORE PRECISION/RELIABILITY: MATHEMATICS SPANISH POOL (FROM SIMULATION STUDY)

Grade	mean # Items	SD($\hat{\theta}$)	Mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
Overall Mathematics					
3	39.8	1.03	0.24	0.26	0.94
4	38.6	1.07	0.27	0.28	0.93
5	39.9	1.20	0.32	0.36	0.91
6	39.0	1.34	0.34	0.36	0.93
7	39.4	1.40	0.39	0.40	0.92
8	38.8	1.53	0.46	0.49	0.90
11	41.8	1.61	0.49	0.54	0.89
Claim 1: Concepts and Procedures					
3	20.0	1.06	0.34	0.36	0.89
4	20.0	1.12	0.37	0.38	0.88
5	20.0	1.28	0.45	0.50	0.84
6	19.0	1.39	0.45	0.48	0.88
7	20.0	1.47	0.50	0.55	0.86
8	20.0	1.57	0.59	0.62	0.84
11	22.0	1.69	0.69	0.75	0.80
Claim 2/4: Problem Solving/Modeling and Data Analysis					
3	10.0	1.42	0.59	0.92	0.58
4	9.1	1.34	0.60	0.77	0.67
5	9.9	1.61	0.59	0.99	0.63
6	10.0	1.79	0.65	1.08	0.64
7	10.0	1.86	0.77	1.18	0.60
8	9.3	2.14	0.80	1.46	0.53
11	9.8	2.07	0.85	1.28	0.62
Claim 3: Communicating Reasoning					
3	9.8	1.45	0.57	0.95	0.57
4	9.4	1.36	0.56	0.77	0.68
5	10.0	1.49	0.61	0.88	0.65
6	10.0	1.70	0.76	1.03	0.63
7	9.4	1.80	0.83	1.16	0.58
8	9.5	1.89	0.95	1.21	0.59
11	10.0	1.86	0.97	1.08	0.66

Table 2-22 through Table 2-27 summarize the average standard errors for the overall scores within true score deciles for ASL and Braille for ELA/literacy and ASL, Braille, Translated Glossaries, and Spanish for mathematics. The averages in deciles 4-10 (i.e., for all examinees above the 30th percentile) range from 0.21 to 0.58 for all grade levels. Average standard errors are higher in the lowest deciles and have a range of 0.36 to 0.99 in decile 1. This is consistent with the fact that the item pools tend to have an average level of difficulty that is higher than the average proficiency of the population of examinees.

TABLE 2-22 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES: ENGLISH LANGUAGE ARTS/ LITERACY AMERICAN SIGN LANGUAGE (ASL) POOL (FROM SIMULATION STUDY)

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
3	0.44	0.31	0.28	0.27	0.26	0.25	0.24	0.25	0.25	0.28	0.30
4	0.45	0.33	0.30	0.29	0.28	0.28	0.27	0.27	0.28	0.30	0.32
5	0.40	0.30	0.28	0.28	0.27	0.28	0.27	0.28	0.29	0.31	0.30
6	0.46	0.33	0.30	0.29	0.28	0.27	0.28	0.28	0.28	0.31	0.32
7	0.47	0.36	0.33	0.31	0.30	0.29	0.29	0.29	0.30	0.33	0.33
8	0.44	0.34	0.31	0.30	0.30	0.30	0.30	0.30	0.30	0.33	0.33
11	0.48	0.37	0.34	0.32	0.31	0.31	0.31	0.31	0.32	0.35	0.35

TABLE 2-23 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES: ENGLISH LANGUAGE ARTS/LITERACY BRAILLE POOL (FROM SIMULATION STUDY)

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
3	0.43	0.31	0.29	0.27	0.26	0.26	0.26	0.26	0.26	0.28	0.30
4	0.45	0.32	0.30	0.29	0.28	0.28	0.27	0.27	0.28	0.31	0.31
5	0.38	0.30	0.28	0.28	0.28	0.28	0.28	0.28	0.29	0.32	0.30
6	0.46	0.33	0.30	0.29	0.29	0.29	0.29	0.29	0.29	0.31	0.32
7	0.47	0.36	0.33	0.30	0.29	0.29	0.29	0.29	0.30	0.34	0.33
8	0.42	0.34	0.32	0.32	0.32	0.31	0.32	0.32	0.32	0.35	0.34
11	0.50	0.39	0.35	0.33	0.32	0.31	0.31	0.31	0.33	0.35	0.36

TABLE 2-24 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES: MATHEMATICS AMERICAN SIGN LANGUAGE (ASL) POOL (FROM SIMULATION STUDY)

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
3	.38	.28	.25	.24	.23	.22	.22	.21	.21	.23	.25
4	.46	.33	.28	.25	.24	.23	.22	.22	.22	.24	.28
5	.60	.42	.35	.31	.28	.26	.24	.22	.21	.22	.35
6	.59	.43	.37	.33	.31	.28	.27	.25	.24	.25	.36
7	.76	.58	.48	.42	.37	.33	.29	.26	.23	.24	.43
8	.82	.65	.54	.49	.43	.38	.34	.29	.26	.25	.48
11	.90	.74	.60	.54	.47	.41	.36	.32	.28	.27	.54

TABLE 2-25 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES: MATHEMATICS BRAILLE POOL (FROM SIMULATION STUDY)

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
3	0.36	0.27	0.24	0.23	0.23	0.22	0.22	0.22	0.22	0.25	0.25
4	0.43	0.31	0.27	0.25	0.24	0.23	0.22	0.22	0.23	0.25	0.28
5	0.58	0.42	0.33	0.31	0.28	0.26	0.24	0.23	0.22	0.23	0.34
6	0.56	0.41	0.35	0.32	0.30	0.28	0.26	0.25	0.24	0.27	0.35
7	0.66	0.52	0.44	0.40	0.35	0.32	0.28	0.26	0.24	0.25	0.41
8	0.71	0.57	0.50	0.46	0.42	0.37	0.33	0.30	0.27	0.27	0.45
11	0.84	0.64	0.55	0.47	0.43	0.38	0.35	0.32	0.30	0.30	0.50

TABLE 2-26 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES: MATHEMATICS TRANSLATED GLOSSARIES POOL (FROM SIMULATION STUDY)

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
3	0.38	0.29	0.26	0.24	0.23	0.22	0.22	0.21	0.21	0.24	0.26
4	0.43	0.32	0.28	0.26	0.25	0.24	0.23	0.23	0.23	0.26	0.29
5	0.65	0.45	0.37	0.32	0.29	0.26	0.24	0.23	0.21	0.22	0.36
6	0.65	0.45	0.37	0.33	0.31	0.29	0.27	0.26	0.24	0.26	0.38
7	0.75	0.55	0.46	0.41	0.36	0.32	0.29	0.26	0.24	0.25	0.44
8	0.80	0.62	0.53	0.48	0.43	0.38	0.35	0.31	0.28	0.28	0.49
11	0.99	0.82	0.68	0.58	0.50	0.41	0.35	0.30	0.27	0.28	0.60

TABLE 2-27 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES: MATHEMATICS SPANISH POOL (FROM SIMULATION STUDY)

Grade	Deciles										Overall
	1	2	3	4	5	6	7	8	9	10	
3	0.37	0.27	0.24	0.23	0.22	0.22	0.21	0.21	0.21	0.23	0.25
4	0.46	0.33	0.28	0.25	0.24	0.23	0.22	0.22	0.22	0.24	0.29
5	0.61	0.43	0.34	0.31	0.28	0.26	0.24	0.22	0.21	0.22	0.35
6	0.61	0.43	0.38	0.33	0.31	0.29	0.27	0.25	0.24	0.25	0.36
7	0.72	0.55	0.45	0.40	0.35	0.31	0.28	0.25	0.23	0.25	0.43
8	0.86	0.67	0.55	0.47	0.42	0.36	0.32	0.29	0.26	0.26	0.51
11	0.89	0.73	0.59	0.53	0.47	0.40	0.35	0.31	0.28	0.27	0.54

Item exposure

TABLE 2-28 shows the distribution of items across simulated test events. Exposure rates represent the number of test events in which items appeared. For example, in Grade 3 ELA/literacy, more than 97% of the items in the pool appeared in 0 to 20 percent of test events. Most items show a desired moderate exposure, and there are relatively few unused items. There are two items in Grade 5 ELA/literacy that were delivered to almost all students. In these cases, the pool contained only one item in a required element. This will be remedied in future tests as new items are added to the pools.

TABLE 2-28 PERCENT OF ITEMS BY EXPOSURE RATE

Grade	Total Items	Exposure Rate					
		Unused	0% 20%	21% 40%	41% 60%	61% 80%	81% 100%
English Language Arts/Literacy							
3	941	3.61	96.39	0.00	0.00	0.00	0.00
4	916	0.22	99.35	0.33	0.11	0.00	0.00
5	921	0.11	99.67	0.11	0.11	0.00	0.00
6	830	2.29	97.23	0.48	0.00	0.00	0.00
7	778	0.13	98.20	1.54	0.00	0.13	0.00
8	839	0.48	97.97	1.43	0.00	0.12	0.00
11	2539	0.35	99.65	0.00	0.00	0.00	0.00
Mathematics							
3	1272	0.39	99.45	0.16	0.00	0.00	0.00
4	1374	0.07	99.85	0.07	0.00	0.00	0.00
5	1295	0.15	99.69	0.15	0.00	0.00	0.00
6	1140	0.09	99.82	0.09	0.00	0.00	0.00
7	1030	0.00	99.90	0.10	0.00	0.00	0.00
8	901	0.00	99.00	1.00	0.00	0.00	0.00
11	2108	1.00	98.91	0.09	0.00	0.00	0.00

Internal Reliability Estimates

Reliability estimates reported in this section are derived from internal, IRT-based estimates of the measurement error in the test scores of examinees (*MSE*) and the observed variance of examinees test scores on the θ-scale $(var(\hat{\theta}))$. The formula for the reliability estimate is

$$\hat{\rho} = 1 - \frac{MSE}{var(\hat{\theta})}, \tag{8}$$

This estimate of reliability is similar to equation 5 in the simulation section except that estimates of measurement error are obtained from the parameter estimates of the items taken by the examinees. This is done by computing the test information for each examinee *i* as:

$$I(\hat{\theta}_i) = \sum_{j=1}^{I} D^2 a_j^2 \left(\frac{\sum_{l=1}^{m_j} l^2 Exp(\sum_{k=1}^{l} Da_j(\hat{\theta}_i - b_{jk}))}{1+\sum_{l=1}^{m_j} Exp(\sum_{k=1}^{l} Da_j(\hat{\theta}_i - b_{jk}))} - \left(\frac{\sum_{l=1}^{m_j} lExp(\sum_{k=1}^{l} Da_j(\hat{\theta}_i - b_{jk}))}{1+\sum_{l=1}^{m_j} Exp(\sum_{k=1}^{l} Da_j(\hat{\theta}_i - b_{jk}))} \right)^2 \right) \tag{9}$$

Where m_j is the maximum possible score point (starting from 0) for the *j*th item, and D is the scale factor, 1.7. Values of a_j and b_{jk} are item parameters for item *j* and score level *k*. The test information is computed using only the items answered by the examinee. The measurement error (SEM) for examinee *i* is then computed as:

$$SEM(\hat{\theta}_i) = \frac{1}{\sqrt{I(\hat{\theta}_i)}}, \tag{10}$$

The upper bound of $SEM(\hat{\theta}_i)$ is set to 2.5. Any value larger than 2.5 is truncated at 2.5. The mean squared error for a group of *N* examinees is then:

$$MSE = N^{-1} \sum_{i=1}^{N} SEM(\hat{\theta}_i)^2, \tag{11}$$

And the variance of the achievement scores is:

$$var(\hat{\theta}) = N^{-1} \sum_{i=1}^{N} \left(\hat{\theta}_i - \bar{\hat{\theta}}\right)^2, \tag{12}$$

where $\bar{\hat{\theta}}$ is the average of the $\hat{\theta}_i$.

The measurement error for a group of examinees is typically reported as the square root of MSE and is denoted RMSE. For the tables in this section, RMSE and $SEM(\hat{\theta}_i)$ are transformed to the reporting scale metric through multiplication by 'a', where '*a*' is the slope used to convert estimates of student achievement on the θ-scale to the reporting scale. The transformation equations for converting estimates of student achievement on the θ-scale to the reporting scale are given in chapter 5.

Tables 2-29 and 2-30 show the reliability of the observed total scores and subscores for, respectively, ELA/literacy and Mathematics. These internally-derived reliability coefficients are comparable to those derived from simulation in tables 2-5 and 2-6. Differences between simulation and empirical results may be due to the fact that actual student performance in 2016-2017 differed from the simulated distributions of student achievement based on the 2014-2015 test results (see table 2-1), and differences between simulations and practice in how extreme scores (HOT and LOT cases) were handled. Not all Smarter Balanced jurisdictions handle extreme scores the same way and as modeled through simulation.

TABLE 2-29 ELA/literacy summative scale marginal reliability estimates

Grade	N	Total score	Claim 1	Claim 2	Claim 3	Claim 4
3	696,653	0.928	0.765	0.716	0.588	0.662
4	699,548	0.923	0.719	0.729	0.632	0.671
5	709,776	0.931	0.721	0.736	0.626	0.721
6	693,625	0.920	0.719	0.725	0.555	0.685
7	685,850	0.924	0.762	0.721	0.592	0.672
8	680,646	0.924	0.747	0.701	0.538	0.676
11	596,954	0.925	0.745	0.733	0.580	0.675

TABLE 2-30 Mathematics summative scale score marginal reliability estimates

Grade	N	Total Score	Claim 1	Claim 2/4	Claim 3
3	867,284	0.946	0.896	0.640	0.687
4	701,496	0.946	0.892	0.674	0.693
5	712,179	0.936	0.890	0.565	0.661
6	693,862	0.937	0.885	0.659	0.696
7	686,593	0.925	0.885	0.610	0.635
8	672,412	0.928	0.884	0.664	0.688
11	575,088	0.910	0.892	0.579	0.570

Table 2-31 and Table 2-32 show that reliability varies by overall score levels. Score levels are represented in these tables as deciles. Students in the first decile are the lowest 10% of the student distribution by achievement score. All students take the same number of items, but the information delivered by the items depends on how difficult the items are for the student. Items that are too easy or too hard provide less information. Items may be classified into student deciles by their difficulty parameter. Information and reliability tends to be highest for deciles that contain the most items. Smarter Balanced pools are difficult relative to the population. Reliability therefore tends to be lower at low deciles (e.g. 1 and 2) and highest at the highest deciles (e.g., 9 and 10).

TABLE 2-31 MARGINAL RELIABILITY OVERALL AND BY DECILE FOR ELA/LITERACY

Grade	N	Var	Overall	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
3	696,653	8,233.1	**0.928**	0.868	0.917	0.930	0.936	0.939	0.942	0.942	0.941	0.938	0.926
4	699,548	9,072.7	**0.923**	0.873	0.916	0.926	0.931	0.932	0.935	0.935	0.934	0.933	0.919
5	709,776	9,857.0	**0.931**	0.881	0.930	0.938	0.942	0.942	0.942	0.940	0.938	0.935	0.920
6	693,625	9,204.4	**0.920**	0.858	0.913	0.925	0.930	0.931	0.934	0.934	0.930	0.929	0.915
7	685,850	10,197.5	**0.924**	0.865	0.918	0.929	0.935	0.936	0.939	0.938	0.936	0.933	0.915
8	680,646	10,133.0	**0.924**	0.877	0.920	0.928	0.933	0.934	0.934	0.934	0.933	0.930	0.918
11	596,954	12,670.7	**0.925**	0.873	0.915	0.928	0.934	0.938	0.938	0.938	0.936	0.931	0.917

TABLE 2-32 MARGINAL RELIABILITY OVERALL AND BY DECILE FOR MATHEMATICS

Grade	N	Var	Overall	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
3	867,284	6,796.1	**0.946**	0.893	0.937	0.946	0.952	0.955	0.957	0.958	0.959	0.958	0.945
4	701,496	7,215.0	**0.946**	0.883	0.936	0.947	0.952	0.956	0.960	0.960	0.961	0.961	0.951
5	712,179	8,684.5	**0.936**	0.839	0.905	0.925	0.938	0.948	0.955	0.961	0.964	0.966	0.959
6	693,862	11,396.4	**0.937**	0.815	0.912	0.936	0.947	0.954	0.958	0.962	0.965	0.966	0.957
7	686,593	12,639.0	**0.925**	0.758	0.887	0.918	0.935	0.946	0.954	0.961	0.966	0.969	0.961
8	672,412	14,563.0	**0.928**	0.812	0.894	0.916	0.928	0.938	0.947	0.956	0.964	0.969	0.963
11	575,088	16,128.5	**0.910**	0.717	0.851	0.890	0.915	0.932	0.945	0.954	0.962	0.968	0.967

*Deciles are based on Full Sample percentiles (Chapter 5)

Tables 2-33 to 2-36 show marginal reliability by demographic group. Because of the differences in average score across demographic groups and the relationship between reliability and student decile, demographic groups with lower average scores tend to have lower reliability than the population as a whole.

TABLE 2-33 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY ETHNIC GROUP-ELA/LITERACY

Grade	Group	N	Var	MSE	Marginal Reliability
3	**All**	**696,65**	**8,233**	**594**	**0.93**
	American Indian/Alaska Native	12,341	6,734	640	0.91
	Asian	56,189	7,795	565	0.93
	Black/African American	40,124	7,264	628	0.91
	Hispanic/Latino Ethnicity	308,13	7,101	617	0.91
	White	243,86	7,541	559	0.93
4	**All**	**699,54**	**9,073**	**695**	**0.92**
	American Indian/Alaska Native	12,501	7,477	741	0.90
	Asian	56,469	8,420	677	0.92
	Black/African American	40,277	8,172	725	0.91
	Hispanic/Latino Ethnicity	310,18	7,908	711	0.91
	White	245,39	8,091	667	0.92
5	**All**	**709,77**	**9,857**	**682**	**0.93**
	American Indian/Alaska Native	12,633	8,537	714	0.92
	Asian	57,769	9,289	687	0.93
	Black/African American	40,549	9,200	714	0.92
	Hispanic/Latino Ethnicity	317,60	8,519	691	0.92
	White	246,57	8,678	660	0.92
6	**All**	**693,62**	**9,204**	**738**	**0.92**
	American Indian/Alaska Native	12,492	8,030	726	0.91
	Asian	58,972	8,584	721	0.92
	Black/African American	39,627	8,471	781	0.91
	Hispanic/Latino Ethnicity	305,40	8,045	771	0.90
	White	243,33	7,885	685	0.91
7	**All**	**685,85**	**10,198**	**772**	**0.92**
	American Indian/Alaska Native	12,329	8,980	776	0.91
	Asian	59,899	8,907	760	0.92
	Black/African American	39,240	9,813	833	0.92
	Hispanic/Latino Ethnicity	296,33	9,045	799	0.91
	White	243,97	8,481	722	0.92
8	**All**	**680,64**	**10,133**	**771**	**0.92**
	American Indian/Alaska Native	12,173	9,059	799	0.91
	Asian	59,749	9,061	756	0.92
	Black/African American	39,439	9,579	818	0.92
	Hispanic/Latino Ethnicity	293,12	8,894	786	0.91
	White	243,97	8,835	742	0.92
11	**All**	**596,95**	**12,671**	**952**	**0.93**
	American Indian/Alaska Native	9,207	10,866	943	0.91
	Asian	58,182	11,230	931	0.92
	Black/African American	32,257	12,479	1,021	0.92
	Hispanic/Latino Ethnicity	261,96	11,643	974	0.92
	White	201,99	10,678	905	0.92

TABLE 2-34 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY ETHNIC GROUP – MATHEMATICS

Grade	Group	N	Var	MSE	Marginal Reliability
3	**All**	**867,284**	**6,796**	**368**	**0.95**
	American Indian/Alaska Native	14,334	5,848	424	0.93
	Asian	67,244	6,232	344	0.95
	Black/African American	43,176	6,363	421	0.93
	Hispanic/Latino Ethnicity	336,001	5,806	389	0.93
	White	319,693	6,142	351	0.94
4	**All**	**701,496**	**7,215**	**387**	**0.95**
	American Indian/Alaska Native	12,143	5,948	443	0.93
	Asian	55,067	6,682	348	0.95
	Black/African American	36,210	6,638	464	0.93
	Hispanic/Latino Ethnicity	295,549	5,995	415	0.93
	White	233,431	6,347	356	0.94
5	**All**	**712,179**	**8,685**	**558**	**0.94**
	American Indian/Alaska Native	12,262	7,153	656	0.91
	Asian	56,441	8,159	429	0.95
	Black/African American	36,646	7,515	714	0.91
	Hispanic/Latino Ethnicity	303,644	7,066	640	0.91
	White	235,099	7,765	476	0.94
6	**All**	**693,862**	**11,396**	**716**	**0.94**
	American Indian/Alaska Native	12,063	9,854	795	0.92
	Asian	57,262	10,239	549	0.95
	Black/African American	35,594	10,671	946	0.91
	Hispanic/Latino Ethnicity	291,508	9,848	840	0.92
	White	231,322	9,499	586	0.94
7	**All**	**686,593**	**12,639**	**943**	**0.93**
	American Indian/Alaska Native	11,932	10,181	986	0.90
	Asian	58,278	11,698	638	0.95
	Black/African American	35,640	10,876	1,279	0.88
	Hispanic/Latino Ethnicity	283,188	10,533	1,166	0.89
	White	232,690	10,534	708	0.93
8	**All**	**672,412**	**14,563**	**1,042**	**0.93**
	American Indian/Alaska Native	11,713	11,427	1,112	0.90
	Asian	57,518	14,141	731	0.95
	Black/African American	35,542	12,139	1,336	0.89
	Hispanic/Latino Ethnicity	279,506	11,996	1,227	0.90
	White	231,122	12,599	832	0.93
11	**All**	**575,088**	**16,129**	**1,448**	**0.91**
	American Indian/Alaska Native	8,635	12,222	1,422	0.88
	Asian	57,263	15,365	860	0.94
	Black/African American	30,394	12,741	2,008	0.84
	Hispanic/Latino Ethnicity	255,299	12,611	1,781	0.86
	White	192,733	14,276	1,089	0.92

TABLE 2-35 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY GROUP- ELA/LITERACY

Grade	Group	N	Var	MSE	Marginal Reliability
3	**All**	**696,653**	**8,233**	**594**	**0.93**
	LEP	163,483	5,707	656	0.89
	IDEA Indicator	6,543	7,562	586	0.92
	Section 504	403,471	7,075	616	0.91
	Economically Disadvantaged	77,990	7,554	726	0.90
4	**All**	**699,548**	**9,073**	**695**	**0.92**
	LEP	144,270	5,906	766	0.87
	IDEA Indicator	8,468	8,100	690	0.92
	Section 504	403,823	7,881	712	0.91
	Economically Disadvantaged	81,308	8,402	842	0.90
5	**All**	**709,776**	**9,857**	**682**	**0.93**
	LEP	128,532	5,971	759	0.87
	IDEA Indicator	10,010	8,671	667	0.92
	Section 504	409,114	8,561	692	0.92
	Economically Disadvantaged	83,798	8,625	849	0.90
6	**All**	**693,625**	**9,204**	**738**	**0.92**
	LEP	101,932	5,542	897	0.84
	IDEA Indicator	10,622	7,652	712	0.91
	Section 504	391,003	8,041	765	0.91
	Economically Disadvantaged	80,121	7,327	948	0.87
7	**All**	**685,850**	**10,198**	**772**	**0.92**
	LEP	85,499	6,081	954	0.84
	IDEA Indicator	11,791	8,439	738	0.91
	Section 504	377,147	9,117	797	0.91
	Economically Disadvantaged	75,594	7,946	989	0.88
8	**All**	**680,646**	**10,133**	**771**	**0.92**
	LEP	71,363	5,557	929	0.83
	IDEA Indicator	12,755	8,576	749	0.91
	Section 504	368,769	8,991	789	0.91
	Economically Disadvantaged	73,172	7,358	943	0.87
11	**All**	**596,954**	**12,671**	**952**	**0.93**
	LEP	53,086	7,327	1,213	0.83
	IDEA Indicator	13,424	11,128	920	0.92
	Section 504	312,898	11,894	976	0.92
	Economically Disadvantaged	54,647	9,638	1,176	0.88

TABLE 2-36 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY GROUP – MATHEMATICS

Grade	Group	N	Var	MSE	Marginal Reliability
3	**All**	**867,284**	**6,796**	**368**	**0.95**
	LEP	184,238	5,374	415	0.92
	IDEA Indicator	10,661	6,239	362	0.94
	Section 504	468,082	5,984	392	0.94
	Economically Disadvantaged	96,743	8,165	500	0.94
4	**All**	**701,496**	**7,215**	**387**	**0.95**
	LEP	139,765	5,181	466	0.91
	IDEA Indicator	8,546	6,619	375	0.94
	Section 504	381,398	6,163	419	0.93
	Economically Disadvantaged	77,082	7,866	570	0.93
5	**All**	**712,179**	**8,685**	**558**	**0.94**
	LEP	124,718	5,582	780	0.86
	IDEA Indicator	10,131	7,865	533	0.93
	Section 504	387,967	7,250	640	0.91
	Economically Disadvantaged	79,816	7,968	899	0.89
6	**All**	**693,862**	**11,396**	**716**	**0.94**
	LEP	98,916	8,227	1,134	0.86
	IDEA Indicator	10,733	9,585	656	0.93
	Section 504	371,605	10,028	834	0.92
	Economically Disadvantaged	75,408	10,559	1,319	0.88
7	**All**	**686,593**	**12,639**	**943**	**0.93**
	LEP	82,838	8,450	1,719	0.80
	IDEA Indicator	11,846	10,706	829	0.92
	Section 504	359,191	10,790	1,140	0.89
	Economically Disadvantaged	71,480	9,962	1,837	0.82
8	**All**	**672,412**	**14,563**	**1,042**	**0.93**
	LEP	68,388	9,328	1,746	0.81
	IDEA Indicator	12,803	12,258	967	0.92
	Section 504	350,316	12,288	1,211	0.90
	Economically Disadvantaged	68,900	10,088	1,784	0.82
11	**All**	**575,088**	**16,129**	**1,448**	**0.91**
	LEP	50,228	10,386	2,754	0.74
	IDEA Indicator	12,401	14,496	1,317	0.91
	Section 504	299,339	13,433	1,738	0.87
	Economically Disadvantaged	49,966	9,819	2,802	0.72

Paper/Pencil Test Reliability

Smarter Balanced supports fixed form paper/pencil tests for use in schools that lack computer capacity or to address potential religious concerns associated with using technology for assessments. Scores on the paper/pencil tests are on the same reporting scale as is used for the online assessments (CRESST, 2015b). The forms used in the 2016-2017 administration are collectively (for all grades) referred to as Form 1 and are the same forms used in the 2015-2016 administration. Tables 2-37 and 2-38 show, for ELA/literacy and mathematics respectively, statistical information pertaining to the items on Form 1 and to the measurement precision of these forms.

TABLE 2-37 RELIABILITY OF PAPER PENCIL TESTS, FORM 3 ENGLISH LANGUAGE ARTS/LITERACY

Gr	Full test					Claim 1		Claim 2		Claim 3		Claim 4	
	N items	Reliability	SEM	Avg. *b*	Avg. *a*	Reliability	SEM	Reliability	SEM	Reliability	SEM	Reliability	SEM
3	50	.905	.268	-.767	.668	.792	.397	.728	.454	.558	.578	.663	.505
4	50	.904	.289	-.225	.621	.797	.419	.694	.514	.598	.589	.628	.567
5	50	.924	.278	.147	.655	.805	.446	.808	.443	.608	.633	.698	.555
6	52	.922	.279	.240	.597	.805	.442	.796	.451	.581	.647	.707	.542
7	51	.918	.307	.932	.564	.799	.480	.791	.489	.602	.675	.661	.623
8	52	.903	.320	.903	.528	.775	.489	.751	.514	.482	.741	.634	.623
11	50	.907	.357	1.45	.489	.787	.540	.725	.613	.558	.778	.698	.643

TABLE 2-38 RELIABILITY OF PAPER PENCIL TEST, FORM 3 MATHEMATICS

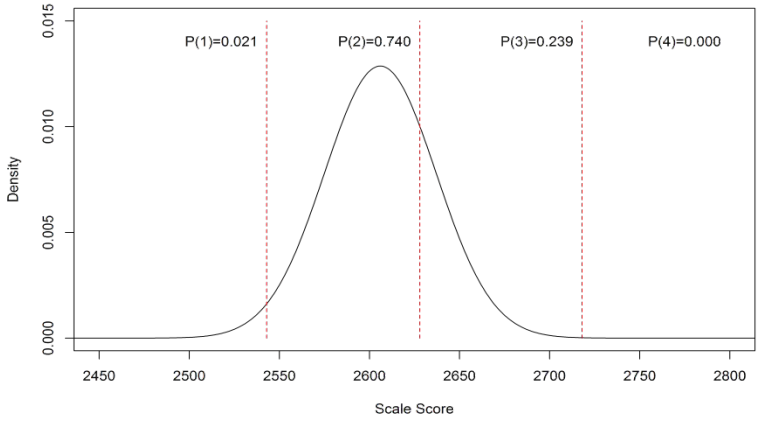
Gr	Full test					Claim 1		Claim 2/4		Claim 3	
	N items	Reliability	SEM	Avg. *b*	Avg. *a*	Reliability	SEM	Reliability	SEM	Reliability	SEM
3	41	.888	.234	-1.344	.894	.752	.349	.763	.341	.569	.459
4	40	.907	.263	-.557	.794	.838	.346	.690	.479	.692	.477
5	41	.901	.318	.200	.639	.811	.439	.770	.485	.618	.624
6	40	.896	.323	1.009	.746	.811	.434	.730	.520	.625	.613
7	41	.899	.382	1.392	.735	.828	.497	.743	.609	.629	.731
8	39	.872	.448	1.780	.552	.796	.564	.603	.788	.641	.749
11	42	.901	.515	2.285	.485	.835	.666	.722	.865	.636	.989

Classification Accuracy

Classification accuracy is a measure of how accurately test scores or sub-scores place students into reporting category levels. The likelihood of inaccurate placement depends on the amount of error associated with scores, especially those nearest cut points, and on the distribution of student achievement. For this report, classification accuracy was calculated in the following manner. For each examinee, analysts used the estimated scale score and its standard error of measurement to obtain a normal approximation of the likelihood function over the range of scale scores. The normal approximation took the scale score estimate as its mean and the standard error of measurement as its standard deviation. The proportion of the area under the curve within each level was then calculated.

illustrates the approach for one examinee in Grade 11 mathematics. In this example, the examinee's overall scale score is 2606 (placing this student in level 2, based on the cut scores for this grade level), with a standard error of measurement of 31 points. Accordingly, a normal distribution with mean of 2606 and standard deviation of 31 was used to approximate the likelihood of the examinee's true level, based on the observed test performance. The area under the curve was computed within each score range in order to estimate the probability that the examinee's true score falls within that level (the red vertical lines identify the cut scores). For the student in , the estimated probabilities were 2.1% for level 1, 74.0% for level 2, 23.9% for level 3, and 0.0% for level four. Since the student's assigned level was level 2, there is an estimated 74.0% chance the student was correctly classified and a 26% (2.1% + 23.9% + 0.0%) chance the student was misclassified.

FIGURE 2-1 ILLUSTRATIVE EXAMPLE OF A NORMAL DISTRIBUTION USED TO CALCULATE CLASSIFICATION ACCURACY



The same procedure was then applied to all students within the sample. Results are shown for 10 cases in the table below (Student 6 is the case illustrated in).

TABLE 2-39 ILLUSTRATIVE EXAMPLE CLASSIFICATION ACCURACY CALCULATION RESULTS

Student	SS	SEM	Level	probability that true score is in level			
				1	2	3	4
1	2751	23	4	0.000	0.000	0.076	0.924
2	2375	66	1	0.995	0.005	0.000	0.000
3	2482	42	1	0.927	0.073	0.000	0.000
4	2529	37	1	0.647	0.349	0.004	0.000
5	2524	36	1	0.701	0.297	0.002	0.000
6	2606	31	2	0.021	0.740	0.239	0.000
7	2474	42	1	0.950	0.050	0.000	0.000
8	2657	26	3	0.000	0.132	0.858	0.009
9	2600	31	2	0.033	0.784	0.183	0.000
10	2672	23	3	0.000	0.028	0.949	0.023
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Within the groups of students assigned to a particular level (Level 1, 2, 3 or 4 for the overall score; Below Standard, At/Near Standard, and Above Standard for the claim scores), we obtained the sums of the proportions over examinees. This gives us estimates of the number of students whose true score falls within a particular level, for each observed performance/achievement level. These sums can then be expressed as a proportion of the total sample.

TABLE 2-40 EXAMPLE OF CROSS-CLASSIFYING TRUE ACHIEVEMENT LEVEL BY OBSERVED ACHIEVEMENT LEVEL.

Observed Level	N	P	Frequency by True Level				Proportion by True Level			
			1	2	3	4	1	2	3	4
Overall										
Level 1	251,896	.451	225,454	26,172	263	8	.404	.047	.000	.000
Level 2	141,256	.253	21,800	100,364	19,080	11	.039	.180	.034	.000
Level 3	104,125	.186	161	14223	81089	8652	.000	.025	.145	.015
Level 4	61,276	.110	47	29	6452	54748	.000	.000	.012	.098
Claim 3										
Below Standard	167810	.300	143536	18323	4961	990	.257	.033	.009	.002
At/Near Standard	309550	.554	93364	102133	89696	24357	.167	.183	.161	.044
Above Standard	81193	.145	94	1214	18949	60936	.000	.002	.034	.109

Taking the table of expected proportions, two correct classification rates are then defined. First, a correct classification rate is provided for each true level (excluding the "At/Near Standard" classification for claims). This is the proportion of students whose true classification matches the observed level, among the subset of students with that observed level Table 2-41 shows the results of applying these calculations to the information in table 2-40.

TABLE 2-41 ILLUSTRATIVE EXAMPLE OF CORRECT CLASSIFICATION RATES

Observed Level	P	Proportion by True Level				Correct Classification	
		1	2	3	4	By level	Overall
Overall							
Level 1	.451	.404	.047	.000	.000	.404/.451=.895	(.404+.180+.145+.098)/1.000=.827
Level 2	.253	.039	.180	.034	.000	.180/.253=.711	
Level 3	.186	.000	.025	.145	.015	.145/.186=.779	
Level 4	.110	.000	.000	.012	.098	.098/.110=.893	
Claim 3							
Below Standard	.300	.257	.033	.009	.002	(.257+.033)/.300=.965	(.257+.033+.034+.109)/(.300+.145)=.971
At/Near Standard	.554	.167	.183	.161	.044	NA	
Above Standard	.145	.000	.002	.034	.109	(.034+.109)/.145=.984	

The overall correct classification rate is the sum of the proportions of students whose true score level matches the observed level. For the claim scores, the overall classification accuracy rate is based only on students whose observed achievement is 'Below standard' or 'Above standard'. The "Below standard" classification is correct when the true score falls within in levels 1 or 2; the "Above" classification is correct when the true score falls within in levels 3 or 4.

In the tables below, accuracy is highest for claim scores, since there are fewer categories to match and categorization is based on SEM, assuring that off-diagonal placements are rare. For overall scores, high and low categories have higher accuracy than middle categories since there is only one adjacent cell. In general, classification accuracy is moderate to high.

English Language Arts/Literacy Classification Accuracy

The classification accuracy for a level is the proportion of students whose true level matches their observed level divided by the proportion of students (P) at the observed level.

Overall classification accuracy is the sum of the proportions of "Below" students whose true level is 1 or 2, plus the proportions of 'Above" students whose true level is 3 or 4 divided by proportion of students in "Below" or "Above".

N is number of observed students in an achievement category. P is proportion of students in an observed category. The total number of students with valid overall scores (at observed levels) may be different from the total number of students with valid claim scores (assigned to Below, At/Near, or Above).

Tables 2-42 through 2-48 show ELA/Literacy classification accuracy for each grade 3 to 8 and 11.

TABLE 2-42 GRADE 3 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	212,123	0.305	0.273	0.031	0	0	0.897	0.809
Level 2	167,986	0.241	0.033	0.175	0.033	0	0.726	
Level 3	152,682	0.219	0	0.035	0.154	0.03	0.702	
Level 4	163,793	0.235	0	0	0.028	0.207	0.879	
Claim 1								
Below	242,314	0.369	0.295	0.067	0.005	0	0.985	0.984
At/Near	264,365	0.402	0.045	0.162	0.145	0.051		
Above	150,701	0.229	0	0.004	0.038	0.187	0.982	
Claim 2								
Below	182,524	0.278	0.231	0.041	0.005	0.001	0.981	0.981
At/Near	326,952	0.497	0.082	0.173	0.157	0.086		
Above	147,904	0.225	0	0.004	0.032	0.188	0.982	
Claim 3								
Below	152,157	0.231	0.195	0.029	0.006	0.002	0.968	0.97
At/Near	371,966	0.566	0.119	0.174	0.157	0.116		
Above	133,257	0.203	0.001	0.005	0.026	0.171	0.973	
Claim 4								
Below	177,720	0.27	0.234	0.029	0.006	0.002	0.973	0.978
At/Near	318,584	0.485	0.106	0.148	0.138	0.092		
Above	161,076	0.245	0	0.004	0.029	0.212	0.984	
All Students	696,584	1						

TABLE 2-43 GRADE 4 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	230,951	0.33	0.299	0.031	0	0	0.905	0.795
Level 2	141,316	0.202	0.034	0.132	0.035	0	0.655	
Level 3	160,116	0.229	0	0.038	0.156	0.034	0.682	
Level 4	167,063	0.239	0	0	0.031	0.208	0.871	
Claim 1								
Below	187,769	0.284	0.246	0.032	0.005	0.001	0.978	0.981
At/Near	314,803	0.477	0.095	0.154	0.16	0.068		
Above	157,627	0.239	0	0.004	0.038	0.196	0.983	
Claim 2								
Below	187,592	0.284	0.251	0.029	0.004	0	0.984	0.982
At/Near	327,281	0.496	0.104	0.146	0.152	0.093		
Above	145,326	0.22	0	0.004	0.029	0.186	0.979	
Claim 3								
Below	179,139	0.271	0.243	0.023	0.005	0.001	0.979	0.976
At/Near	359,754	0.545	0.142	0.142	0.146	0.115		
Above	121,306	0.184	0.001	0.004	0.024	0.155	0.972	
Claim 4								
Below	165,243	0.25	0.224	0.02	0.005	0.001	0.975	0.978
At/Near	334,727	0.507	0.133	0.131	0.142	0.101		
Above	160,229	0.243	0	0.004	0.03	0.208	0.981	
All Students	699,446	1						

TABLE 2-44 GRADE 5 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	216,003	0.304	0.276	0.029	0	0	0.906	0.809
Level 2	145,275	0.205	0.032	0.14	0.033	0	0.684	
Level 3	201,514	0.284	0	0.036	0.216	0.032	0.759	
Level 4	146,935	0.207	0	0	0.029	0.178	0.86	
Claim 1								
Below	204,947	0.305	0.258	0.04	0.006	0.001	0.978	0.979
At/Near	306,738	0.457	0.075	0.155	0.186	0.041		
Above	160,179	0.238	0	0.004	0.058	0.176	0.981	
Claim 2								
Below	172,390	0.257	0.223	0.029	0.004	0	0.984	0.983
At/Near	326,053	0.485	0.091	0.146	0.183	0.065		
Above	173,421	0.258	0	0.004	0.053	0.2	0.982	
Claim 3								
Below	157,336	0.234	0.207	0.022	0.005	0.001	0.976	0.976
At/Near	376,608	0.561	0.136	0.147	0.185	0.093		
Above	137,920	0.205	0.001	0.004	0.039	0.161	0.975	
Claim 4								
Below	192,463	0.286	0.255	0.026	0.005	0.001	0.979	0.981
At/Near	300,560	0.447	0.093	0.134	0.171	0.049		
Above	178,841	0.266	0	0.004	0.059	0.203	0.983	
All Students	709,727	1						

TABLE 2-45 GRADE 6 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	176,714	0.255	0.226	0.028	0	0	0.889	0.803
Level 2	182,510	0.263	0.034	0.192	0.038	0	0.73	
Level 3	219,652	0.317	0	0.04	0.245	0.032	0.774	
Level 4	114,627	0.165	0	0	0.026	0.139	0.841	
Claim 1								
Below	199,057	0.303	0.245	0.052	0.006	0	0.98	0.981
At/Near	316,967	0.483	0.067	0.183	0.195	0.039		
Above	139,876	0.213	0	0.004	0.059	0.15	0.982	
Claim 2								
Below	179,515	0.274	0.216	0.054	0.004	0	0.986	0.982
At/Near	332,534	0.507	0.056	0.183	0.206	0.062		
Above	143,851	0.219	0	0.004	0.057	0.157	0.978	
Claim 3								
Below	145,759	0.222	0.182	0.036	0.005	0.001	0.977	0.966
At/Near	388,563	0.592	0.097	0.184	0.208	0.103		
Above	121,578	0.185	0.002	0.006	0.041	0.135	0.953	
Claim 4								
Below	153,632	0.234	0.204	0.025	0.005	0.001	0.977	0.98
At/Near	322,225	0.491	0.098	0.152	0.188	0.053		
Above	180,043	0.274	0	0.004	0.072	0.198	0.983	
All Students	693,503	1						

TABLE 2-46 GRADE 7 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	175,282	0.256	0.229	0.026	0	0	0.898	0.811
Level 2	160,402	0.234	0.03	0.168	0.035	0	0.719	
Level 3	240,943	0.351	0	0.04	0.28	0.031	0.798	
Level 4	109,060	0.159	0	0	0.026	0.133	0.838	
Claim 1								
Below	193,589	0.299	0.245	0.048	0.005	0	0.982	0.983
At/Near	291,432	0.449	0.058	0.17	0.198	0.023		
Above	163,460	0.252	0	0.004	0.083	0.165	0.983	
Claim 2								
Below	156,786	0.242	0.2	0.037	0.004	0	0.984	0.982
At/Near	333,466	0.514	0.068	0.176	0.219	0.052		
Above	158,229	0.244	0	0.005	0.08	0.159	0.98	
Claim 3								
Below	157,926	0.244	0.206	0.032	0.005	0.001	0.978	0.972
At/Near	371,552	0.573	0.103	0.171	0.215	0.084		
Above	119,003	0.184	0.001	0.005	0.052	0.125	0.964	
Claim 4								
Below	150,324	0.232	0.203	0.024	0.005	0.001	0.976	0.979
At/Near	312,853	0.482	0.095	0.145	0.198	0.044		
Above	185,304	0.286	0	0.005	0.085	0.195	0.981	
All Students	685,687	1						

TABLE 2-47 GRADE 8 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	164,372	0.242	0.214	0.028	0	0	0.884	0.81
Level 2	175,295	0.258	0.032	0.191	0.034	0	0.741	
Level 3	235,896	0.347	0	0.038	0.277	0.031	0.8	
Level 4	104,922	0.154	0	0	0.026	0.128	0.832	
Claim 1								
Below	195,066	0.303	0.235	0.063	0.005	0	0.982	0.982
At/Near	285,440	0.444	0.041	0.177	0.205	0.021		
Above	163,013	0.253	0	0.005	0.092	0.156	0.982	
Claim 2								
Below	153,960	0.239	0.191	0.045	0.004	0	0.984	0.982
At/Near	340,656	0.529	0.07	0.185	0.218	0.057		
Above	148,903	0.231	0	0.004	0.067	0.16	0.98	
Claim 3								
Below	129,972	0.202	0.166	0.03	0.005	0.001	0.974	0.97
At/Near	395,173	0.614	0.115	0.183	0.222	0.094		
Above	118,374	0.184	0.001	0.005	0.047	0.13	0.966	
Claim 4								
Below	153,537	0.239	0.199	0.033	0.006	0.001	0.971	0.977
At/Near	312,209	0.485	0.085	0.163	0.2	0.037		
Above	177,773	0.276	0	0.005	0.086	0.185	0.982	
All Students	680,485	1						

TABLE 2-48 GRADE 11 ELA/LITERACY CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	107,269	0.18	0.158	0.022	0	0	0.88	0.812
Level 2	127,103	0.213	0.025	0.157	0.031	0	0.736	
Level 3	198,705	0.333	0	0.036	0.259	0.037	0.779	
Level 4	163,877	0.275	0	0	0.036	0.238	0.867	
Claim 1								
Below	121,148	0.204	0.155	0.044	0.004	0	0.978	0.982
At/Near	266,986	0.449	0.042	0.174	0.2	0.033		
Above	206,168	0.347	0	0.005	0.097	0.244	0.985	
Claim 2								
Below	120,236	0.202	0.162	0.036	0.004	0	0.982	0.983
At/Near	275,144	0.463	0.058	0.164	0.188	0.053		
Above	198,922	0.335	0	0.005	0.075	0.254	0.984	
Claim 3								
Below	96,872	0.163	0.134	0.024	0.004	0.001	0.972	0.972
At/Near	334,517	0.563	0.096	0.171	0.203	0.093		
Above	162,913	0.274	0.001	0.007	0.06	0.206	0.971	
Claim 4								
Below	104,819	0.176	0.146	0.025	0.005	0.001	0.968	0.977
At/Near	271,758	0.457	0.072	0.146	0.183	0.056		
Above	217,725	0.366	0	0.006	0.087	0.273	0.982	
All Students	596,954	1						

Mathematics Classification Accuracy

The classification accuracy for a level is the proportion of students whose true level matches their observed level divided by the proportion of students (P) at the observed level.

Overall classification accuracy is the sum of the proportions of "Below" students whose true level is 1 or 2, plus the proportions of 'Above" students whose true level is 3 or 4 divided by proportion of students in "Below" or "Above".

N is number of observed students in an achievement category. P is proportion of students in an observed category. The total number of students with valid overall scores (at observed levels) may be different from the total number of students with valid claim scores (assigned to Below, At/Near, or Above).

Tables 2-49 through 2-55 show classification accuracy for each grade 3 to 8 and 11.

TABLE 2-49 GRADE 3 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	222,909	0.257	0.23	0.027	0	0	0.896	0.829
Level 2	208,839	0.241	0.03	0.178	0.032	0	0.741	
Level 3	249,310	0.287	0	0.033	0.229	0.026	0.795	
Level 4	186,226	0.215	0	0	0.023	0.192	0.892	
Claim 1								
Below	262,285	0.316	0.241	0.072	0.003	0	0.991	0.99
At/Near	280,004	0.338	0.017	0.155	0.158	0.007		
Above	286,700	0.346	0	0.003	0.107	0.236	0.99	
Claim 2/4								
Below	216,265	0.261	0.212	0.039	0.007	0.002	0.963	0.974
At/Near	381,667	0.46	0.077	0.171	0.184	0.029		
Above	231,057	0.279	0	0.004	0.077	0.198	0.985	
Claim 3								
Below	189,286	0.228	0.191	0.029	0.006	0.001	0.967	0.978
At/Near	403,052	0.486	0.101	0.167	0.183	0.035		
Above	236,651	0.285	0	0.004	0.072	0.21	0.987	
All Students	867,284	1						

TABLE 2-50 GRADE 4 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	181,401	0.259	0.231	0.027	0	0	0.894	0.84
Level 2	219,683	0.313	0.03	0.253	0.031	0	0.808	
Level 3	175,851	0.251	0	0.03	0.198	0.022	0.79	
Level 4	124,561	0.178	0	0	0.02	0.158	0.89	
Claim 1								
Below	262,802	0.396	0.255	0.138	0.003	0	0.992	0.991
At/Near	210,983	0.318	0.004	0.159	0.148	0.007		
Above	189,234	0.285	0	0.003	0.091	0.191	0.99	
Claim 2/4								
Below	218,899	0.33	0.247	0.075	0.006	0.001	0.976	0.979
At/Near	300,509	0.453	0.042	0.202	0.174	0.036		
Above	143,611	0.217	0	0.003	0.053	0.16	0.983	
Claim 3								
Below	213,405	0.322	0.239	0.074	0.007	0.001	0.974	0.979
At/Near	291,636	0.44	0.046	0.198	0.167	0.029		
Above	157,978	0.238	0	0.003	0.061	0.173	0.985	
All Students	701,496	1						

TABLE 2-51 GRADE 5 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	257,764	0.362	0.326	0.036	0	0	0.901	0.836
Level 2	194,892	0.274	0.033	0.213	0.028	0	0.777	
Level 3	121,709	0.171	0	0.026	0.123	0.022	0.72	
Level 4	137,814	0.194	0	0	0.019	0.174	0.9	
Claim 1								
Below	309,397	0.458	0.341	0.114	0.004	0	0.992	0.991
At/Near	197,090	0.292	0.009	0.15	0.115	0.017		
Above	168,493	0.25	0	0.002	0.049	0.199	0.99	
Claim 2/4								
Below	256,338	0.38	0.297	0.068	0.01	0.005	0.962	0.97
At/Near	284,623	0.422	0.06	0.186	0.132	0.043		
Above	134,019	0.199	0	0.003	0.034	0.162	0.986	
Claim 3								
Below	250,624	0.371	0.303	0.059	0.008	0.002	0.975	0.977
At/Near	301,204	0.446	0.075	0.183	0.127	0.061		
Above	123,152	0.182	0	0.003	0.025	0.155	0.983	
All Students	712,179	1						

TABLE 2-52 GRADE 6 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	230,954	0.333	0.303	0.03	0	0	0.91	0.831
Level 2	199,903	0.288	0.032	0.223	0.033	0	0.775	
Level 3	135,704	0.196	0	0.03	0.141	0.024	0.722	
Level 4	127,301	0.183	0	0	0.02	0.163	0.89	
Claim 1								
Below	279,922	0.426	0.32	0.103	0.003	0	0.992	0.991
At/Near	214,131	0.326	0.008	0.157	0.138	0.023		
Above	162,980	0.248	0	0.003	0.053	0.192	0.988	
Claim 2/4								
Below	248,463	0.378	0.301	0.067	0.007	0.002	0.974	0.978
At/Near	283,562	0.432	0.049	0.194	0.145	0.043		
Above	125,008	0.19	0	0.003	0.035	0.153	0.985	
Claim 3								
Below	236,021	0.359	0.293	0.057	0.007	0.001	0.976	0.98
At/Near	287,166	0.437	0.069	0.181	0.134	0.053		
Above	133,846	0.204	0	0.003	0.032	0.169	0.985	
All Students	693,862	1						

TABLE 2-53 GRADE 7 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	232,743	0.339	0.304	0.035	0	0	0.896	0.83
Level 2	188,494	0.275	0.034	0.208	0.033	0	0.757	
Level 3	138,203	0.201	0	0.028	0.152	0.022	0.754	
Level 4	127,153	0.185	0	0	0.018	0.167	0.901	
Claim 1								
Below	277,943	0.427	0.321	0.102	0.003	0	0.992	0.991
At/Near	200,693	0.308	0.01	0.156	0.13	0.012		
Above	172,437	0.265	0	0.002	0.066	0.197	0.991	
Claim 2/4								
Below	222,258	0.341	0.277	0.052	0.009	0.003	0.963	0.973
At/Near	288,897	0.444	0.083	0.183	0.142	0.036		
Above	139,918	0.215	0	0.003	0.042	0.17	0.988	
Claim 3								
Below	178,769	0.275	0.231	0.035	0.007	0.002	0.968	0.975
At/Near	339,277	0.521	0.114	0.191	0.158	0.058		
Above	133,027	0.204	0	0.003	0.037	0.164	0.984	
All Students	686,593	1						

TABLE 2-54 GRADE 8 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	259,788	0.386	0.346	0.04	0	0	0.896	0.826
Level 2	162,946	0.242	0.037	0.175	0.031	0	0.72	
Level 3	115,289	0.171	0	0.026	0.123	0.022	0.719	
Level 4	134,389	0.2	0	0	0.018	0.182	0.908	
Claim 1								
Below	279,022	0.434	0.352	0.078	0.004	0	0.991	0.991
At/Near	199,874	0.311	0.023	0.149	0.12	0.02		
Above	164,308	0.255	0	0.002	0.046	0.207	0.992	
Claim 2/4								
Below	247,490	0.385	0.318	0.057	0.008	0.002	0.974	0.979
At/Near	249,647	0.388	0.073	0.155	0.124	0.036		
Above	146,067	0.227	0	0.003	0.04	0.185	0.988	
Claim 3								
Below	203,827	0.317	0.272	0.037	0.006	0.001	0.976	0.981
At/Near	302,317	0.47	0.109	0.167	0.132	0.062		
Above	137,060	0.213	0	0.003	0.031	0.18	0.988	
All Students	672,412	1						

TABLE 2-55 GRADE 11 MATHEMATICS CLASSIFICATION ACCURACY

Observed Level	N	P	True Level				Correct Classification	
			1	2	3	4	by level	overall
Overall								
Level 1	241,688	0.42	0.378	0.042	0	0	0.9	0.833
Level 2	140,011	0.243	0.036	0.175	0.032	0	0.719	
Level 3	117,031	0.204	0	0.026	0.16	0.017	0.786	
Level 4	76,358	0.133	0	0	0.013	0.119	0.899	
Claim 1								
Below	285,135	0.497	0.397	0.096	0.004	0	0.993	0.992
At/Near	158,659	0.277	0.013	0.14	0.12	0.004		
Above	129,818	0.226	0	0.002	0.081	0.143	0.992	
Claim 2/4								
Below	220,315	0.384	0.323	0.046	0.012	0.003	0.961	0.969
At/Near	254,985	0.445	0.106	0.164	0.146	0.029		
Above	98,312	0.171	0	0.002	0.044	0.125	0.987	
Claim 3								
Below	157,743	0.275	0.236	0.029	0.008	0.002	0.964	0.973
At/Near	315,952	0.551	0.166	0.176	0.161	0.047		
Above	99,917	0.174	0	0.002	0.039	0.133	0.987	
All Students	575,088	1						

Standard Errors of Measurement (SEMs)

Information in this section is based on actual test results from the 2016-2017 administration. Deciles in table 2-56 were defined by ranking students from highest to lowest achievement score and dividing the students into ten equal-sized groups according to rank. Decile 1 contains the 10% of students with the lowest achievement scores. Decile 10 contains the 10% of students with the highest achievement scores. The standard error of measurement (SEM) reported for a decile in table 2-56 is the average SEM among examinees at that decile.

The SEM is described conceptually and mathematically in the "Simulation Studies" (equation 4) and "Observed Reliability" (equation 11) sections of this chapter. SEMs in this section are computed using equation 11.

Table 2-56 is highly similar to table 2-7 with regard to SEM trends by grade and by decile within grade. The SEMs in table 2-7 are based on simulated examinees and are reported in terms of the θ-scale. The SEMs in table 2-56 are based on the performance of real examinees and are reported in terms of the reporting scale metric. Scores on student reports are obtained by transforming student achievement estimates from the θ-scale to the reporting scale using the transformation equations in Chapter 5. The multiplication factors of 85.8 and 79.3 for ELA/literacy and mathematics respectively in these transformation equations can be applied to the standard errors in table 2-7 to obtain values approximately equal to those in table 2-56.

Aside from differences due to multiplication factors, slight differences between results in tables 2-7 and 2-56 may be due to the fact that actual student performance in 2016-2017 differed from the simulated distributions of student achievement based on the 2014-2015 test results (see table 2-1), and differences between simulations and practice in how extreme scores (HOT and LOT cases) were handled. Not all Smarter Balanced jurisdictions handle extreme scores the same way.

Trends in table 2-56 are similar to the trends in table 2-7. At every grade, measurement error (SEM) is higher for lower deciles than for higher deciles. At deciles 1 and 2 and sometimes up to decile 4 for mathematics, the SEM is generally higher than the overall SEM. These trends reflect the fact that the item pools are difficult relative to student achievement, moreso in mathematics than in ELA and moreso at upper grades than at lower grades in mathematics.

TABLE 2-56 OVERALL SEM AND CONDITIONAL STANDARD ERRORS OF MEASUREMENT (CSEM) BY DECILE, ELA/LITERACY AND MATHEMATICS 2015-2016 ADMINISTRATION

Content Area	Grade	Overall SEM	1	2	3	4	5	6	7	8	9	10
ELA/Literacy	3	24.1	32.4	26.1	24	22.9	22.4	21.8	21.8	22	22.6	24.6
	4	26.1	33.3	27.5	25.9	25	24.9	24.2	24.2	24.4	24.6	27.1
	5	25.8	33.5	26.2	24.7	24	23.9	23.9	24.4	24.7	25.3	28
	6	26.9	35.4	28.2	26.3	25.3	25.1	24.6	24.7	25.4	25.6	27.9
	7	27.5	36.4	28.9	26.8	25.8	25.6	24.9	25.2	25.5	26.1	29.4
	8	27.6	34.9	28.5	27	26.1	25.9	25.8	25.8	26	26.7	28.8
	11	30.6	39.7	32.8	30.3	28.9	28	27.9	27.9	28.4	29.5	32.4
Mathematics	3	18.9	26.4	20.6	19.1	18.1	17.5	17.1	16.8	16.8	17	19.2
	4	19.2	28.3	21.5	19.6	18.5	17.8	17.1	16.9	16.8	16.9	18.7
	5	22.8	36.8	28.6	25.5	23.2	21.3	19.7	18.5	17.7	17.2	18.9
	6	25.5	44.8	31.6	27.1	24.5	22.9	21.7	20.7	20	19.7	21.9
	7	28.7	53.9	37.5	32.1	28.6	26.1	24	22.1	20.6	19.9	22
	8	30.9	51.4	39.1	35	32.3	30.1	27.7	25.2	23	21.4	23.2
	11	35.4	66.3	48.8	41.9	36.8	33	29.8	27.1	24.6	22.5	23

Table 2-57 and table 2-58 show the average SEM near the achievement level cut scores.

TABLE 2-57 CONDITIONAL STANDARD ERRORS OF MEASUREMENT NEAR (±10 POINTS) ACHIEVEMENT LEVEL CUT SCORES, GRADES 3-8 & 11 ELA/LITERACY 2015-2016 ADMINISTRATION

Grade	cut 1/2			cut 2/3			cut 3/4		
	N	M	SD	N	M	SD	N	M	SD
3	50570	23.33	1.31	56664	21.78	1.07	49358	22.15	1.15
4	47937	25.04	1.33	56718	24.38	1.21	50538	24.43	1.25
5	46704	24.08	0.95	54161	23.88	0.8	47427	24.75	0.85
6	42977	26.19	1.58	57165	24.63	1.25	42448	25.45	1.32
7	37892	26.77	0.98	54525	25.36	1.07	39890	25.93	0.99
8	40287	26.92	0.89	49989	25.87	0.71	39902	26.74	0.88
11	22880	31.81	1.15	37517	28.59	0.9	41521	28.08	0.84

TABLE 2-58 CONDITIONAL STANDARD ERRORS OF MEASUREMENT NEAR (±10 POINTS) OF ACHIEVEMENT LEVEL CUT SCORES, GRADES 3-8 & 11 MATHEMATICS 2015-2016 ADMINISTRATION

Grade	cut 1/2			cut 2/3			cut 3/4		
	N	M	SD	N	M	SD	N	M	SD
3	68994	19.04	0.84	86281	17.18	0.78	67003	16.78	0.76
4	54251	19.49	0.81	66217	17.02	0.75	45844	16.81	0.8
5	55893	23	1.01	54615	18.68	1.02	43761	17.53	0.97
6	45464	24.87	1.09	54955	21	0.89	40888	19.69	0.81
7	43210	28.94	1.69	47267	22.79	1.19	35476	19.93	1.01
8	45136	31.52	1.44	38513	25.71	1.03	31661	21.98	1.06
11	34344	33.98	2.35	32908	26.8	1.36	20302	22.21	1.09

Figure 2-2 to Figure 2-15 plot the SEM for the overall and claim scale scores for Grades 3 through 8 and 11 for ELA/literacy and mathematics. Together with Table 2-45 and Table 2-46, the figures show that the SEM tends to minimize around cut scores for levels 3 and 4.

FIGURE 2-2 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 3 ELA/LITERACY 2015-2016 ADMINISTRATION

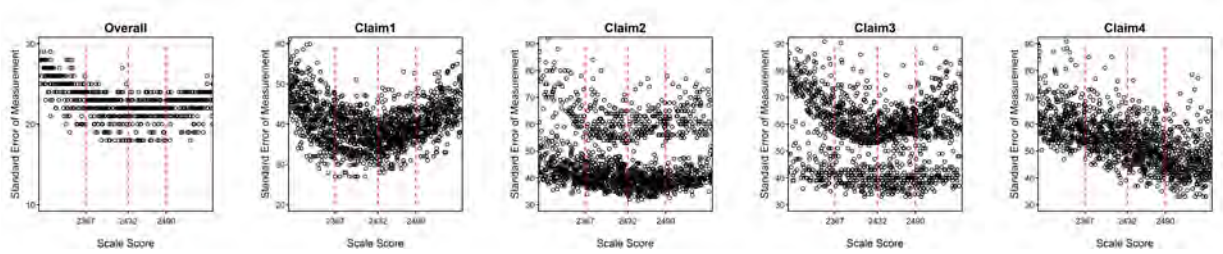


FIGURE 2-3 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 4 ELA/LITERACY 2015-2016 ADMINISTRATION

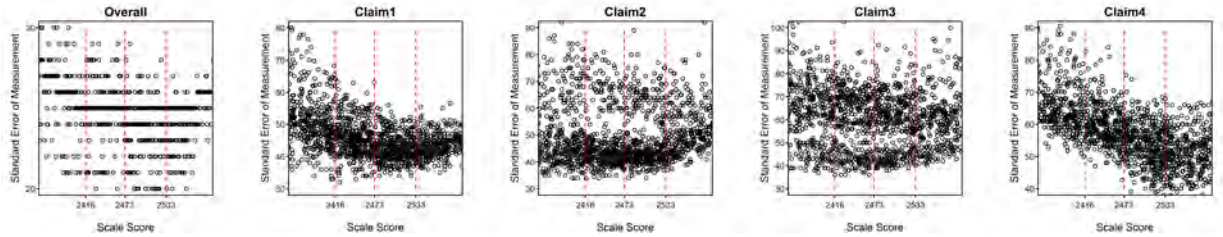


FIGURE 2-4 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 5 ELA/LITERACY 2015-2016 ADMINISTRATION

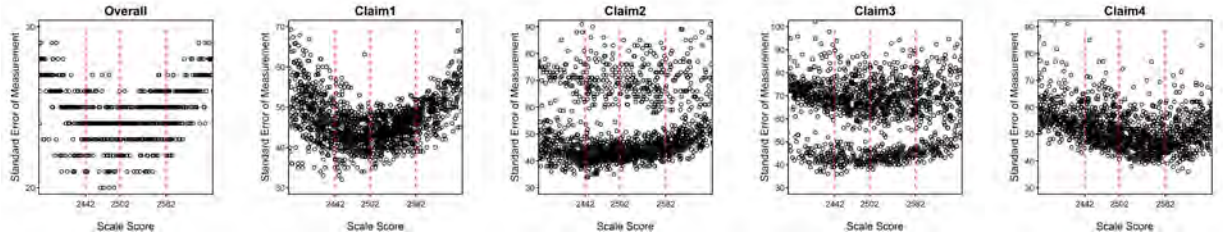


FIGURE 2-5 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 6 ELA/LITERACY 2015-2016 ADMINISTRATION

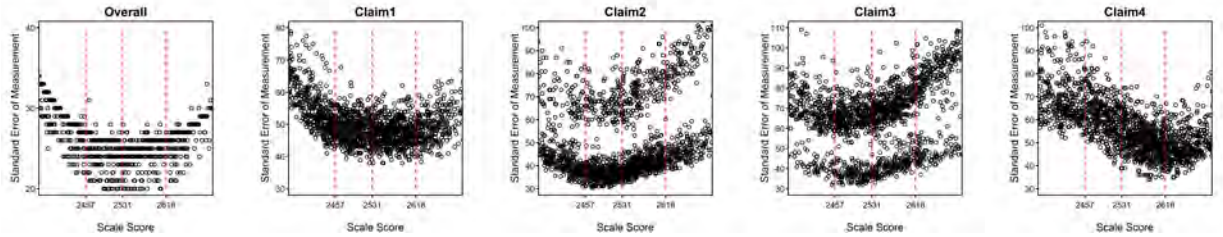


FIGURE 2-6 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 7 ELA/LITERACY 2015-2016 ADMINISTRATION

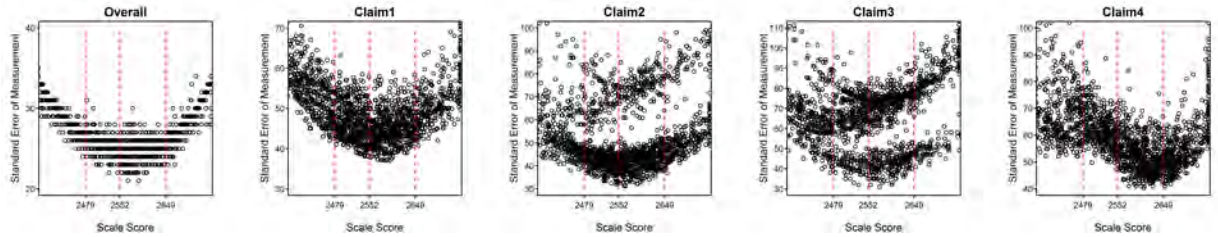


FIGURE 2-7 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 8 ELA/LITERACY 2015-2016 ADMINISTRATION

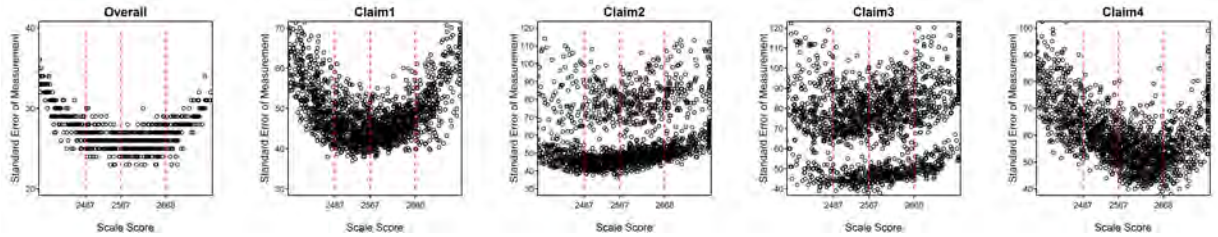


FIGURE 2-8 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 11 ELA/LITERACY 2015-2016 ADMINISTRATION

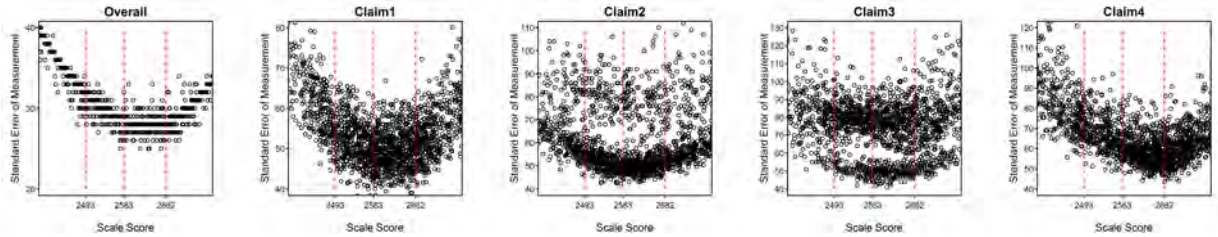


FIGURE 2-9 Conditional standard errors of measurement for overall and claim scale scores, grade 3 mathematics 2015-2016 (note: Claim2 refers to combined Claims 2 and 4)

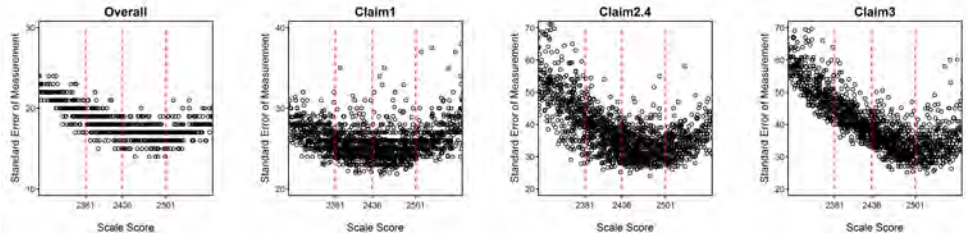


FIGURE 2-10 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 4 MATHEMATICS 2015-2016 ADMINISTRATION (NOTE: CLAIM2 REFERS TO COMBINED CLAIMS 2 AND 4)

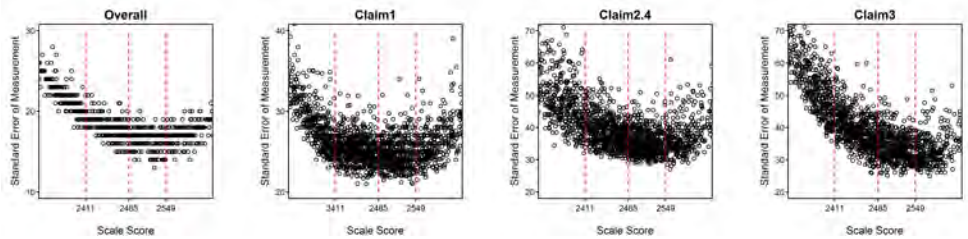


FIGURE 2-11 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 5 MATHEMATICS 2015-2016 ADMINISTRATION (NOTE: CLAIM2 REFERS TO COMBINED CLAIMS 2 AND 4)

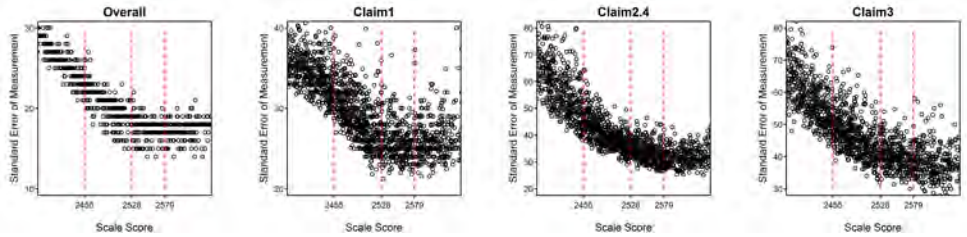


FIGURE 2-12 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 6 MATHEMATICS 2015-2016 ADMINISTRATION (NOTE: CLAIM2 REFERS TO COMBINED CLAIMS 2 AND 4)

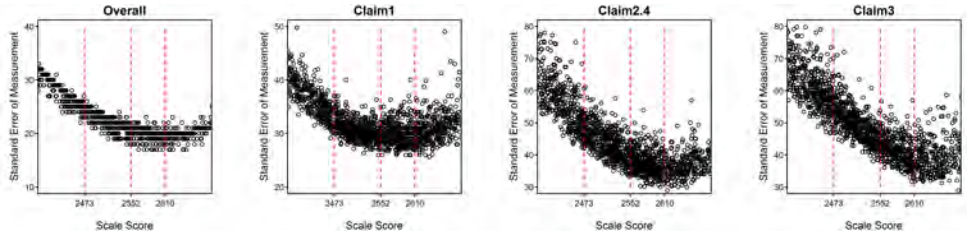


FIGURE 2-13 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 7 MATHEMATICS 2015-2016 ADMINISTRATION (NOTE: CLAIM2 REFERS TO COMBINED CLAIMS 2 AND 4)

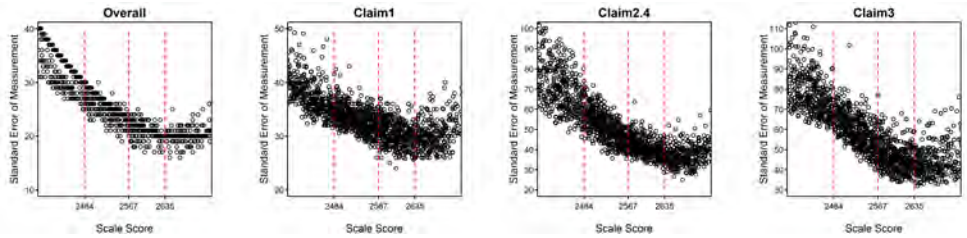


FIGURE 2-14 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 8 MATHEMATICS 2015-2016 ADMINISTRATION (NOTE: CLAIM2 REFERS TO COMBINED CLAIMS 2 AND 4)

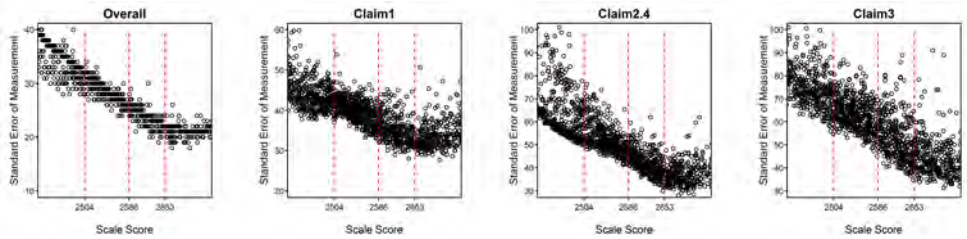
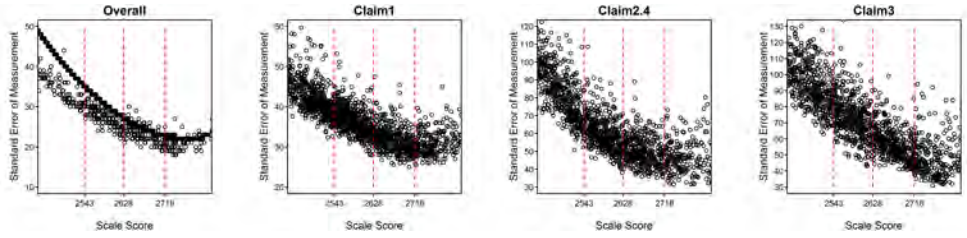


FIGURE 2-15 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 11 MATHEMATICS 2015-2016 ADMINISTRATION (NOTE: CLAIM2 REFERS TO COMBINED CLAIMS 2 AND 4)



References

American Institutes for Research (2014a). *Smarter Balanced scoring specification: 2014-2015 administration*. Los Angeles, CA: Smarter Balanced Assessment Consortium.

American Institutes for Research. (2014b). *Testing procedures for adaptive item-selection algorithm*. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

American Institutes for Research. (October, 2016). *Smarter Balanced summative assessments simulation results*. University of California Santa Cruz College of Extension: Smarter Balanced Assessment Consortium. Retrieved from https://portal.smarterbalanced.org/library/en/2016-17-summative-assessments-simulation-results.pdf

Cohen, J. & Albright, L. (2014). *Smarter Balanced adaptive item selection algorithm design report*. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterapp.org/specs/AdaptiveAlgorithm.html

Edelen, M. O., Stucky, B. D., & Chandra, A. (2015). Quantifying 'problematic' DIF within an IRT framework: Application to a cancer stigma index. *Quality of Life Research*, 24(1), 95-103.

National Center for Research on Evaluation, Standards, & Student Testing. (2015a*). Simulation-based evaluation of the smarter balanced summative assessments*. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

National Center for Research on Evaluation, Standards, & Student Testing. (2015b). *Initial report on the calibration of paper and pencil forms*. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

National Center for Research on Evaluation, Standards, & Student Testing. (February, 2017). S*imulation-based evaluation of the 2016-2017 Smarter Balanced summative assessments: general and accommodated item pools*. University of California Santa Cruz College of Extention: Smarter Balanced Assessment Consortium. Retrieved from https://portal.smarterbalanced.org/library/en/2016-2017-simulation-based-evaluation-summative-item-pools.pdf

Chapter 3: Test Fairness

Introduction

The Smarter Balanced Assessment Consortium (Smarter Balanced) has designed the assessment system to provide all eligible students with a fair assessment and equitable opportunity to participate in the Smarter Balanced Assessment. Ensuring test fairness is a fundamental part of validity, starting with test design, and is an important feature built into each step of the test development process, such as item writing, test administration, and scoring. The 2014 *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014, p. 49) state, "The term fairness has no single technical meaning, and is used in many ways in public discourse." It also suggests that fairness to all individuals in the intended population is an overriding and fundamental validity concern. As indicated in the *Standards for Educational and Psychological Testing* (2014, p. 63), "The central idea of fairness in testing is to identify and remove construct-irrelevant barriers to maximal performance for any examinee."

The Smarter Balanced system is designed to provide a valid, reliable, and fair measure of student achievement based on the Common Core State Standards[2] (CCSS). The validity and fairness of the measures of student achievement are influenced by a multitude of factors; central among them are:

- a clear definition of the construct—the knowledge, skills, and abilities—that are intended to be measured,
- the development of items and tasks that are explicitly designed to assess the construct that is the target of measurement,
- delivery of items and tasks that enable students to demonstrate their achievement of the construct
- capture and scoring of responses to those items and tasks.

Smarter Balanced uses several processes to address reliability, validity, and fairness. The fairness construct is defined in the CCSS which were developed during a state-led effort that was launched in 2009 by state leaders, including governors and state commissioners of education from 48 states, two territories and the District of Columbia, through their membership in the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO). The CCSS is a set of high-quality academic standards in mathematics and English language arts/literacy (ELA/literacy) that outline what a student should know and be able to do at the end of each grade. The standards were created to ensure that all students graduate from high school with the skills and knowledge necessary for post-secondary success. The CCSS have been adopted by all Consortium members. The *Smarter Balanced Content Specifications for the Summative Assessment of the CCSS for English Language Arts/Literacy* and the *Smarter Balanced Content Specifications for the Summative Assessment of the CCSS for Mathematics*, developed by Smarter Balanced (2015a; 2015b), define the knowledge, skills, and abilities to be assessed and their relationship to the CCSS. In doing so, these documents describe the major constructs—identified as "Claims"—within ELA/literacy and mathematics for which evidence of student achievement is gathered and which forms the basis for reporting student performance. Each claim is

[2] http://www.corestandards.org/

accompanied by a set of assessment targets that provide more detail about the range of content and Depth of Knowledge levels. The targets serve as the building blocks of test blueprints. Much of the evidence presented in this chapter pertains to fairness to students during the testing process and design elements and procedures that serve to minimize measurement bias (i.e., DIF). Fairness in item and test design processes and the design of accessibility supports (i.e., universal tools, designated supports and accommodations) in content development are also addressed.

Definitions for Validity, Bias, Sensitivity, and Fairness

Some key concepts for the ensuing discussion concern validity, bias, and fairness and are described as follows.

Validity

Validity is the extent to which the inferences and actions made based on test scores are appropriate and backed by evidence (Messick, 1989). It constitutes the central notion underlying the development, administration and scoring of a test, as well as the uses and interpretations of test scores. Validation is the process of accumulating evidence to support each proposed score interpretation or use. Evidence in support of validity is extensively discussed in Chapter 2.

Attention to bias, sensitivity and fairness in test development

According to the *Standards for Educational and Psychological Testing*, bias is "construct underrepresentation or construct-irrelevant components of tests scores that differentially affect the performance of different groups of test takers and consequently the reliability/precision and validity of interpretations and uses of their test scores." (AERA, APA, & NCME, 2014, p. 216). "Sensitivity" refers to an awareness of the need to avoid explicit bias in assessment. In common usage, reviews of tests for bias and sensitivity help ensure that test items and stimuli are fair for various groups of test takers, (AERA, APA, & NCME, 2014, p. 64).

The goal of fairness in assessment is to assure that test materials are as free as possible from unnecessary barriers to the success of diverse groups of students. Smarter Balanced developed the *Bias and Sensitivity Guidelines* (ETS, 2012) to help ensure that the assessments are fair for all groups of test takers, despite differences in characteristics including, but not limited to, disability status, ethnic group, gender, regional background, native language, race, religion, sexual orientation, and socioeconomic status. Unnecessary barriers can be reduced by following some fundamental rules:

- measuring only knowledge or skills that are relevant to the intended construct
- not angering, offending, upsetting, or otherwise distracting test takers, and
- treating all groups of people with appropriate respect in test materials.

These rules help ensure that the test content is fair for test takers as well as acceptable to the many stakeholders and constituent groups within Smarter Balanced member organizations. The more typical view is that bias and sensitivity guidelines apply primarily to the review of test items. However, fairness must be considered in all phases of test development and use. Smarter Balanced strongly relied on the *Bias and Sensitivity Guidelines* in the development of the Smarter Balanced assessments, particularly in item writing and review. Items must comply with the *Bias and Sensitivity*

Guidelines in order to be included in the Smarter Balanced assessments. Use of the *Guidelines* will help the Smarter Balanced assessments comply with Chapter 3, Standard 3.2 of the *Standards for Educational and Psychological Testing*. Standard 3.2 states that "Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics such as linguistic, communicative, cognitive, cultural, physical or other characteristics." (AERA, APA, & NCME, 2014, p. 64).

Smarter Balanced assessments were developed using the principles of evidence-centered design (ECD). ECD requires a chain of evidence-based reasoning that links test performance to the Claims made about test takers. Fair assessments are essential to the implementation of ECD. If test items are not fair, then the evidence they provide means different things for different groups of students. Under those circumstances, the Claims cannot be equally supported for all test takers, which is a threat to validity. As part of the validation process, all items are reviewed for bias and sensitivity using the *Bias and Sensitivity Guidelines* prior to being presented to students. This helps ensure that item responses reflect only knowledge of the intended content domain, are free of offensive or distracting material and portray all groups in a respectful manner. When the guidelines are followed, item responses provide evidence that supports assessment claims.

Smarter Balanced Item Development

Smarter Balanced has established item development practices that maximize access for all students, including English Language Learners (ELLs), students with disabilities, and ELs with disabilities, but not limited to those groups. Three resources—the Smarter Balanced *Item and Task Specifications* (2015c), the Smarter Balanced *General Accessibility Guidelines* (Measured Progress/ETS, 2012), and the Smarter Balanced *Bias and Sensitivity Guidelines* (ETS, 2012)—are used to guide the development of items and tasks to ensure that they accurately measure the targeted constructs. Recognizing the diverse characteristics and needs of students who participate in the Smarter Balanced assessments, the states worked together through the Smarter Balanced Test Administration and Student Access Work Group to incorporate research and practical lessons learned through Universal Design, accessibility tools, and accommodations (Thompson, Johnstone, & Thurlow, 2002).

A fundamental goal is to design an assessment that is accessible for all students, regardless of English language proficiency, disability, or other individual circumstances. The intent is to ensure that the following steps were achieved for Smarter Balanced.

- Design and develop items and tasks to ensure that all students have access to the items and tasks designed to measure the targeted constructs. In addition, deliver items, tasks, and the collection of student responses in a way that maximizes validity for each student.
- Adopt the conceptual model embodied in the Accessibility and Accommodations Framework that describes accessibility resources of digitally delivered items/tasks and acknowledges the need for some adult-monitored accommodations. The model also characterizes accessibility resources as a continuum ranging from those available to all students to ones that are implemented under adult supervision available only to those students with a documented need.

- Implement the use of an individualized and systematic needs profile for students, or Individual Student Assessment Accessibility Profile (ISAAP), that promotes the provision of appropriate access and tools for each student. Smarter created an ISAAP process that helps education teams systematically select the most appropriate accessibility resources for each student and the ISAAP tool, which helps teams note the accessibility resources chosen.

- Prior to any item development and item review, Smarter Balanced staff train item writers and reviewers on the Smarter Balanced General Accessibility Guidelines (Measured Progress/ETS, 2012) and Bias and Sensitivity Guidelines (ETS, 2012). As part of item review, individuals with expertise in accessibility, bias and sensitivity review each item and compare it against a checklist for (a) accessibility and, (b) bias and sensitivity. Items must pass each criterion on both checklists to be eligible for field testing. By relying on Universal Design to develop the items and requiring that individuals with expertise in bias, sensitivity and accessibility review the items throughout the iterative process of development, Smarter Balanced ensures that the items are appropriate for a wide range of students.

Guidelines for General Accessibility

In addition to implementing the principles of Universal Design during item development, Smarter Balanced meets the needs of English Learners (ELs) by addressing language aspects during development, as described in the *Smarter Balanced Guidelines for Accessibility for English Language Learners* (Young, Pitoniak, King, & Ayad, 2012). ELs have not yet acquired proficiency in English. The use of language that is not fully accessible can be regarded as a source of invalidity that affects the resulting test score interpretations by introducing construct-irrelevant variance. Although there are many validity issues related to the assessment of ELs, the main threat to validity when assessing content knowledge stems from language factors that are not relevant to the construct of interest. The goal of these EL guidelines was to minimize factors that are thought to contribute to such construct-irrelevant variance. Adherence to these guidelines helped ensure that, to the greatest extent possible, the Smarter Balanced assessments administered to ELs measure the intended targets. The EL *Guidelines* were intended primarily to inform Smarter Balanced assessment developers or other educational practitioners, including content specialists and testing coordinators.

For assessments, an important distinction is between content-related language that is the target of instruction versus language that is not content-related. For example, the use of words with specific technical meaning, such as “slope” when used in algebra or “population” when used in biology, should be used to assess content knowledge for all students. In contrast, greater caution should be exercised when including words that are not directly related to the domain. ELs may have had cultural and social experiences that differ from those of other students. Caution should be exercised in assuming that ELs have the same degree of familiarity with concepts or objects occurring in situational contexts. The recommendation was to use contexts or objects based on classroom or school experiences rather than ones that are based outside of school. For example, in constructing mathematics items, it is preferable to use common school objects, such as books and pencils, rather than objects in the home, such as kitchen appliances, to reduce the potential for construct-irrelevant variance associated with a test item. When the construct of interest includes a language component, the decisions regarding the proper use of language becomes more nuanced. If the construct assessed is the ability to explain a mathematical concept, then the decisions depend on how the construct is defined. If the construct includes the use of specific language skills, such as the ability

to explain a concept in an innovative context, then it is appropriate to assess these skills. In ELA/literacy, there is greater uncertainty as to item development approaches that faithfully reflect the construct while avoiding language inaccessible for ELs. The decisions of what best constitutes an item can rely on the content standards, definition of the construct, and the interpretation of the claims and assessment targets. For example, if interpreting the meanings in a literary text is the skill assessed, then using the original source materials is acceptable. However, the test item itself—as distinct from the passage or stimulus—should be written so that the task presented to a student is clearly defined using accessible language. Since ELs taking Smarter Balanced content assessments likely have a range of English proficiency skills, it is also important to consider the accessibility needs across the entire spectrum of proficiency. Since ELs by definition have not attained complete proficiency in English, the major consideration in developing items is ensuring that the language used is as accessible as possible. The use of accessible language does not guarantee that construct-irrelevant variance will be eliminated, but it is the best strategy for helping ensure valid scores for ELs and for other students as well.

Using clear and accessible language is a key strategy that minimizes construct-irrelevant variance in items. Language that is part of the construct being measured should not be simplified. For non-content-specific text, the language of presentation should be as clear and as simple as is practical. The following guidelines for the use of accessible language were proposed as guidance in the development of test items. This guidance was not intended to violate other principles of good item construction. From the ELL *Guidelines* (Young, Pitoniak, King, & Ayad, 2012, pp. 2-3), some general principles for the use of accessible language were proposed as follows.

- Design test directions to maximize clarity and minimize the potential for confusion.
- Use vocabulary widely accessible to all students, and avoid unfamiliar vocabulary not directly related to the construct (August, Carlo, & Snow, 2005; Bailey, Huang, Shin, Farnsworth, & Butler, 2007).
- Avoid the use of syntax or vocabulary that is above the test's target grade level (Borgioli, 2008). The test item should be written at a vocabulary level no higher than the target grade level, and preferably at a slightly lower grade level, to ensure that all students understand the task presented (Young, 2008).
- Keep sentence structures as simple as is possible while expressing the intended meaning. In general, ELs find a series of simpler, shorter sentences to be more accessible than longer, more complex sentences (Pitoniak, Young, Martiniello, King, Buteux, & Ginsburgh, 2009).
- Consider the impact of cognates (words with a common etymological origin) when developing items and false cognates. These are word pairs or phrases that appear to have the same meaning in two or more languages, but do not. Spanish and English share many cognates, and because the large majority of ELs speak Spanish as their first language (nationally, more than 75%), the presence of cognates can inadvertently confuse students and alter the skills being assessed by an item. Examples of false cognates include: billion (the correct Spanish word is millones; not billón, which means *trillion*); deception (engaño; not decepción, which means disappointment); large (grande; not largo, which means long); library (biblioteca; not librería, which means bookstore).

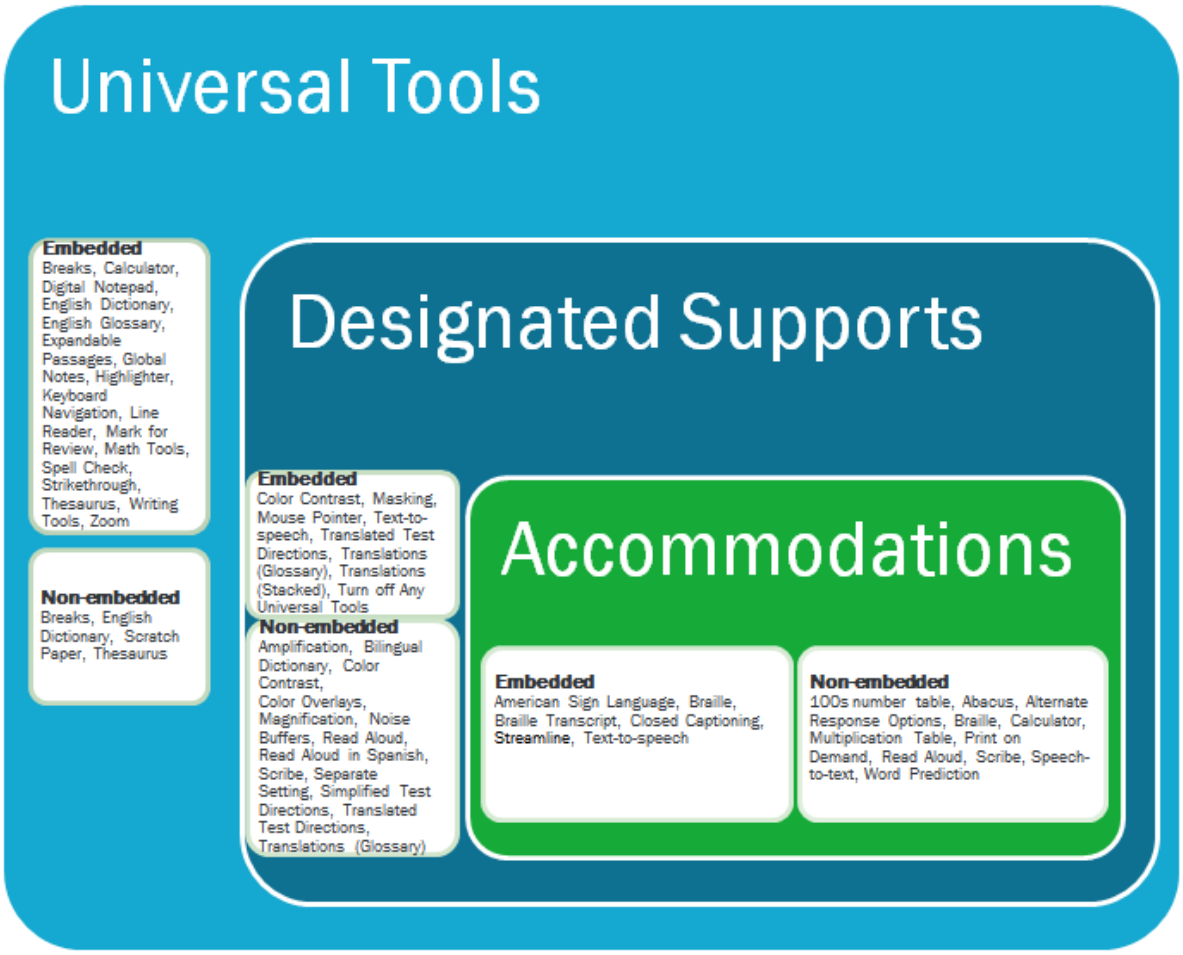
- Do not use cultural references or idiomatic expressions (such as "being on the ball") that are not equally familiar to all students (Bernhardt, 2005).
- Avoid sentence structures that may be confusing or difficult to follow, such as the use of passive voice or sentences with multiple clauses (Abedi & Lord, 2001; Forster & Olbrei, 1973; Schachter, 1983).
- Do not use syntax that may be confusing or ambiguous, such as using negation or double negatives in constructing test items (Abedi, 2006; Cummins, Kintsch, Reusser, & Weimer, 1988).
- Minimize the use of low-frequency, long, or morphologically complex words and long sentences (Abedi, 2006; Abedi, Lord & Plummer, 1995).
- Teachers can use multiple semiotic representations to convey meaning to students in their classrooms. Assessment developers should also consider ways to create questions using multi-semiotic methods so that students can better understand what is being asked (Kopriva, 2010). This might include greater use of graphical, schematic, or other visual representations to supplement information provided in written form.

Smarter Balanced Accessibility and Accommodations Framework

In addition to focusing on accessibility, bias, and sensitivity during item development, Smarter Balanced also maximizes accessibility through test delivery. Smarter Balanced works with members to maintain the original conceptual framework (Smarter Balanced, 2014a) that continues to serve as the basis underlying the usability, accessibility, and accommodations (Figure 3-1). This figure portrays several aspects of the Smarter Balanced assessment resources—universal tools (available for all students), designated supports (available when indicated by an adult or team), and accommodations as documented in an Individualized Education Program (IEP) or 504 plan. It also displays the additive and sequentially inclusive nature of these three aspects. Universal tools are available to all students, including those receiving designated supports and those receiving accommodations. Designated supports are available only to students who have been identified as needing these resources (as well as those students for whom the need is documented). Accommodations are available only to those students with documentation of the need through a formal plan (e.g., IEP, 504). Those students also may access designated supports and universal tools.

A universal tool or a designated support may also be an accommodation, depending on the content or grade. This approach is consistent with the emphasis that Smarter Balanced has placed on the validity of assessment results coupled with access. Universal tools, designated supports, and accommodations are all intended to yield valid scores. Use of universal tools, designated supports, and accommodations result in scores that count toward participation in statewide assessments. Also shown in Figure 0-1 are the universal tools, designated supports, and accommodations for each category of accessibility resources. There are both embedded and non-embedded versions of the universal tools, designated supports, or accommodations depending on whether they are provided as digitally delivered components of the test administration or provided locally separate from the test delivery system.

FIGURE 3-1 CONCEPTUAL MODEL UNDERLYING THE SMARTER BALANCED USABILITY, ACCESSIBILITY, AND ACCOMMODATIONS GUIDELINES.



Meeting the Needs of Traditionally Underrepresented Populations

Members decided to make accessibility resources available to all students based on need rather than eligibility status or student subgroup categorical designation. This reflects a belief among Consortium states that unnecessarily restricting access to accessibility resources threatens the validity of the assessment results and places students under undue stress and frustration. Additionally, accommodations are available for students who qualify for them. The Consortium utilizes a needs-based approach to providing accessibility resources. A description as to how this benefits ELs, students with disabilities, and ELs with disabilities is presented here.

How the Framework Meets Needs of Students Who Are ELs

Students who are ELs have needs that are unique from those students with disabilities, including language-related disabilities. The needs of ELs are not the result of a language-related disability, but instead are specific to the student's current level of English language proficiency. The needs of students who are ELs are diverse and are influenced by the interaction of several factors, including their current level of English language proficiency, their prior exposure to academic content and

language in their native language, the languages to which they are exposed outside of school, the length of time they have participated in the U.S. education system, and the language(s) in which academic content is presented in the classroom. Given the unique background and needs of each student, the conceptual framework is designed to focus on students as individuals and to provide several accessibility resources that can be combined in a variety of ways. Some of these digital tools, such as using a highlighter to highlight key information, are available to all students, including those at various stages of English language development. Other tools, such as the audio presentation of items or glossary definitions in English, may also be assigned to any student, including those at various stages of English language development. Still other tools, such as embedded glossaries that present translations of construct irrelevant terms, are intended for those students whose prior language experiences would allow them to benefit from translations into another spoken language. Collectively, the conceptual framework for usability, accessibility, and accommodations embraces a variety of accessibility resources that have been designed to meet the needs of students at various stages in their English language development.

How the Framework Meets Needs of Students with Disabilities

Federal law requires that students with disabilities who have a documented need receive accommodations that address those needs, and that they participate in assessments. The intent of the law is to ensure that all students have appropriate access to instructional materials and are held to the same high standards. When students are assessed, the law ensures that students receive appropriate accommodations during testing so they can appropriately demonstrate what they know and can do so that their achievement is measured accurately.

The *Accessibility and Accommodations Framework* addresses the needs of students with disabilities in three ways. First, it provides for the use of digital test items that are purposefully designed to contain multiple forms of the item, each developed to address a specific access need. By allowing the delivery of a given access form of an item to be tailored based on each student's access need, the Framework fulfills the intent of federal accommodation legislation. Embedding universal accessibility digital tools, however, addresses only a portion of the access needs required by many students with disabilities. Second, by embedding accessibility resources in the digital test delivery system, additional access needs are met. This approach fulfills the intent of the law for many, but not all, students with disabilities, by allowing the accessibility resources to be activated for students based on their needs. Third, by allowing for a wide variety of digital and locally provided accommodations (including physical arrangements), the Framework addresses a spectrum of accessibility resources appropriate for math and ELA/literacy assessment. Collectively, the Framework adheres to federal regulations by allowing a combination of universal design principles, universal tools, designated supports and accommodations to be embedded in a digital delivery system and through local administration assigned and provided based on individual student needs. Therefore, a student who is both an ELL and a student with a disability benefits from the system, because they may be eligible to have access to resources from any of the 3 categories as necessary to create an assessment tailored to their individual need.

The Individual Student Assessment Accessibility Profile (ISAAP)

Typical practice frequently required schools and educators to document, a priori, the need for specific student accommodations and then to document the use of those accommodations after the

assessment. For example, most programs require schools to document a student's need for a large-print version of a test for delivery to the school. Following the test administration, the school documented (often by bubbling in information on an answer sheet) which of the accommodations, if any, a given student received, whether the student actually used the large-print form, and whether any other accommodations, such as extended time, were provided. Traditionally, many programs have focused only on those students who have received accommodations and thus may consider an accommodation report as documenting accessibility needs. The documentation of need and use establishes a student's accessibility needs for assessment.

For most students, universal digital tools will be available by default in the Smarter Balanced test delivery system and need not be documented. These tools can be deactivated if they create an unnecessary distraction for the student. Other embedded accessibility resources that are available for any student needing them must be documented prior to assessment. To capture specific student accessibility needs, the Smarter Balanced Assessment System has established an individual student assessment accessibility profile (ISAAP). The ISAAP Tool is designed to facilitate selection of the universal tools, designated supports and accommodations that match student access needs for the Smarter Balanced assessments, as supported by the *Smarter Balanced Usability, Accessibility, and Accommodations Guidelines*. The ISAAP Tool[3] should be used in conjunction with the *Smarter Balanced Usability, Accessibility and Accommodations Guidelines* and state regulations and policies related to assessment accessibility as a part of the ISAAP process. For students requiring one or more accessibility resource, schools will be able to document this need prior to test administration. Furthermore, the ISAAP can include information about universal tools that may need to be eliminated for a given student. By documenting need prior to test administration, a digital delivery system will be able to activate the specified options when the student logs in to an assessment. In this way, the profile permits school-level personnel to focus on each individual student, documenting the accessibility resources required for valid assessment of that student in a way that is efficient to manage.

The conceptual framework shown in Figure 0-1 provides a structure that assists in identifying which accessibility resources should be made available for each student. In addition, the conceptual framework is designed to differentiate between universal tools available to all students and accessibility resources that must be assigned before the administration of the assessment. Consistent with recommendations from Shafer and Rivera (2011), Thurlow, Quenemoen, and Lazarus (2011), Fedorchak (2012), and Russell (2011), Smarter Balanced is encouraging school-level personnel to use a team approach to make decisions concerning each student's ISAAP. Gaining input from individuals with multiple perspectives, including the student, will likely result in appropriate decisions about the assignment of accessibility resources. Consistent with these recommendations one should avoid selecting too many accessibility resources for a student. The use of too many unneeded accessibility resources can decrease student performance.

The team approach encouraged by Smarter Balanced does not require the formation of a new decision-making team, and the structure of teams can vary widely depending on the background and needs of a student. A locally convened student support team can potentially create the ISAAP. For most students who do not require accessibility tools or accommodations, an initial decision by a

[3] http://isaap.smarterbalanced.org/

teacher may be confirmed by a second person (potentially the student). In contrast, for a student who is an English language learner and has been identified with one or more disabilities, the IEP team should include the English language development specialist who works with the student, along with other required IEP team members and the student, as appropriate. The composition of teams is not being defined by Smarter Balanced; it is under the control of each school and is subject to state and Federal requirements.

Usability, Accessibility, and Accommodations Guidelines

Smarter Balanced (2014b) has developed *Usability, Accessibility, and Accommodations Guidelines* (UAAG) that are intended for school-level personnel and decision-making teams, particularly Individualized Education Program (IEP) teams, as they prepare for and implement the Smarter Balanced assessment. The UAAG provide information for classroom teachers, English development educators, special education teachers, and related services personnel to use in selecting and administering universal tools, designated supports, and accommodations for those students who need them. The UAAG are also intended for assessment staff and administrators who oversee the decisions that are made in instruction and assessment. The Smarter Balanced UAAG emphasize an individualized approach to the implementation of assessment practices for those students who have diverse needs and participate in large-scale content assessments. This document focuses on universal tools, designated supports, and accommodations for the Smarter Balanced content assessments of ELA/literacy and mathematics. At the same time, it supports important instructional decisions about accessibility for students who participate in the Smarter Balanced assessments. It recognizes the critical connection between accessibility in instruction and accessibility during assessment. The UAAG are also incorporated into the Smarter Balanced Test Administration Manuals.

According to the UAAG (Smarter Balanced, 2014b, p. 2), all eligible students (including students with disabilities, ELs, and ELs with disabilities) should participate in the assessments. In addition, the performance of all students who take the assessment are measured with the same criteria. Specifically, all students enrolled in grades 3 to 8 and 11 are required to participate in the Smarter Balanced mathematics assessment except students with the most significant cognitive disabilities who meet the criteria for the mathematics alternate assessment based on alternate achievement standards (approximately 1% or less of the student population).

All students enrolled in grades 3 to 8 and 11 are required to participate in the Smarter Balanced English language/literacy assessment except:

- students with the most significant cognitive disabilities who meet the criteria for the English language/literacy alternate assessment based on alternate achievement standards (approximately 1% or fewer of the student population), and
- ELLs who are enrolled for the first year in a U.S. school. These students will participate in their state's English language proficiency assessment.

Federal laws governing student participation in statewide assessments include the Elementary and Secondary Education Act (ESEA)—reauthorized as the No Child Left Behind Act (NCLB) of 2001, the Individuals with Disabilities Education Improvement Act of 2004 (IDEA), and Section 504 of the Rehabilitation Act of 1973 (reauthorized in 2008).

Since the Smarter Balanced assessment is based on the CCSS, the universal tools, designated supports, and accommodations that are appropriate for the Smarter Balanced assessment may be different from those that state programs utilized previously. For the summative assessments, state participants can only make available to students the universal tools, designated supports, and accommodations consistent with the Smarter Balanced UAAG. According to the UAAG (Smarter Balanced, 2014b p. 1), when the implementation or use of the universal tool, designated support, or accommodation is in conflict with a member state's law, regulation, or policy, a state may elect not to make it available to students.

The Smarter Balanced universal tools, designated supports, and accommodations currently available for the Smarter Balanced assessments have been prescribed. The specific universal tools, designated supports, and accommodations approved by Smarter Balanced may undergo change if additional tools, supports, or accommodations are identified for the assessment based on state experience or research findings. The Consortium has established a standing committee, including members from Consortium and staff, that reviews suggested additional universal tools, designated supports, and accommodations to determine if changes are warranted. Proposed changes to the list of universal tools, designated supports, and accommodations are brought to consortium members for review, input, and vote for approval. Furthermore, states may issue temporary approvals (i.e., one summative assessment administration) for individual, unique student accommodations. It is expected that states will evaluate formal requests for unique accommodations and determine whether the request poses a threat to the measurement of the construct. Upon issuing temporary approval, the petitioning state can send documentation of the approval to the Consortium. The Consortium will consider all state-approved temporary accommodations as part of the annual Consortium accommodations review process. The Consortium will provide to member states a list of the temporary accommodations issued by states that are not Consortium-approved accommodations.

Provision of Specialized Tests or Pools

Smarter Balanced provides a full item pool and a series of specialized items pools that allow students who are eligible for them to access the tests with a minimum of barriers. These accessibility resources are considered embedded accommodations or embedded designated supports. The specialized pools that were available in 2015-16 are shown in Table 3-1.

TABLE 3-1 SPECIALIZED TESTS AVAILABLE TO QUALIFYING STUDENTS IN 2015-16

Subject	Test instrument
ELA/literacy	ASL adaptive online (Listening only)
ELA/literacy	Closed Captioning adaptive online (Listening only)
ELA/literacy	Braille adaptive online
ELA/literacy	Braille paper pencil
Math	Translated glossaries adaptive online
Math	Stacked Spanish adaptive online
Math	ASL adaptive online
Math	Braille adaptive online
Math	Braille fixed form online
Math	Spanish paper pencil
Math	Braille paper pencil
Math	Translated glossaries paper pencil

The technical quality of these tests is reported in Chapter 2. Online fixed forms and paper/pencil forms use the same item pools and share their psychometric properties. Given the small populations, the measurement properties for the adaptive tests in American Sign Language, Braille and Spanish is primarily gained through simulations.

Differential Item Functioning (DIF)

As part of the validity evidence from internal structure, differential item functioning (DIF) analyses were conducted on items using data from field testing. This section presents the evidence to support the frameworks' claims. DIF analyses are used to identify those items for which identifiable groups of students (e.g., males, females) with the same underlying level of ability have different probabilities of answering an item correctly or obtaining a given score level. Students data are aggregated according to the reported subgroups (e.g. ethnicity, gender, English Language Proficiency, etc.) for DIF analyses. Students in each subgroup are then ranked relative to their total test score (conditioning on ability). Students in the focal group (e.g., females) are then compared to students in the reference group (e.g., males) relative to their performance on individual items. It is part of the Smarter Balanced framework to have ongoing study and review of findings to inform iterative, data-driven decisions.

If items are more difficult for some groups of students than for other groups of students, the items may not necessarily be unfair. For example, if an item were intended to measure the ability to

comprehend a reading passage in English, score differences between groups based on real differences in comprehension of English would be valid and, therefore, fair. As Cole and Zieky (2001, p. 375) noted, "If the members of the measurement community currently agree on any aspect of fairness, it is that score differences alone are not proof of bias." Fairness does not require that all groups have the same average item score. Evaluations of validity include examination of differences in responses for groups of students matched on overall ability. An item would be unfair if the source of the difficulty were not a valid aspect of the item. For example, an item would be unfair if members of a group of test takers were distracted by an aspect of the item that they found highly offensive. If the difference in difficulty reflected real and relevant differences in the group's level of mastery of the tested CCSS, the item could be considered fair.

Method of Assessing DIF

Differential item functioning (DIF) analyses are performed on items using data gathered in the field test stage. DIF analyses are used to identify those items for which different groups of students (e.g., males versus females) with the same underlying level of ability have different probabilities of answering an item correctly. To perform a DIF analysis, student data are aggregated according to the reported subgroups (e.g., ethnicity, gender, etc.). Students in each subgroup are then matched on their total test score (conditioning on ability). The focal group's performance on the item is compared to the performance of reference group, conditional on total score. The definitions for the focal and references groups used are given in Table 0-2. A DIF analysis asks, "If we compare focal-group and reference-group students of the same overall ability (as indicated by their performance on the full test), are any test items appreciably more difficult for one group compared with another group?" As explained above, DIF in this context is viewed as only an indicator of possible bias.

Items flagged for DIF are subsequently reviewed by content experts and bias/sensitivity committees to determine the source and meaning of performance differences. An item flagged for DIF may be measuring something different from the intended construct. However, it is important to recognize that DIF-flagged items might be related to actual differences in relevant knowledge and skills, or may have been flagged due to chance variation in the DIF statistic (known as statistical Type I error). Final decisions about the resolution of item DIF are made by the multi-disciplinary panel of content experts.

TABLE 3-2 DEFINITION OF FOCAL AND REFERENCE GROUPS

Group Type	Focal Groups	Reference Groups
Gender	Female	Male
Ethnicity	African American	White
	Asian/Pacific Islander	
	Native American/Alaska Native	
	Hispanic	
Special Populations	Limited English Proficient (LEP)	English Proficient
	Individualized Education Program (IEP)	No IEP
	Title 1	Not Title 1

TABLE 3-3 DIF FLAGGING LOGIC FOR SELECTED-RESPONSE ITEMS

DIF Category	Definition
A (negligible)	MH D-DIF is less than 1 in absolute value or MH Chi-square is not significantly different from 0 at p<.05.
B (slight to moderate)	Items that do not meet the definitions of A or C. Positive values of MH D-DIF are classified as "B+" and negative values as "B-".
C (moderate to large)	MH D-DIF is at least 1.5 in absolute value and MH Chi-square is significantly different from 0 at p<.05. Positive values of MH D-DIF are classified as "C+" and negative values as "C-"

TABLE 3-4 DIF FLAGGING LOGIC FOR CONSTRUCTED RESPONSE ITEMS

DIF Category	Definition
A (negligible)	Mantel p-value >0.05 or chi-square |SMD/SD| ≤ 0.17
B (slight to moderate)	Mantel chi-square p-value <0.05 and |SMD/SD| >0.17, but ≤0.25
C (moderate to large)	Mantel chi-square p-value <0.05 and |SMD/SD| > 0.25

Items are classified into three DIF categories of "A," "B," or "C." DIF. Category A items contain negligible DIF, category B items exhibit moderate DIF, and category C items have significant values of DIF. Positive values favor the focus group, and negative values are in favor of the reference group. The positive and negative values are reported for C-DIF item flagging. DIF analyses are not conducted if the sample size for either the reference group or the focal group is less than 100 or if the sample size for the two combined groups is less than 400. In subsequent tables, category A levels of DIF are not flagged as they are too small to have perceptible interpretation.

Different DIF analysis procedures are used for dichotomous items (items with 0/1 score categories; selected-response items) and polytomous items (items with more than two score categories; constructed-response items). Statistics from two DIF detection methods are computed. The Mantel-Haenszel procedure (Mantel & Haenszel, 1959) is used for dichotomously-scored (0/1) items. The standardized mean difference (SMD) procedure (Dorans & Kulick, 1983, 1986) is used for items that have more than two score levels (polytomously-scored items). Both types of items are classified into DIF categories of A, B, and C, as described in Table 3-3.

For dichotomous items, the statistic described by Holland and Thayer (1988), known as Mantel-Haenszel D-DIF (MH D-DIF), is reported. This statistic is reported on the delta scale, which is a normalized transformation of item difficulty (p-value) with a mean of 13 and a standard deviation of 4. Items that are not significantly different based on the MH D-DIF ($p > 0.05$) are considered to have similar performance between the two studied groups; these items are considered to be functioning appropriately. For items where the statistical test indicates significant differences ($p < 0.05$), the effect size is used to determine the direction and severity of the DIF. The formula for the estimate of constant odds ratio is

$$\alpha_{MH} = \frac{\left(\sum_m \frac{R_{rm} W_{fm}}{N_m}\right)}{\left(\sum_m \frac{R_{fm} W_{rm}}{N_m}\right)}, \quad (1)$$

where

R_{rm} = number in reference group at ability level m answering the item right;

W_{fm} = number in focal group at ability level m answering the item wrong;

R_{fm} = number in focal group at ability level m answering the item right;

W_{rm} = number in reference group at ability level m answering the item wrong; and

N_m = total group at ability level m.

This value can then be used as follows (Holland & Thayer, 1988):

$$MH\ D\text{-}DIF = -2.35 \ln[\alpha_{MH}], \quad (2)$$

The Mantel-Haenszel chi-square statistic used to classify items into the three DIF categories is

$$MH\ CHISQ = \frac{(\left|\sum_m R_{rm} - \sum_m E(R_{rm})\right| - \frac{1}{2})^2}{\sum_m Var(R_{rm})}, \quad (3)$$

where

$$E(R_{rm}) = N_{rm}R_{Nm} / N_m; Var(R_{rm}) = \frac{N_{rm}N_{fm}R_{Nm}W_{Nm}}{N_m^2(N_m - 1)}, \quad (4)$$

N_{rm} and N_{fm} are the numbers of examinees in the reference and focal groups, respectively, R_{Nm} and W_{Nm} are the number of examinees who answered the item correctly and incorrectly, respectively. The classification logic used for flagging items is based on a combination of absolute differences and significance testing. Items that are not statistically different based on the MH D-DIF ($p > 0.05$) are considered to have similar performance between the two studied groups; these items are considered to be functioning appropriately. For items where the statistical test indicates significant differences ($p < 0.05$), the effect size is used to determine the direction and severity of the DIF, which translates into the three categories of DIF and the "+" and "-" signs attached to these letters.

The standardized mean difference compares item performance of two subpopulations adjusting for differences in the distributions of the two subpopulations. The standardized mean difference statistic can be divided by the total standard deviation to obtain a measure of the effect size. A negative value of the standardized mean difference shows that the item is more difficult for the focal group, whereas a positive value indicates that it is more difficult for the reference group. The standardized mean difference used for polytomous items is defined as:

$$SMD = \sum p_{FK}m_{FK} - \sum p_{FK}m_{RK}, \quad (5)$$

where p_{Fk} is the proportion of the focal group members who are at the kth level of the matching variable, m_{Fk} is the mean score for the focal group at the kth level, and m_{Rk} is the mean item score for the reference group at the kthlevel. The standardized mean difference is divided by the total item group standard deviation to get a measure of the effect size. The classification logic for polytomous items is based on a combination of absolute differences and significance testing, as shown in Table 0-4. Items that are not statistically different are considered to have similar performance between the two studied groups; these items are considered to be functioning appropriately.

DIF Results for the Summative Pools

Table 3-5 and Table 3-6 show DIF analysis results for items in the 2016-2017 ELA/Literacy and mathematics summative item pools. The numbers of items with moderate or significant levels of DIF (B or C DIF) in the summative pools were relatively small. All of these items had previously undergone bias reviews. After the DIF analyses, content editors inspected B and C DIF items before including them in operational tests administrations. This inspection is to assure that statistical differences are not caused by any content issues of bias or sensitivity. Only items approved by a multi-disciplinary panel of experts are eligible to move into operational pools.

TABLE 3-1 NUMBER OF DIF ITEMS IN 16/17 SUMMATIVE POOL FLAGGED BY CATEGORY (ELA/LITERACY, GRADES 3-8 AND 11)

Grade	DIF Category	Focal group/Referent Group							
		Female/ Male	Asian/ White	Black/ White	Hispanic/ White	NativeAmeri-can/White	IEP/Non-IEP	LEP/Non-LEP	Title1/Non-Title1
3	N/A	0	177	79	2	655	81	31	0
3	A	917	699	811	912	272	829	879	933
3	B-	6	28	35	20	11	23	26	8
3	B+	14	32	16	7	3	8	5	0
3	C-	0	1	0	0	0	0	0	0
3	C+	4	4	0	0	0	0	0	0
4	N/A	0	139	97	5	605	78	22	0
4	A	872	709	771	875	295	794	849	904
4	B-	15	19	33	33	11	34	39	11
4	B+	22	40	12	3	5	10	5	1
4	C-	1	3	1	0	0	0	1	0
4	C+	6	6	2	0	0	0	0	0
5	N/A	0	150	93	1	664	74	40	0
5	A	850	692	783	873	243	805	821	905
5	B-	22	37	25	32	11	32	46	12
5	B+	32	36	15	11	2	6	9	3
5	C-	2	0	1	2	0	3	2	0
5	C+	14	5	3	1	0	0	2	0
6	N/A	0	153	53	0	543	46	64	0
6	A	786	617	736	792	275	756	699	819
6	B-	20	31	31	25	5	24	60	7
6	B+	18	19	9	8	6	2	4	3
6	C-	2	5	0	4	0	1	2	0
6	C+	3	4	0	0	0	0	0	0
7	N/A	0	159	95	4	458	72	117	0
7	A	711	573	652	748	304	684	621	771
7	B-	27	19	22	22	9	19	35	6
7	B+	33	19	7	4	7	3	5	1
7	C-	2	1	1	0	0	0	0	0
7	C+	5	7	1	0	0	0	0	0
8	N/A	0	148	134	0	587	109	138	0
8	A	752	615	665	797	245	690	632	830
8	B-	31	37	26	25	4	35	57	8
8	B+	40	27	10	12	2	3	11	0
8	C-	4	2	2	1	0	1	0	0
8	C+	11	9	1	3	0	0	0	0
11	N/A	81	749	924	88	2389	1143	1421	81
11	A	2298	1634	1552	2290	133	1341	1041	2377
11	B-	103	77	27	127	1	24	49	65
11	B+	42	61	23	23	6	19	18	7
11	C-	6	3	2	1	1	1	0	0
11	C+	0	6	2	1	0	2	1	0

TABLE 3-2 NUMBER OF DIF ITEMS IN 16/17 SUMMATIVE POOL FLAGGED BY CATEGORY (MATHEMATICS, GRADES 3-8 AND 11)

Grade	DIF Category	Focal group/Referent Group							
		Female/ Male	Asian/ White	Black/ White	Hispanic/ White	NativeAmeri-can/White	IEP/Non-IEP	LEP/Non-LEP	Title1/Non-Title1
3	N/A	0	137	19	0	1197	2	0	0
3	A	1242	1006	1145	1189	74	1221	1222	1259
3	B-	13	30	45	38	0	33	21	10
3	B+	16	77	58	43	1	13	24	3
3	C-	0	4	0	0	0	1	1	0
3	C+	1	18	5	2	0	2	4	0
4	N/A	0	140	104	0	1186	16	1	0
4	A	1328	1095	1187	1314	177	1301	1302	1353
4	B-	21	27	38	29	2	51	36	16
4	B+	23	93	37	26	7	4	28	4
4	C-	0	3	2	1	0	1	4	0
4	C+	1	15	5	3	1	0	2	0
5	N/A	0	142	99	0	1043	1	20	0
5	A	1251	999	1124	1259	237	1214	1229	1285
5	B-	25	19	28	20	4	55	22	8
5	B+	19	116	44	15	7	22	20	2
5	C-	0	5	0	1	1	2	4	0
5	C+	0	14	0	0	3	1	0	0
6	N/A	0	106	203	0	1076	79	50	0
6	A	1088	919	890	1093	62	1015	1045	1118
6	B-	19	22	18	25	0	35	24	14
6	B+	28	58	21	15	0	9	19	6
6	C-	2	4	2	1	0	0	0	0
6	C+	1	29	4	4	0	0	0	0
7	N/A	0	175	229	0	962	122	110	0
7	A	992	750	776	975	68	868	887	999
7	B-	23	8	4	38	0	24	17	27
7	B+	15	72	18	17	0	15	15	4
7	C-	0	3	1	0	0	0	0	0
7	C+	0	22	2	0	0	1	1	0
8	N/A	0	219	140	0	863	81	198	0
8	A	883	615	723	867	36	784	668	890
8	B-	13	18	21	19	1	20	24	11
8	B+	5	37	15	14	1	11	9	0
8	C-	0	3	1	1	0	1	0	0
8	C+	0	9	1	0	0	4	2	0
11	N/A	526	1772	1149	526	2643	1965	2186	526
11	A	2022	777	1453	2038	14	653	440	2089
11	B-	41	12	18	47	1	17	11	22
11	B+	56	69	31	42	0	22	18	18
11	C-	5	1	4	0	0	0	1	0
11	C+	8	27	3	5	0	1	2	3

Test Fairness and Implications for Ongoing Research

There are many features of the Smarter Balanced assessments that support equitable assessment across all groups of students. The assessments are developed using the principles of evidence-centered design and universal test design. Test accommodations are provided for students with disabilities, and language-tools and supports were developed for ELs. The Consortium developed a set of guidelines to facilitate accessibility to the assessments. In addition to these general accessibility guidelines embedded in the conceptual framework, procedures for item writing and reviewing and guidelines for creating audio, sign language, and tactile versions of the items were implemented. Smarter Balanced developed guidelines for item development that aim toward reducing construct-irrelevant language complexities for English language learners (Young, Pitoniak, King, & Ayad, 2012) and comprehensive guidelines for bias and sensitivity (ETS, 2012), and a rubric specifically geared towards scoring language complexity (Cook & MacDonald, 2013). In addition, measurement bias was investigated using DIF methods. This evidence underscores the commitment to fair and equitable assessment for all students, regardless of their gender, cultural heritage, disability status, native language, and other characteristics. Irrespective of these proactive development activities designed to promote equitable assessments, further validity evidence that the assessments are fair for all groups of students should be provided. To evaluate the degree to which the Smarter Balanced assessments are fulfilling the purpose of valid, reliable, and fair information that is equitable for all students, several types of additional evidence are recommended based on the relevant types listed in the AERA, APA, & NCME (2014) *Standards*. Validity studies are described here as well as ones that can be addressed in the ongoing research agenda for Smarter Balanced.

References

Abedi, J., & Ewers, N (2013). Accommodations for English language learners and students with disabilities: A research-based decision algorithm. Available from https://portal.smarterbalanced.org/library/en/v1.0/accommodations-for-english-language-learners-and-students-with-disabilities-a-research-based-decision-algorithm.pdf

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219-234.

Abedi, J., Lord, C., & Plummer, J. (1995). *Language background as a variable in NAEP mathematics performance* (CSE Technical Report 429). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disability Research and Practice, 20(1),* 50-57.

Bailey, A. L., Huang, B. H., Shin, H W., Farnsworth, T., & Butler, F. A., (2007) *Developing academic English language proficiency prototypes for 5th grade reading: Psychometric and linguistic profiles of tasks* (CSE Technical Report 727). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, *25,* 133–150.

Borgioli, G. (2008). Equity for English language learners in mathematics classrooms. *Teaching Children Mathematics, 15,* 185-191.

Cole, N.S., & Zieky, M. J. (2001). The New Faces of Fairness. *Journal of Educational Measurement. 38,* 4.

Cook, H.G. & McDonald, R. (2013). Tool to Evaluate Language Complexity of Test Items. Wisconsin Center for Education Research. Retrieved from http://wcer.wisc.edu/docs/working-papers/Working_Paper_No_2013_05.pdf

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 405-438.

Dorans, N. J., & Kulick, E. (1983). Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach (Research Report No. RR-83-09). Princeton, NJ: Educational Testing Service

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. Journal of Educational Measurement, 23, 355–368.

ETS. (2012). *Smarter Balanced Assessment Consortium: Bias and Sensitivity Guidelines.* Princeton, NJ: ETS. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/BiasandSensitivityGuidelines.pdf

Fedorchak, G. (2012). *Access by design—Implications for equity and excellence in education*. Draft paper prepared for the Smarter Balanced Assessment Consortium.

Forster, K. I. & Olbrei, I. (1973). Semantic heuristics and syntactic trial. *Cognition, 2,* 319-347.

Kopriva, R. (2010, September). *Building on student strengths or how to test ELs against challenging math (and science) standards when they don't have the English yet.* Common Core State Standards Implementation Conference.

Mantel N. & Haenszel, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute,* 22, 719-748

Measured Progress/ETS Collaborative. (April, 2012). Smarter Balanced Assessment Consortium: General Accessibility Guidelines. https://portal.smarterbalanced.org/library/en/general-accessibility-guidelines.pdf

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement.* Washington, DC: American Council on Education.

Pitoniak, M., Young, J. W., Martiniello, M., King, T., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners.* Princeton, NJ: Educational Testing Service.

Russell, M. (2011). *Digital Test Delivery: Empowering Accessible Test Design to Increase Test Validity for All Students*. Paper prepared for Arabella Advisors.

Schachter, P. (1983). *On syntactic categories.* Bloomington, IN: Indiana University Linguistics Club.

Shafer W., L., & Rivera, C. (2011). Are EL needs being defined appropriately for the next generation of computer-based tests? *AccELLerate*, *3*(2), 12–14.

Smarter Balanced Assessment Consortium. (2014a). *Accessibility and Accommodations Framework*. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/09/Accessibility-and-Accommodations-Framework.pdf

Smarter Balanced Assessment Consortium. (2014b, November 5). *Usability, Accessibility, and Accommodations Guidelines.*

Smarter Balanced Assessment Consortium. (2015a). *Content specifications for the summative assessment of the common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Retrieved from https://portal.smarterbalanced.org/library/en/english-language-artsliteracy-content-specifications.pdf

Smarter Balanced Assessment Consortium. (2015b). *Content specifications for the summative assessment of the common core state standards for mathematics*. Retrieved from https://portal.smarterbalanced.org/library/en/english-language-artsliteracy-content-specifications.pdf

Smarter Balanced Assessment Consortium. (2015c). *Item and task specifications*. Retrieved from http://www.smarterbalanced.org/assessments/development/

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Quenemoen, R. F., & Lazarus, S. S. (2011). *Meeting the Needs of Special Education Students: Recommendations for the Race to the Top Consortia and States*. Paper prepared for Arabella Advisors.

Young, J., Pitoniak, M. J., King, T. C., & Ayad, E. (2012). *Smarter Balanced Assessment Consortium: Guidelines for Accessibility for English Language Learners*. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/GuidelinesforAccessibilityforELL.pdf

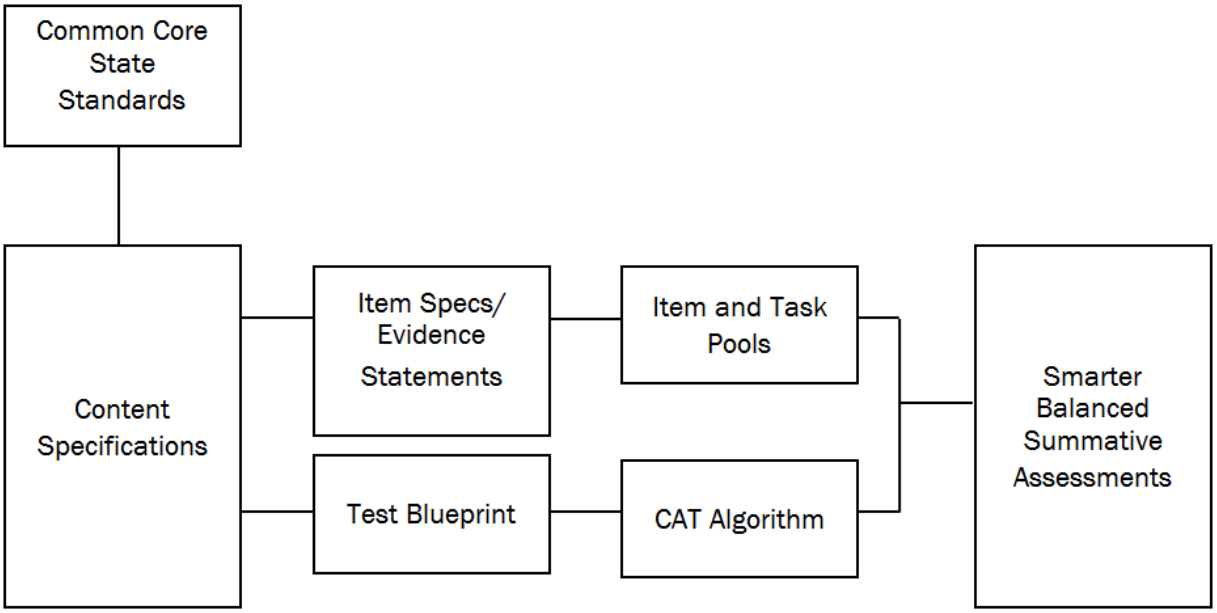
Young, J. W. (2008, December). Ensuring valid content tests for English language learners. *R&D Connections, No. 8*. Princeton, NJ: Educational Testing Service.

Chapter 4: Test Design

Introduction

Test design entails developing a test philosophy (i.e., Theory of Action), identifying test purposes, and determining the targeted examinee populations, test specifications, item pool design, and other features such as test delivery (Schmeiser & Welch, 2006). The Smarter Balanced Theory of Action, test purposes, and the targeted examinee population were outlined in the introduction of this report. Other elements of test design are further emphasized here, such as the interim assessments. In developing and maintaining a system of assessments, the goal of Smarter Balanced is to ensure that the assessment's measurement properties reflect industry standards for content, rigor, and performance. A key step in this direction is to ensure that the Smarter Balanced assessments are aligned with the Common Core State Standards (CCSS). Figure 4-1 briefly encapsulates the Smarter Balanced content structure.

FIGURE 4-1 COMPONENTS OF SMARTER BALANCED TEST DESIGN



A Brief Description of Smarter Balanced Content Structure

The Common Core State Standards are the content standards in English language arts/literacy (ELA/literacy) and mathematics that many states have adopted. Because the CCSS were not specifically developed for assessment, they contain extensive rationale and information concerning instruction. Therefore, adopting previous practices used by many state programs, Smarter Balanced content experts produced Content Specifications in ELA/Literacy and mathematics, which distill assessment-focused elements from the CCSS. The Smarter Balanced *Content Specifications for the Summative Assessment of the CCSS for English Language Arts/Literacy* (2015c) and *Content Specifications for the Summative Assessment of the CCSS for Mathematics* (2015d) were expressly created to guide the structure and content of assessment development. Within each of the two subject areas in grades 3 to 8 and high school, there are four broad claims. Within each claim, there are a number of assessment targets. The claims in ELA/literacy and mathematics are given in Table 4-1.

TABLE 4-1 CLAIMS FOR ELA/LITERACY AND MATHEMATICS

Claim	ELA/Literacy	Mathematics
1	Reading	Concepts and Procedures
2	Writing	Problem Solving
3	Speaking/Listening	Communicating Reasoning
4	Research	Modeling and Data Analysis

Currently, only the listening part of ELA/literacy Claim 3 is assessed. In mathematics, Claims 2 and 4 are reported together as a single subscore, so there are only three reporting categories for mathematics, but four claims.

Because of the breadth in coverage of the individual claims, targets within each claim were needed to define more specific performance expectations. The relationship between targets and Common Core State Standards elements is made explicit in the Smarter Balanced content specifications (2015c; 2015d).

The Smarter Balanced *Item and Task Specifications* (2015e) for ELA/literacy and mathematics provide guidance on how to translate the Smarter Balanced Content Specifications into actual assessment items. In addition, guidelines for bias and sensitivity, accessibility and accommodations, and style help item developers and reviewers ensure consistency and fairness across the item bank. The specifications and guidelines were reviewed by member states, school districts, higher education representatives, and other stakeholders. The item specifications describe the evidence to be elicited and provide sample task models to guide the development of items that measure student performance relative to the target.

The Smarter Balanced *ELA/Literacy Summative Assessment Blueprint* (2015a) and *Mathematics Summative Assessment Blueprint* (2015b) describe the content of the English language arts/literacy and math summative assessments for grades 3 to 8 and high school administered in 2016-2017—and how that content will be assessed. The blueprints also describe the composition of the two assessment components, computer adaptive test (CAT) and performance task (PT), and how their results will be combined for score reporting. For the computer adaptive component, specific items administered to each student are uniquely determined based on an item-selection algorithm that includes content constraints that correspond to the test blueprint. The performance tasks (PTs) act in concert with the computer adaptive test (CAT) items to fulfill the blueprint. Developed with broad input from member states, partners, and stakeholders, the summative test blueprints reflect the depth and breadth of the performance expectations of the CCSS. Smarter Balanced Governing Members adopted the preliminary test blueprints in 2012. The summative test blueprints developed subsequently contain refinements and revisions based on the analyses of the Pilot and Field Tests.

Synopsis of Assessment System Components

The summative assessment for each content area consists of two parts: a computer adaptive test (CAT) and a performance task (PT). The PT is administered on a computer but is not computer adaptive. The summative assessment is administered according to the guidance provided in the Smarter Balanced *State Procedures Manual* (2014). The summative assessment scores

- accurately describe student achievement and can describe growth of student learning as part of program evaluation and school, district, and state accountability systems;
- provide valid, reliable, and fair measures of students' progress toward, and attainment of, the knowledge and skills required to be college- and career-ready;
- Measure the breadth and depth of the CCSS across the full spectrum of student ability by incorporating a variety of item types (including items and tasks scored by expert raters) that are supported by a comprehensive set of accessibility resources;
- capitalize on the strengths of computer adaptive testing—efficient and precise measurement across the full range of student achievement; and
- utilize performance tasks to provide a measure of the student's ability to integrate knowledge and skills.

Evidence-Centered Design in Constructing Smarter Balanced Assessments

Evidence-centered design (ECD) is an approach to the creation of educational assessments in terms of reasoning about evidence (arguments) concerning the intended constructs. The ECD begins with identification of claims, or inference users want to make concerning student achievement. Evidence needed to support those claims is then specified, and finally, items/tasks capable of eliciting that information are designed (Mislevy, Steinberg, & Almond, 2003). Explicit attention is paid to the potential influence of unintended constructs. ECD accomplishes this in two ways. The first is by incorporating an overarching concept of assessment as an argument from imperfect evidence. This argument makes explicit the claims (the inferences that one intends to make based on scores) and the nature of the evidence that supports those claims (Hansen & Mislevy, 2008; Mislevy & Haertel, 2006). The second is by distinguishing the activities and structures involved in the assessment enterprise in order to exemplify an assessment argument in operational processes. By making the underlying evidentiary argument more explicit, the framework makes operational elements more amenable to examination, sharing, and refinement. Making the argument more explicit also helps designers meet diverse assessment needs caused by changing technological, social, and legal environments (Hansen & Mislevy, 2008; Zhang, Haertel, Javitz, Mislevy, Murray, & Wasson, 2009). The ECD process entails five types of activities, or layers, of assessment. The activities focus on 1) the identification of the substantive domain to be assessed; 2) the assessment argument; 3) the structure of assessment elements such as tasks, rubrics, and psychometric models; 4) the implementation of these elements; and 5) the way they function in an operational assessment, as described below.

- **Domain Analysis.** In this first layer, domain analysis involves determining the specific content to be included in the assessment. Smarter Balanced uses the Common Core State Standards as its content domain for mathematics and ELA/literacy. Domain analysis was conducted by

the developers of the CCSSs, who first developed college- and career-readiness standards, to address what students are expected to know and be able to do by the time they graduate from high school. This was followed by development of the K-12 standards, which address expectations for students in elementary through high school.

- **Domain Modeling.** In domain modeling, a high-level description of the overall components of the assessment is created and documented. For Smarter Balanced, the components include computer-adaptive summative and interim assessments in mathematics and ELA/literacy. The domain framework was developed by organizing the CCSS into domain areas that form the structure of test blueprints and reporting categories. This overall structure was created in the course of Smarter Balanced content specification development.
- **The Conceptual Assessment Framework.** Next, the conceptual assessment framework is developed. For Smarter Balanced, this step was accomplished in developing the Smarter Balanced content specifications, which identify major claim structure, targets within claims, and the relationship of those elements to underlying content of the CCSS. In this step, the knowledge, skills, and abilities to be assessed (i.e. *intended constructs*, *targets of assessment*), the evidence that needs to be collected, and the features of the tasks that will elicit the evidence are specified in detail. Ancillary constructs that may be required to respond correctly to an assessment task but are not the intended target of the assessment are also specified (e.g., reading skills in a mathematics examination). By identifying any ancillary knowledge, skills, and abilities (KSAs), construct-irrelevant variance can be identified a priori and minimized during item and task development—potential barriers created by the ancillary KSAs can be removed or their effects minimized through the provision of appropriate access features. The item and task specifications describe the evidence required to support claims about the assessment targets and also identify any ancillary constructs.
- **Implementation.** This layer involves the development of the assessment items or tasks using the specifications created in the conceptual assessment framework just described. In addition, scoring rubrics are created and the scoring process is specified. Smarter Balanced items, performance tasks, and associated scoring rubrics were developed starting in the spring of 2012.
- **Delivery.** In this final layer, the processes for the assessment administration and reporting are created. The delivery system describes the adaptive algorithm, collection of student evidence, task assembly, and presentation models required for the assessment and how they function together. The ECD elements chosen lead to the best evaluation of the construct for the intended test purposes. Test delivery and test scoring are discussed below.

Test Blueprints

Test specifications and blueprints define the knowledge, skills, and abilities intended to be measured on each student's test event. A blueprint also specifies how skills are sampled from a set of content standards (i.e., the CCSS). Other important factors such as Depth of Knowledge (DOK) are also specified. Specifically, a test blueprint is a formal document that guides the development and assembly of an assessment by explicating the following types of essential information:

- content (claims and assessment targets) that is included for each assessed subject and grade, across various levels of the system (student, classroom, school, district, state);
- the relative emphasis of content standards generally indicated as the number of items or percentage of points per claim and assessment target;
- item types used or required, which communicate to item developers how to measure each claim and assessment target, and to teachers and students about learning expectations; and
- Depth of Knowledge (DOK), indicating the complexity of item types for each claim and assessment target.

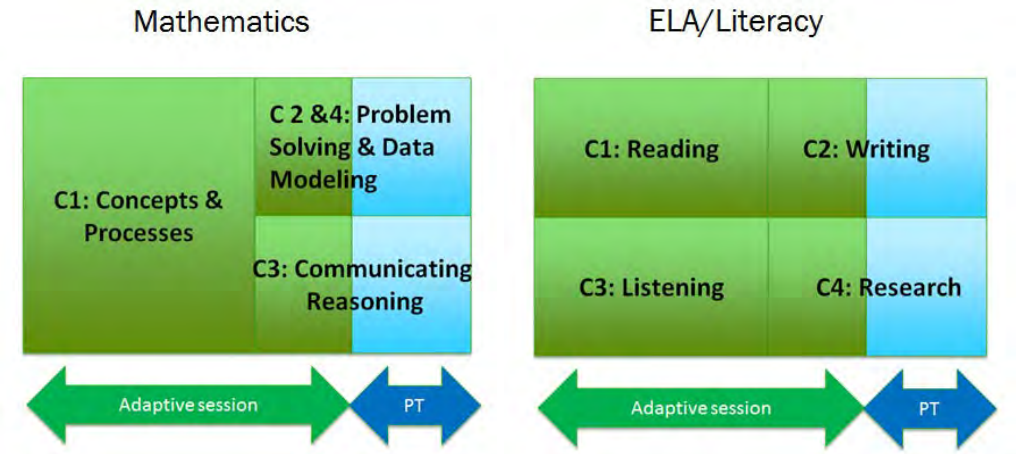
The test blueprint is an essential guide for both assessment developers and for curriculum and instruction. For assessment developers, the blueprint and related test-specification documents define how the test will ensure coverage of the full breadth and depth of content and how it will maintain fidelity to the intent of the CCSS on which the Smarter Balanced assessment is based. Full content alignment is necessary in order to ensure that educational stakeholders can make valid, reliable, and unbiased inferences about student, classroom, school, and state performance. At the instructional level, the test blueprint provides a guide to the relative importance of competing content demands and suggests how the content is demonstrated, as indicated by item type and depth-of-knowledge. In summary, an assessment blueprint provides clear development specifications and signals to the broader education community both the full complexity of the CCSS and how performance on these standards is substantiated.

Part of the innovative aspect of the Smarter Balanced assessments is that the test blueprints sample the content domain using both a computer adaptive component (CAT) and a performance task (PT). The test blueprints can be inspected to determine the contribution of the CAT and PT components in a grade and content area toward the construct intended to be measured. Another aspect of the assessments is the provision of a variety of both machine-scored and human-scored item types. The contribution of these item types is specified in the Smarter Balanced test blueprints.

In February 2015, the Governing Members of the Smarter Balanced Assessment Consortium adopted blueprints for the summative assessments of mathematics and ELA/literacy for grades 3 to 8 and high school (Smarter Balanced, 2015a; Smarter Balanced, 2015b). These were fully implemented in the 2014-15 school year and were in effect in the 2016-2017 school year.

The summative assessment is composed of the CAT and PT components. Responses from both components are combined to cover the test blueprint in a grade and content area and are used to produce the overall and claim scale scores. Figure 4-2 is a conceptual diagram of how claims are distributed across the adaptive and performance task parts of the tests.

FIGURE 4-2 CLAIM DISTRIBUTION IN TEST BLUEPRINTS



Operational Summative Assessment Blueprints and Specifications.

For each designated grade range (3 to 5, 6 to 8, and high school), the test blueprints present specifications for numbers of items and points by claim score\reporting category, content category, type of stimuli, CAT component, and performance task. (Figure 4-3 shows the mathematics blueprint for the grade 6 to 8 range and is discussed below.) Details are given separately for each grade and include claim, assessment target, DOK, assessment type (CAT/PT), and the total number of items (Smarter Balanced, 2015a; Smarter Balanced, 2015b). (Figures 4-4 and 4-5 show this more detailed blueprint for grade 6 mathematics and are discussed below.) Assessment targets are nested within claims and represent a more detailed specification of content. Note that in addition to the nested hierarchical structure, each blueprint also specifies a number of rules applied at global or claim levels. Most of these specifications are in the footnotes, which constitute important parts of the test designs.

The CAT algorithm selects items necessary to conform to the test blueprint and at the same time meet requirements for measurement precision. In establishing requirements for content and measurement precision in the CAT component, designers took advantage of the adaptive pool to allow more variety than would be present in a fixed form test. For example, when the number of content targets in a domain area is large, blueprints allow choice within target clusters rather than limiting the number of targets. Since all content targets are represented in the pool, any student can potentially get any target while the full set of content constraints is still maintained.

To assist in blueprint interpretation, an example of a mathematics summative blueprint is given in Figure 4-3 through Figure 4-5. Figure 4-3 presents higher level blueprint requirements pertaining to the grades 6 to 8 band. The high level requirements are expressed in terms of the number of stimuli (for items that are associated with a common stimulus) and items per claim within the CAT and PT segments of the assessment. In all figures, footnotes are an important component of the blueprint.

FIGURE 4-3 OVERVIEW OF MATHEMATICS GRADE 6-8 SUMMATIVE BLUEPRINT

Blueprint Table Mathematics Grades 6 8 Estimated Total Testing Time: 3:30 (with Classroom Activity)[1]						
Claim/Score Reporting Category	Content Category[2]	Stimuli		Items		Total Items by Claim[3]
		CAT	PT	CAT[4]	PT[5]	
1. Concepts and Procedures	Priority Cluster	0	0	12-15	0	16-20
	Supporting Cluster	0		4-5		
2. Problem Solving 4. Modeling and Data Analysis[6]	Problem Solving	0	1	6	2-4	8-10
	Modeling and Data Analysis	0				
3. Communicating Reasoning	Communicating Reasoning	0		8	0-2	8-10

All times are estimates. Actual times may vary.
[2] For more information on content categories, see the Content Specifications document at http://www.smarterbalanced.org/assessments/development
[3] While the range for the total items by Claim for Problem Solving/Modeling and Data Analysis and Communicating Reasoning indicates 8-10 items in each reporting category, the total number of items across these two reporting categories for any individual test event is 18-20.
[4] In grades 6-8, up to one CAT item per student may require hand-scoring (from either Claim 3 or Claim 4), which may be AI-scored with an application that yields comparable results by meeting or exceeding reliability and validity criteria for hand-scoring.
[5] Each PT contains 4-6 total items. Up to four PT items may require hand-scoring.
[6] Claim 2 (Problem Solving) and Claim 4 (Modeling and Data Analysis) have been combined because of content similarity and to provide flexibility for item development. There are still four claims, but only three claim scores will be reported with the overall math score.

Figures 4-4 and 4-5 present more detailed blueprint requirements for grade six at the target level. Note that some targets are clustered together. For example, Claim 1 calls for 14 items from targets E, F, A, G, B, and D. Note that five to six items come from targets E and F, while only two items come from targets G and B. This represents the appropriate content emphasis, while allowing flexibility in item choice. The detailed blueprint shows how performance tasks and CAT components work in conjunction. The DOK requirements are applied at the claim level and are stated in the footnotes.

FIGURE 4-4 BLUEPRINT FOR GRADE 6 SHOWING DETAILED CONTENT STRUCTURE (ASSESSMENT TARGETS), PAGE 1 OF 2

Target Sampling Mathematics Grade 6						
Claim	Content Category	Assessment Targets	DOK	Items		Total
				CAT	PT	
1. Concepts and Procedures	Priority Cluster	E. Apply and extend previous understandings of arithmetic to algebraic expressions.	1	5-6	0	16-19
		F. Reason about and solve one-variable equations and inequalities.	1, 2			
		A. Understand ratio concepts and use ratio reasoning to solve problems.	1, 2	3-4		
		G. Represent and analyze quantitative relationships between dependent and independent variables.	2	2		
		B. Apply and extend previous understandings of multiplication and division to divide fractions by fractions.	1, 2			
		D. Apply and extend previous understandings of numbers to the system of rational numbers.	1, 2	2		
	Supporting Cluster	C. Compute fluently with multi-digit numbers and find common factors and multiples.	1, 2	4-5		
		H. Solve real-world and mathematical problems involving area, surface area, and volume.	1, 2			
		I. Develop understanding of statistical variability.	2			
		J. Summarize and describe distributions.	1, 2			

— DOK: Depth of Knowledge, consistent with the Smarter Balanced Content Specifications.
— The CAT algorithm will be configured to ensure the following:
- For Claim 1, each student will receive at least 7 CAT items at DOK 2 or higher.
- For Claim 3, each student will receive at least 2 CAT items at DOK 3 or higher.
- For combined Claims 2 and 4, each student will receive at least 2 CAT items at DOK 3 or higher.

FIGURE 4-5 BLUEPRINT FOR GRADE 6 SHOWING DETAILED CONTENT STRUCTURE (ASSESSMENT TARGETS), PAGE 2 OF 2

Target Sampling Mathematics Grade 6						
Claim	Content Category	Assessment Targets	DOK	Items		Total Items
				CAT	PT	
2. Problem Solving 4. Modeling and Data Analysis	Problem Solving (drawn across content domains)	A. Apply mathematics to solve well-posed problems arising in everyday life, society, and the workplace.	2, 3	2	1–2	8-10
		B. Select and use appropriate tools strategically. C. Interpret results in the context of a situation. D. Identify important quantities in a practical situation and map their relationships (e.g., using diagrams, two-way tables, graphs, flow charts, or formulas).	1, 2, 3	1		
	Modeling and Data Analysis (drawn across content domains)	A. Apply mathematics to solve problems arising in everyday life, society, and the workplace. D. Interpret results in the context of a situation.	2, 3	1	1–3	
		B. Construct, autonomously, chains of reasoning to justify mathematical models used, interpretations made, and solutions proposed for a complex problem. E. Analyze the adequacy of and make improvements to an existing model or develop a mathematical model of a real phenomenon.	2, 3, 4	1		
		C. State logical assumptions being used. F. Identify important quantities in a practical situation and map their relationships (e.g., using diagrams, two-way tables, graphs, flow charts, or formulas).	1, 2, 3	1		
		G. Identify, analyze, and synthesize relevant external resources to pose or solve problems.	3, 4	0		
3. Communicating Reasoning	Communicat-ing Reasoning (drawn across content domains)	A. Test propositions or conjectures with specific examples. D. Use the technique of breaking an argument into cases.	2, 3	3	0-2	8-10
		B. Construct, autonomously, chains of reasoning that will justify or refute propositions or conjectures. E. Distinguish correct logic or reasoning from that which is flawed, and—if there is a flaw in the argument—explain what it is.	2, 3, 4	3		
		C. State logical assumptions being used. F. Base arguments on concrete referents such as objects, drawings, diagrams, and actions. G. At later grades, determine conditions under which an argument does and does not apply. (For example, area increases with perimeter for squares, but not for all plane figures.)	2, 3	2		

— DOK: Depth of Knowledge, consistent with the Smarter Balanced Content Specifications.
— The CAT algorithm will be configured to ensure the following:
- For Claim 1, each student will receive at least 7 CAT items at DOK 2 or higher.
- For Claim 3, each student will receive at least 2 CAT items at DOK 3 or higher.
- For combined Claims 2 and 4, each student will receive at least 2 CAT items at DOK 3 or higher.

CAT and Performance Task Test Components

Part of the Smarter Balanced Theory of Action is to leverage appropriate technology and innovation. The use of CAT methodologies helps ensure that students across the range of proficiency have an assessment experience with items well targeted to their skill level. Adaptive testing allows average-, very low-, and very high-performing students to stay engaged in the assessment because they respond to items specifically targeted to their skill level. Non-adaptive performance tasks measure a student's ability to integrate knowledge and skills across multiple standards. No order is imposed on the components; either the CAT or PT portion can be administered to students first.

CAT tests are more efficient in that they provide a higher level of score precision than fixed form tests with the same number of items. For the CAT component, there are both content constraints (e.g., a long reading passage in ELA/literacy must be administered) as well as psychometric criteria that must be optimized for each student. Performance tasks are intended to measure multiple standards in a coherent task that requires use of integrated skill sets. Performance tasks measure capacities such as essay writing, research skills, and complex analysis, which are not as easy to assess with individual, discrete items.

Adaptive Test Design and Algorithm

This section describes the method used in the Smarter Balanced system to satisfy the blueprint and provide optimal precision. The implementation described here is released under the Creative Commons Attribution Only, No Derivatives license. This document is a summary with supplemental explanations and examples of explicit functionality found in the separate, *Smarter Balanced Adaptive Item Selection Algorithm Design Report* by Jon Cohen and Larry Albright (2014). Interested readers can refer to the more detailed document for more technical information and specific formulas the algorithm employs.

For the operational test, an item-level, fully adaptive CAT component is administered in ELA/literacy and mathematics. The adaptive part delivers blueprints in a manner that efficiently minimizes measurement error and maximizes information. To assert that test results are comparable across the consortium, members must adopt an algorithm that delivers the published blueprint. Three potential scenarios through which this could be accomplished are listed below:

- Members may deliver Smarter Balanced assessments using the open source software for both the test delivery system and adaptive algorithm.
- Members may use the open source software for one component and a service provider solution for the other (e.g., open source test delivery system, and a vendor's algorithm that can be appropriately configured).
- Members may use service provider solutions for both components.

Members who use service providers for one or both components may have to respond independently to federal peer review requirements concerning test characteristics that depend on the test delivery platform, including algorithms for item selection and blueprint fulfillment.

Automated test assembly for a CAT depends on a number of factors to produce optimal tests. These depend on the quality of the item bank, reasonableness of the test constraints and precision targets,

and the degree to which content or other qualitative attributes of items are salient and can be defined as constraints.

In general, an adaptive algorithm is a method used to carry out a blueprint design by acting on an item pool. The algorithm finds the items expected to compose the best test for each student, selecting items from the pool that match blueprint demands while using information from a student's responses to find the most accurate score. The blueprint describes in detail the content and other attributes for each student's test. Both the algorithm and items in the pool must support blueprints in order to deliver accurate, efficient tests.

Item attributes

Specified in blueprints and needed to run the algorithm include depth of knowledge, response type, scoring type, common stimulus membership and mathematical domain. Complete information about these elements must be available to algorithm software for all items. The minimum and maximum number of items in each element is specified in the adaptive software, serving as a constraint to balance aspects such as blueprint coverage with test length. Each element can be given weights to guide the item selection process in an optimal fashion. By allowing for the specification of weights, the general algorithm can be customized for varying conditions of population and pool distribution. For example, weights can be shifted to emphasize measurement precision or content coverage, depending on policy priorities. Final weights are established during the last stages of test design when all item parameters are known and simulation results are available.

Item measurement data

In addition to the blueprint attributes listed above, each item has a set of parameters that provide measurement information. The purpose of the algorithm is to satisfy the content blueprint while providing the most accurate student score, in the most efficient manner. In measurement terms, the most information is obtained when the difficulty of the item is close to the performance level of the student. At the beginning of the test, item difficulty and discriminating power are known, and student ability is unknown. The job of the algorithm is to find out the student's ability in the content area being assessed by comparing the student's performance to the known, statistical attributes of the items.

Test Operation Walkthrough

Preparation

The test delivery system must have in place a sufficient item pool with the full set of parameters and metadata. Smarter pools contain all items for the intended grade level and items from adjacent grades that address on-grade content. Items from upper grades address content the student has had an opportunity to learn. Items from lower grades are screened for age-appropriateness. Initially, the pool is filtered to include only items that were assigned to the examinee's grade. Under certain circumstances (described below) the filter is dropped and items assigned to an adjacent grade (higher or lower, depending on the student's performance) are added if the content they represent is also appropriate for the examinee's grade. The adaptive engine needs to be populated with all hierarchical and content elements, including the minimum and maximum number of items allowed for each facet of the blueprint.

Initialization.

Adaptive tests require methods for avoiding overuse of items. In the 2014-15 summative tests, the algorithm was configured to choose each test's initial item randomly from the pool. The initial claim is chosen at random as long as passages and hand-scored items are not presented first. The algorithm then cycles through the claims. This strategy is still in use as of the 2016-2017 assessment.

Item selection.

The initialization and selection processes attempt to avoid underuse and overuse of items – to control item exposure. Exposure control enhances item security and discourages copying and cheating by presenting a variety of items. It also leads to more efficient pool use, assuring that all items developed to cover the content are used. Rather than choosing the single best item for initialization and selection, which would cause some items to be used repeatedly and others rarely or never, the algorithm selects randomly from targeted sets of items. To prevent overuse of highly discriminating items, the discrimination (a) parameter is not taken into account in selection ranking. The optimal size of the first content-based set and the subsequent subset, which takes information into account, was determined through simulation with actual pool parameters.

Once the student's response to the first item is scored, the selection process is launched and will be repeated for every subsequent response. The software uses the set of weights described earlier to determine a group of items with the best match to the blueprint, excluding items from target groups that have reached the maximum number of items specified in the blueprint and items previously seen by the examinee. When this mini pool (subset of items) has been chosen, the information value is calculated for each item using the current student ability estimate and known item parameters. Overall item value is calculated using both information and content data. The item set is then sorted according to overall value and a set of the most preferred items are identified. The item to be administered is chosen randomly from within this set. After each response to a machine-scored item is scored, the student ability estimate is updated. The selection procedure is then repeated until the blueprint has been satisfied.

The algorithm proceeds in this manner until a percentage of the test (coverage in mathematics, 61%; ELA/literacy, 62%.) has been administered, sampling items from all claim areas. At this point the distance of the estimated score from the level 3 cut score is evaluated. If the student is either above or below the level 3 cut score with probability $p<.000001$, the item pool is expanded to include items from no more than two adjacent grades in the direction of the difference. In grade 3, the pool is expanded only for students above the level 3 cut score and includes items from adjacent upper grades only; in grade 11 the pool is expanded only for students below the level 3 cut score and includes items from adjacent lower grades only. Items from adjacent grades have been screened for appropriateness by content experts to assure that they represent the content of the target grade. For the remainder of the test, both on-grade and off-grade items can be administered. The item with the best content and measurement characteristics is chosen from the pool.

Early in the development process, Consortium members determined that students should be allowed to go back to earlier questions, review their answers and revise their answers if necessary. Students can go back and change their answers within a test segment. When this occurs for machine-scored items, the ongoing student score estimate is updated with the new response. This has implications

for test design and delivery. If a student takes a test over the course of two or more days, answers from previous days cannot be changed. In mathematics, some items permit the use of a calculator, while others forbid calculator use. Mathematics tests are consequently divided into two sections, one for non-calculator items, and one that permits calculator use. Students can change answers within sections but not across different test sections.

Note that blueprints call for the administration of human-scored items during the adaptive part of the test. The blueprints specify that these items may be AI scored with an application that yields comparable results by meeting or exceeding reliability and validity criteria for hand-scoring. These items are chosen based on their information value just like machine-scored items. However, the adaptive engine is designed to work asynchronously from hand-scoring. Because the response to the item is not immediately scored, the adaptive engine proceeds using the most recent estimate of student ability and selects the next item accordingly.

Termination

The test ends when the blueprint has been met. At that point, student machine-scored responses are retained.

Test Scoring

The method of combining item scores to produce test scores and sub-scores is presented in detail in the *Smarter Balanced Scoring Specification* document (AIR, 2014). Scores are calculated using maximum likelihood estimation (MLE) applied at the overall and sub-score levels. No special weights for claims, item types or performance tasks are applied. Desired domain representations is achieved by the numbers of items and points specified in the blueprints.

Item and Task Development

In order to build a summative assessment that measured the intended claims, the Consortium's test development cycle was iterative, involving experts from various education-related fields, and was based on assessment-related research and best practices.

Item and Task Specifications

The item and task specifications bridge the distance from the content specifications and Achievement Levels to the assessment itself. While the content specifications establish the Consortium's claims and the types of evidence that is needed to support these claims, more specificity is needed to develop items and tasks that measured the claims.

The first iteration of the item and task specifications was developed in 2011. In early 2012, the Consortium held a series of showcases where the contractors introduced the item and task specifications and collected feedback from member states. Using this feedback, the item and tasks specifications were revised during the first quarter of 2012.

Using the revised item and task specifications, a small set of items was developed and administered in fall 2012 during a small-scale trial. This provided the Consortium with the first opportunity to administer and score the new item types. During the small-scale trials, the Consortium also conducted cognitive laboratories to better understand how students respond to various types of

items (AIR, 2013). The cognitive laboratories used a think-aloud methodology in which students speak their thoughts while working on a test item. The item and task specifications were again revised based on the findings of the cognitive laboratories and the small-scale trial. These revised specifications were used to develop items for the 2013 pilot test, and they were again revised based on 2013 pilot test results and subsequent reviews by content experts.

The Smarter Balanced *Item and Task Specifications* (2015e) are designed to ensure that assessment items measure the assessment's claims. Indeed, the purpose of item and task specifications is to define the characteristics of items and tasks that will provide evidence to support one or more claims. To do this, the item and task specifications delineate types of evidence that should be elicited for each claim within a grade level. Then, the specifications provide explicit guidance on how to write items in order to elicit the desired evidence.

Item and task specifications provide guidelines on how to create items specific to each claim and assessment target through the use of task models. In mathematics a task model provides a description of an item/task's key features. These task models describe the knowledge, skills, and processes being measured by each of the item types aligned to particular targets. In addition, task models sometimes provide examples of plausible distractors. Exemplar items are provided within every task model. In ELA/literacy these functions are carried out through item specifications.

Task models were developed for each grade level and target to delineate the expectations of knowledge and skill to be represented through test questions at each grade. In addition, both ELA/literacy and mathematics item and stimulus specifications provide guidance about grade appropriateness of task and stimulus materials (the materials that a student must refer to in working on a test question). The task and stimulus models also provide information on characteristics of stimuli or activities to avoid because they are not germane to the knowledge, skill, or process being measured.

Guidelines concerning what to avoid in item writing are important because the underscore the Consortium's efforts to use universal design principles to develop items that are accessible to the widest range of students possible. As the name suggests, the concept of universal design aims to create items that accurately measure the assessment target for all students. At the same time, universal design recognizes that one solution rarely works for all students. Instead, this framework acknowledges "the need for alternatives to suit many different people." (Rose & Meyer, 2000, p. 4).

To facilitate the application of universal design principles, item writers are trained to consider the full range of students who may answer a test question. A simple example of this is the use of vocabulary that is expected to be known by all third-grade students versus only those third-grade students who play basketball. Almost all third-grade students are familiar with activities (e.g., recess) that happen during their school day, while only a subset of these students will be familiar with basketball terms like "double dribble," "layup," "zone defense," or "full-court press."

Item specifications discuss accessibility issues unique to the creation of items for a particular claim and/or assessment target. Accessibility issues concern supports that various groups of students may need to access item content. By considering the supports that may be needed for each item, item writers are able to create items that can be adapted to a variety of needs.

The use of universal design principles allows the Consortium to collect evidence on the widest possible range of students. By writing items that adhere to item and task specifications, the

Consortium is assured that assessments measure the claims and assessment targets established in content specifications as well as the knowledge, skills, and processes found in the CCSS for *all* students for whom the assessment is appropriate.

Performance Task Design

The Race to the Top Assessment Program Application for the Smarter Balanced Assessment Consortium (June, 2010) highlights the importance of performance tasks to "provide a measure of the student's ability to integrate knowledge and skills across multiple standards—a key component of college and career readiness" (p. 42). The development of an assessment system that fulfills this goal requires an understanding of how the world is changing and what skills are required to compete in an increasingly global economy. Research suggests that measuring college and career readiness will increasingly require the use of performance-based assessments (Fadel, Honey, & Pasnik, 2007).

A key component of college and career readiness is the ability to integrate knowledge and skills across multiple content standards. Smarter Balanced derives inferences concerning this ability through performance tasks. Performance assessments are intended to represent students' competence in applying the knowledge and cognitive skills needed to solve substantive, meaningful problems. Performance assessments give students opportunities to demonstrate their ability to find, organize, or use information to solve problems, undertake research, frame and conduct investigations, analyze and synthesize data, and apply learning to novel situations.

A Smarter Balanced performance task involves 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. Stimuli include a variety of information forms (e.g., readings, video clips, data), as well as an assignment or problem situation. As shown in the test blueprints, performance tasks are an integral part of the Smarter Balanced test design. When a performance task is assigned and given in its entirety, it fulfills a specific role in the test blueprint for a grade and content area. Performance tasks are intended to challenge students in applying their knowledge and skills to complex, contextually rich problems. These activities are meant to measure capacities such as depth of understanding, writing or research skills, mathematical modeling and complex analysis. They consist of collections of questions and activities coherently connected to a single scenario. The performance tasks are administered online via computer (not computer adaptive) and require one to two class periods to complete.

Performance tasks were constructed so they can be delivered effectively in the school/classroom environment (Dana and Tippins, 1993). Requirements for task specifications included, but were not limited to, pre-assessment classroom activities, materials and technology needs, and allotted time for assessment. Performance tasks adhere to specifications used by item writers to develop new tasks that focus on different content but are comparable in contribution to the blueprint.

All Smarter Balanced performance tasks consist of three basic components: stimulus presentation, information processing, and scorable product(s) or performance(s). "Information processing" means student interactions with the stimulus materials and their content. It could include note taking, data generation, and any other activities that increase students' understanding of the stimulus content or the assignment. All activities within a task must have a rationale for inclusion (e.g., to increase understanding, for scaffolding, as early steps in product creation or for product creation).

In ELA/literacy, each performance task comprises a targeted research effort in which students read sources and respond to two to three research items, followed by an essay. During the research component, students may take notes to which they may later refer. Students then write a full essay drawing from source material and research notes. Claim level results in writing and research are based on both CAT and performance task item responses.

In mathematics, each performance task comprises a set of stimulus materials and a follow-up item set consisting of six items in Claims 2, 3, and 4. These are combined with CAT items in Claims 2, 3 and 4 to satisfy the blueprint and create a Claim 3 score and a combined Claim 2 and 4 score. Performance tasks address an integrated scenario in middle and high school and a common theme in grades 3 to 5.

The Item/task Pool Specification

An **item pool** refers to a collection of test questions (known as items) that supports the test blueprint for a particular content area and grade. The Consortium takes multiple steps to ensure the quality of the items in our item pool. Building on the ongoing process of developing item/task specifications and test blueprints, the Consortium uses an iterative process for creating and revising each item as well as the collection of items. In the initial item development process, the Consortium tested items and refined its approach to item development through three steps: small-scale tryouts, a large pilot test, and a large field test. Details of the pilot and field tests are found in the Smarter Balanced *2013 - 2014 Technical Report* (2016). During each phase, the Consortium used cognitive laboratories to understand the strategies that students used to respond to the items. By incorporating this tiered and iterative approach, the item and task specifications that guided the development of the final operational pool were improved based on lessons learned during tryouts.

Using test blueprints, measurement experts specified the number and distribution of items to be written. Pools of items/tasks were written specifically to support proportions of items and intended difficulty distribution in the operational blueprint. Teachers were integrally involved in the creation of the item/task pool from beginning to end. Some participated in the processes described in the flow charts that appear in the Appendix A. Others developed items through a rigorous item writing process, and yet others reviewed the items for accuracy and appropriateness of the content knowledge and skill level required to respond to the items. Teams of content experts reviewed items for potential issues of bias in favor of or against any demographic group of students, and for accessibility for students with disabilities and English language learners. Content, bias, and accessibility reviews were conducted prior to administration to any students. Following pilot and field test administrations, items were again reviewed if pilot or field test data indicated a potential problem. Finally, teachers participated in range finding and scoring of constructed-response items/tasks to ensure that the items/tasks could be properly scored given their scoring rubrics.

In this section, we will examine the primary role that educators play in creating the field-test item pool by writing, reviewing, and scoring items. This section will end by examining the current composition of the item pool.

Item Writing

The Consortium works with educators throughout the test development cycle to develop items. All K-12 participants:

- are certified/licensed to teach ELA/literacy and/or mathematics in a K-12 public school;
- are currently teaching in a public school within a Smarter Balanced Governing State;
- have taught ELA/literacy and/or mathematics in grades 3 through 8 and/or high school within the past three years (second-grade teachers are also recruited to participate in the development of grade 3 items and/or tasks);
- have previously reviewed part or all of the CCSS for the content area for which they are writing items and/or performance tasks;
- have submitted a statement of interest that describes their interest in developing Smarter Balanced items and/or performance tasks as well as their qualifications for doing so;
- have completed training and achieved qualifications through a certification process.

Qualifications for Higher Education Faculty include:

- current employment with, or recent retirement from, a college or university located within a Smarter Balanced Member State;
- having taught developmental and/or entry-level courses in English, composition, mathematics, statistics or a related discipline within the last 3 years;
- having previously reviewed part or all of the CCSS for the content area in which they are interested in writing items and/or performance tasks;
- having completed training and achieving qualifications through the certification process.

The Consortium's staff train contractors and educators on the item specifications, ELA/literacy stimulus specifications, as well as guidelines for accessibility, bias and sensitivity, as described in the next section.

Prior to the spring 2013 pilot test, the Consortium engaged 136 educators in K-12 and higher education from 19 member states to write items. Prior to the spring 2014 field test, 184 educators in K-12 and higher education from 16 member states participated in item writing. The items developed in this process were used in the 2014 field test and in the 2015 embedded field test. These items account for all of the items used in the 2016-2017 summative assessment.

Training

For the development of all operational items in the 2016-2017 summative assessment, educators participated in a series of facilitated, online webinars in order to qualify as item writers. To facilitate participation, the Consortium scheduled multiple sessions in different time zones, including evening sessions. In addition to the facilitated sessions, the Consortium provided training modules that covered background on the Consortium, assessment design principles, and detailed information about item and performance task development. All modules were available in three formats: a

PowerPoint presentation with notes, a streaming presentation with narration that could be viewed online, and a downloadable audio/video presentation.

For all item writing, including more recent processes, item writers are specifically trained on the Consortium's content and item specifications, stimulus specifications, sensitivity and bias guidelines, and general accessibility guidelines. Training on these specifications and guidelines helps ensure that item writers are trained to write items that allow the widest possible range of students to demonstrate their knowledge, skills, and cognitive processes with regard to the content. This means that item writers need to understand the content for which they were writing items as well as accessibility and sensitivity issues that might hinder students' ability to answer an item. Item writers are also trained to be aware of issues that might unintentionally bias an item for or against a particular group.

Educator Participation

Educators were the primary developers of items used operationally in the 2016-2017 assessment. The active involvement of educators is critical to the success of the item writing activities. Educators engage with students on a daily basis, and they understand the ways in which students can demonstrate their knowledge. Their involvement in item writing helps ensure that the assessment system is accurate and efficient, and provides valid evidence of student learning.

State-Managed Item Development

The Consortium invites member states to participate in a separate effort to write items. This voluntary effort, known as State-Managed Item Development, is conducted to build the capacity of states to write items and to support the overall sustainability of the Consortium. To this end, six states (HI, ID, MI, WA, WV, and WY) participated in the state-managed field test item development opportunity. During this opportunity, educators within the six states developed approximately 3,100 items in mathematics and ELA/literacy across grades 3 through 8 and high school.

Item Reviews

Once items are written, groups of educators review items prior to field testing. The reviews are guided as follows.

Accessibility, Bias/Sensitivity, and Content Reviews

Panels of educators review all items, performance tasks, and item stimuli for accessibility, bias/sensitivity, and content. Item stimuli refer to the reading passages used on the ELA/literacy assessments or to the stimulus materials provided in the performance tasks in both mathematics and ELA/literacy. Prior to the spring 2013 pilot test, 122 ELA/literacy educators and 106 mathematics educators reviewed items and performance tasks for accessibility, bias/sensitivity, or content, and 60 educators reviewed the ELA/literacy stimuli. Prior to the spring 2014 field test, 107 ELA/literacy educators and 157 mathematics educators from 14 states reviewed items and performance, and 95 educators from 13 states reviewed the ELA/literacy stimuli.

The educator qualifications for the accessibility, bias/sensitivity, and content reviews are the same as the educator qualifications for item writing except that participants are not required to submit a

statement of interest. In addition, it is preferred (but not required) that educators have previous experience reviewing items, tasks, and/or stimuli.

During the committee reviews, educators specifically compare the items against the Quality Criteria for accessibility and for Bias and Sensitivity. The reviewers identify and resolve or reject any item, stimulus, or performance task that does not pass the criteria. This review removes any aspect that may negatively impact a student's ability to access stimuli, items, or performance tasks, or to elicit valid evidence about an assessment target. Items flagged for accessibility, bias/sensitivity, and/or content concerns are either revised to address the issues identified by the panelists or removed from the item pool.

The committee also compares each stimulus, item, and performance task against the ELA/literacy or mathematics quality criteria. This review focuses on developmental appropriateness and alignment of stimuli, items, and performance tasks to the content specifications and appropriate depths of knowledge. Panelists in the content review also check the accuracy of the content, answer keys, and scoring materials. Items flagged for content concerns are either revised or removed from the item pool.

Details about the item development process in ELA/literacy and mathematics are found in Appendix A. These are the steps each item goes through before it can be presented to students.

Field Testing

After items pass the content and accessibility, bias and sensitivity reviews, they become eligible for field testing. Details of the 2014 field test can be found in Chapters 7, 8 and 9 of the 2014 Technical Manual. Briefly, the field test was a comprehensive test that both established subject matter scales and provided item statistics and parameters.

Embedded field testing is conducted in each operational administration beginning with the 2014-2015 assessment. All operational items in the 2016-2017 summative assessment that weren't field tested in 2014, came from the 2014-2015 embedded field test (2015 EFT). All of the 2015 EFT items were CAT items.

CAT EFT items are administered randomly in positions 5 through N-5, where N is the number of items required in the blueprint. Two items are embedded in the mathematics test and up to four items are embedded in the ELA/literacy test. Passage-based items, or items associated with a common stimulus are embedded in sets. Stimulus-based sets consist of three or four items. A stimulus-based set and/or individual EFT items totaling four items are selected randomly at the beginning of each student's test and the positions of the set and/or individual items are selected randomly. EFT items are thus exposed to random samples of the student population.

Item Scoring

For those items that cannot be machine scored, the Consortium engages content experts in range finding activities. Range finding improves the consistency and validity of scoring for the assessment. During range finding, educators focus on the performance tasks for mathematics and ELA/literacy. The participants review student responses against item rubrics, validate the rubrics' accuracy, and select the anchor papers that would be used by scorers during operational scoring of test items. In mathematics, educators also review constructed response items for grades 7, 8, and high school.

Following the 2013 pilot test, 102 participants from 20 states were engaged in range finding. After the spring 2014 field test, 104 educators participated in range finding. After the 2014-2015 embedded field test 34 educators participated in range finding.

The educator qualifications for range finding are the same as the educator qualifications for item writing. It is preferred (but not required) that educators have previous range finding experience.

To verify correct scoring for machine-scored items, a rubric validation activity is conducted. For multiple choice items, this is a simple key check. For other item types, such as grid interaction items (drag-and-drop), matching tables or equation entry, the procedure involves looking at a sample of student raw responses (screen coordinates or keystrokes) and assuring that the raw response was scored correctly. In the course of this process, reviewers may find unexpected responses that require adjustment of the scoring procedure to account for a wider response range. Item scoring software is then changed accordingly.

Data Reviews

The consortium analyzes field test data to determine the statistical quality of the items and to guide further item processing. In a process called data review, educators review items that have been identified on the basis of statistical criteria. Statistical criteria that flag items for data review are shown in Table 4-1.

In data review, the educators look for possible content flaws, bias, and other features that might explain the statistical qualities of the items and make recommendations for revising, accepting, or rejecting the items. Content experts on the staff of Smarter Balanced and its contractors reviewed all items along with the educators recommendations and make final decisions about the items. The educator qualifications for participating in the data reviews are the same as the educator qualifications for item writing except that participants were not required to submit a statement of interest.

TABLE 4-2 ITEM FLAGGING BASED ON STATISTICAL CRITERIA

Flag	Definition
A	High difficulty (*p*-value less than 0.10)
B	Polytomous items with percentage obtaining any score category less than three percent of total N
C	Polytomous items with higher criterion score mean for students in a lower score-point category
D	Selected response items with proportionally more high-proficient students selecting a distractor over the key
F	Selected response items with higher criterion score mean for students choosing a distractor than the mean for those choosing the key
H	Low difficulty (*p*-value greater than 0.95)
P	Selected response items with positive distractor point-biserial correlation
R	Low item-total correlation (point-biserial correlation less than 0.30)
DIF	Differential item functioning
IR	Hand-scored items that do not meet criteria for inter-rater agreement

Items with no statistical flags are eligible for use in the operational pools. Flagged items move into operational pools if they are not rejected or revised in data review. Not operational items approved for operational use are used in summative test pools. Other uses for operational items include achievement level setting, interim assessments, practice tests, and released items. Items for these uses may also be drawn from the pool of items flagged for, or revised in, data review.

Table 4-3 shows how items that were field tested in the 2014 stand-alone field test fell out among various categories of item quality and use. A total of 9,383 ELA/Literacy and 10,052 Mathematics items were field tested. Several reasons are given for not using these items for summative assessments along with the counts of items for each reason. Ultimately, a total of 5,310 ELA/literacy and 7,028 Mathematics items were used in the 2014-2015 summative assessment.

TABLE 4-3 OUTCOMES OF ITEMS IN THE 2014 STAND-ALONE FIELD TEST

Content Area	Grade	Initial Field Test Pool	Reasons for Not Including in Summative Pool				14 15 Summative Pool
			Content Issues	Small Sample Size	Poor Item Statistics	Interim/ Other uses	
English Language Arts/Literacy	3	1,045	30	31	88	242	654
	4	965	17	32	60	203	653
	5	975	23	45	84	178	645
	6	984	23	30	82	236	613
	7	1,033	27	31	100	286	589
	8	1,010	20	40	114	242	594
	HS	3,371	61	658	281	809	1,562
	Total	**9383**	**201**	**867**	**809**	**2196**	**5310**
Mathe-matics	3	1,163	1	0	48	157	957
	4	1207	9	0	68	198	932
	5	1108	2	0	63	130	913
	6	1115	8	0	89	164	854
	7	1,037	5	0	90	175	767
	8	1,036	9	0	133	159	735
	HS	3,386	75	797	488	156	1,870
	Total	**10052**	**109**	**797**	**979**	**1139**	**7028**

Table 4-4 shows the dispensation of items that were embedded as field test items in the CAT segment of the 2014-2015 summative assessment (2015 EFTs). Over 10,000 items were field tested. Approximately one-third of the field tested items in each subject area were flagged for one or more of the reasons given in Table 4-2, plus insufficient data for a small number of items. Approximately 20% of the flagged items were rejected out of hand due to insufficient data or values of key statistics listed in Table 4-2 that were considered too poor for use even if they were approved for use in the data review process. The remainder of the flagged items were submitted for data review. Approximately 62% of the ELA/literacy items and 71% of the mathematics items sent through data review were approved for operational use.

The numbers of items ultimately approved for operational use are shown by grade and subject in the last column of Table 4-4. These included the not-flagged items and the number of items accepted in data review.

TABLE 4-4 OUTCOMES OF ITEMS IN THE 2014-2015 EMBEDDED FIELD TEST (2015 EFT)

Content Area	Grade	2015 EFT Items	Not Flagged	Flagged				Approved for Summative Pool
				Total Flagged	Rejected	Data Review		
						Revise/Reject	Accept	
English Language Arts/Literacy	3	654	475	179	21	60	98	573
	4	632	456	176	26	49	101	557
	5	643	455	188	29	51	108	563
	6	633	394	239	36	75	128	522
	7	629	390	239	46	80	113	503
	8	607	395	212	51	73	88	483
	HS	2,133	1397	736	162	225	349	1746
	Total	**5,931**	**3,962**	**1,969**	**371**	**613**	**985**	**4,947**
Mathe-matics	3	564	431	133	15	32	86	517
	4	659	543	116	17	28	71	614
	5	616	486	130	19	25	86	572
	6	674	453	221	47	26	148	601
	7	684	377	307	77	64	166	543
	8	691	379	312	76	67	169	548
	HS	923	460	463	141	127	195	655
	Total	**4,811**	**3,129**	**1,682**	**392**	**369**	**921**	**4,050**

Summative Item Pool

This section describes the 2016-2017 summative item pool. This pool included for the first time a large number of items that came from embedded field testing – the 2015 embedded field test (2015 EFT). All of the 2015 EFT-sourced items were CAT items. Table 4-5 shows the number and characteristics of the items in the summative pool by source and segment. Segments are CAT or PT (performance task). Sources are the field testing event, which is either the 2014 field test or the 2015 EFT. Approximately 40% of the CAT ELA/literacy items (2,965) and 28% of the CAT mathematics items (2,569) came from the 2015 EFT.

In both subjects, the 2015 EFT-sourced items were similar in discrimination (a-parameter) to the 2014 FT-sourced items, but somewhat easier as shown by their lower b-parameter. The fact that they were slightly easier was intentional. The 2014 FT was more difficult than desired in comparison to student achievement. Staff therefore chose somewhat easier items for operational use from the available (not flagged) 2015 EFT pool.

The PT segment of the 16-17 pool was comprised exclusively of items from the 2014 FT and was therefore virtually the same as in previous operational assessments. Items in the PT segment are grouped into distinct sets of items that are delivered intact to students in a randomized fashion. The number of item sets per grade is shown in table 4-10.

The addition of relatively large numbers of somewhat easier CAT items did not appreciably alter the psychometric characteristics of the tests that students received. Taking items from both sources together (2014 FT and 2015 EFT), the overall difficulty of the item pool did not change appreciably. (See tables 4-11 and 4-12.) Also, by its very nature, computer adaptive testing tends to deliver tests that are targeted on student performance regardless of differences in student performance and regardless of differences in the overall difficulty of the item pool. A larger number of easier items in the pool would not appreciably improve measurement precision at lower achievement levels compared to previous years because there were already enough items in the pool at lower achievement levels in previous years. A larger number of items may improve (reduced) item exposure, however. Chapters 3 and 5 contains more specific information about the psychometric characteristics of the tests students received.

TABLE 4-5 COMPOSITION OF 2016-2017 SUMMATIVE ITEM POOL BY SEGMENT AND SOURCE

Subject	Grade	Source of CAT						Source of PT		
		2014 FT			2015 EFT			2014 FT		
		N	a-parm	b-parm	N	a-parm	b-parm	N	a-parm	b-parm
English Language Arts/Literacy	3	522	0.66	-0.45	357	0.70	-0.63	62	0.71	0.21
	4	493	0.59	0.12	342	0.62	-0.05	81	0.64	0.45
	5	473	0.61	0.52	355	0.62	0.27	93	0.71	0.74
	6	477	0.54	1.03	292	0.60	0.64	61	0.87	0.92
	7	444	0.53	1.14	255	0.57	1.07	79	0.81	1.15
	8	448	0.53	1.27	303	0.56	1.08	88	0.72	1.26
	11	1374	0.50	1.73	1061	0.51	1.56	104	0.59	1.84
	Total/Avg:	**4231**	**0.55**	**0.95**	**2965**	**0.58**	**0.78**	**568**	**0.71**	**1.00**
Mathematics	3	808	0.84	-0.83	384	0.83	-0.98	80	0.89	-0.52
	4	807	0.82	-0.08	472	0.83	-0.31	95	0.85	-0.03
	5	776	0.78	0.69	434	0.76	0.20	85	0.76	1.01
	6	711	0.70	1.08	357	0.69	0.76	72	0.73	0.80
	7	651	0.71	1.79	292	0.70	1.46	87	0.89	1.58
	8	584	0.60	2.35	259	0.55	1.34	58	0.88	1.81
	11	2251	0.55	2.54	371	0.65	2.01	61	0.66	2.67
	Total/Avg:	**6588**	**0.68**	**1.34**	**2569**	**0.73**	**0.53**	**538**	**0.81**	**0.94**

The number of items per claim is shown for each subject area in table 4-6. All claims in both subjects were represented by sufficient numbers of items.

TABLE 4-6 COMPOSITION OF SUMMATIVE ITEM POOLS

	CLAIMS				
GRADE	1	2	3	4	Total
ELA/Literacy					
3	325	274	184	161	944
4	265	281	192	184	922
5	303	276	163	184	926
6	241	256	161	179	837
7	234	248	170	138	790
8	225	281	186	158	850
11	859	708	560	427	2554
Total	2452	2324	1616	1431	7823
Mathematics					
3	785	127	224	138	1274
4	839	151	236	152	1378
5	778	119	229	172	1298
6	745	107	185	103	1140
7	642	114	157	120	1033
8	597	72	148	84	901
11	1839	184	453	208	2684
Total	6225	874	1632	977	9708

The Consortium develops many different types of items beyond the traditional multiple-choice item. This is done to measure claims and assessment targets with varying degrees of complexity by allowing students to respond in a variety of ways rather than simply recognizing a correct response. These different item types are listed in Table 4-7. The frequency of item types by claim within grade and subject is shown in Table 4-8 and Table 4-9. Note that each Essay written is associated with two items. Essays are scored on three traits, two of which are combined, resulting in two items for each essay.

TABLE 4-7 ITEM TYPES FOUND IN THE SUMMATIVE ITEM POOLS

Item Types	ELA/literacy	Mathematics
Multiple Choice (MC)	X	X
Multiple Select (MS)	X	X
Evidence-Based Selected Response (EBSR)	X	
Match Interaction (MI)	X	X
Hot Text (HTQ)	X	
Short Answer Text Response (SA)	X	X
Essay/Writing Extended Response (WER)	X	
Equation Response (EQ)		X
Grid Item Response (GI)		X
Table Interaction (TI)		X

TABLE 4-8 DISTRIBUTION OF ELA/LITERACY ITEM TYPES BY GRADE AND CLAIM

Grade	Claim	Item Type							Total
		EBSR	HTQ	MC	MI	MS	SA	WER	
ELA/Literacy									
3	1	48	57	164	0	43	13	0	325
3	2	0	58	114	0	55	19	14	260
3	3	47	0	76	20	41	0	0	184
3	4	0	22	60	12	41	26	0	161
3	Total	95	137	414	32	180	58	14	930
4	1	48	46	102	0	48	21	0	265
4	2	0	55	121	0	46	23	18	263
4	3	48	0	84	21	39	0	0	192
4	4	0	21	65	15	48	35	0	184
4	Total	96	122	372	36	181	79	18	904
5	1	58	48	110	0	58	29	0	303
5	2	0	46	107	0	61	22	20	256
5	3	44	0	70	19	30	0	0	163
5	4	0	27	55	12	52	38	0	184
5	Total	102	121	342	31	201	89	20	906
6	1	42	60	75	0	48	16	0	241
6	2	0	52	90	0	64	22	14	242
6	3	44	0	75	20	22	0	0	161
6	4	0	16	62	11	62	28	0	179
6	Total	86	128	302	31	196	66	14	823
7	1	34	50	86	0	45	19	0	234
7	2	0	48	81	0	59	22	19	229
7	3	49	0	74	15	32	0	0	170
7	4	0	34	25	11	31	37	0	138
7	Total	83	132	266	26	167	78	19	771
8	1	42	48	70	0	46	19	0	225
8	2	0	47	94	0	75	25	20	261
8	3	25	0	118	6	37	0	0	186
8	4	0	34	39	14	32	39	0	158
8	Total	67	129	321	20	190	83	20	830
11	1	161	175	232	0	209	82	0	859
11	2	0	162	216	0	235	47	24	684
11	3	104	0	303	19	134	0	0	560
11	4	0	90	160	29	101	47	0	427
11	Total	265	427	911	48	679	176	24	2530
All	Total	794	1196	2928	224	1794	629	129	7694

TABLE 4-9 DISTRIBUTION OF MATHEMATICS ITEM TYPES BY GRADE AND CLAIM

Grade	Claim	Item Type							Total
		EQ	GI	MC	MI	MS	SA	TI	
Mathematics									
3	1	486	64	120	79	3	0	33	785
3	2	75	21	13	6	7	5	0	127
3	3	13	66	71	22	29	23	0	224
3	4	50	20	30	6	12	8	12	138
3	Total	624	171	234	113	51	36	45	1274
4	1	457	83	108	179	0	0	12	839
4	2	92	14	31	7	4	1	2	151
4	3	25	81	59	11	31	28	1	236
4	4	39	21	56	4	8	16	8	152
4	Total	613	199	254	201	43	45	23	1378
5	1	427	47	219	84	1	0	0	778
5	2	85	13	10	2	3	0	6	119
5	3	21	70	72	14	21	28	3	229
5	4	69	37	24	5	5	16	16	172
5	Total	602	167	325	105	30	44	25	1298
6	1	362	71	66	94	133	0	19	745
6	2	69	15	5	3	8	2	5	107
6	3	20	49	42	20	31	23	0	185
6	4	53	13	8	2	6	9	12	103
6	Total	504	148	121	119	178	34	36	1140
7	1	359	39	56	68	120	0	0	642
7	2	81	7	7	6	11	0	2	114
7	3	23	43	30	11	20	30	0	157
7	4	65	26	15	2	8	1	3	120
7	Total	528	115	108	87	159	31	5	1033
8	1	251	44	150	69	70	0	13	597
8	2	42	12	4	4	2	0	8	72
8	3	15	51	19	17	21	25	0	148
8	4	30	20	11	5	6	8	4	84
8	Total	338	127	184	95	99	33	25	901
11	1	667	316	439	299	111	0	7	1839
11	2	87	40	27	11	13	1	5	184
11	3	49	146	137	55	32	33	1	453
11	4	92	29	47	15	12	8	5	208
11	Total	895	531	650	380	168	42	18	2684
All	Total	4104	1458	1876	1100	728	265	177	9708

Each grade's item pool is large enough to support the summative blueprint. Unlike a traditional paper-and-pencil test where all students take the same items, students taking the CAT take items and tasks targeted to their ability level. This means that the Consortium needs to develop a large number of items in order to deliver tests that simultaneously meet the blueprint and are at a level of difficulty that is tailored to the performance of each student.

In addition to the items for the CAT, the Consortium also developed performance tasks. All students take performance tasks designed to measure a student's ability to integrate knowledge and skills across multiple claims and assessment targets. Each ELA/literacy performance task has a set of related stimuli presented with two or three research items and an essay. Each Mathematics performance task has 4 to 6 items relating to a central problem or stimulus. The PT items are organized into distinct sets that are delivered intact to students. The number of PT item sets per grade and subject in the 16-17 summative assessment is shown in table 4-10. The sets are delivered in randomized fashion to students rather than adaptively.

TABLE 4-10 NUMBER OF PERFORMANCE TASKS BY GRADE AND SUBJECT.

Grade	ELA/literacy	Mathematics
3	14	14
4	18	19
5	20	15
6	14	12
7	19	16
8	20	12
11	24	12

The distribution of item parameters by grade and claim are shown in tables 4-11 (ELA/literacy) and 4-12 (mathematics). Note that there is a wide range of difficulty in each category. This enables the algorithm (described previously in this chapter) to find the best items for each student. As such, adaptive tests provide more precise measurement for all levels of student performance than would be provided with a fixed form test of the same length. This is accomplished through having a bank of previously calibrated items to deliver during the adaptive portion of the test. In addition, fixed, randomly-assigned performance tasks add information to student performance.

TABLE 4-11 ITEM DIFFICULTY (B-PARAMETER) AND DISCRIMINATION (A-PARAMETER), ELA/LITERACY.

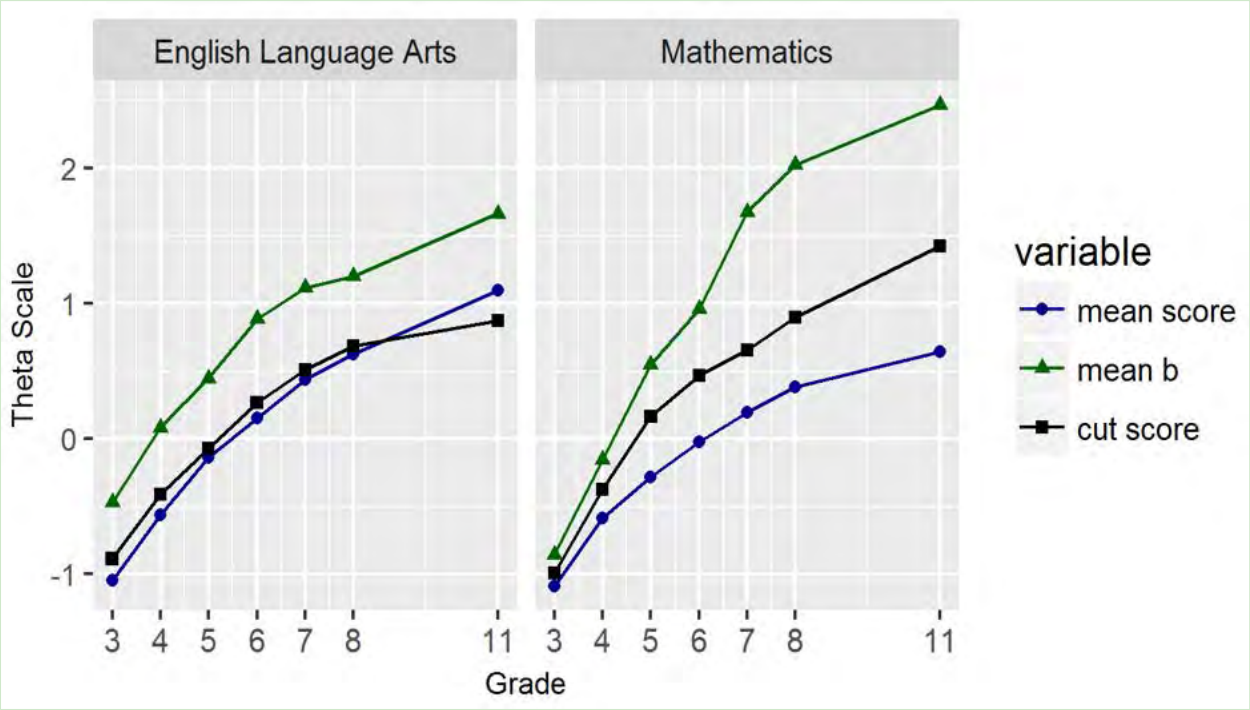
Grade	Claim	# of Items	b-parameter			a-parameter
			Mean	Min	Max	Mean
ELA/LITERACY						
3	1	325	-0.561	-2.725	2.743	0.72
	2	274	-0.775	-2.896	2.810	0.70
	3	184	-0.178	-2.920	3.816	0.55
	4	161	-0.144	-2.216	3.032	0.70
	Total	944	-0.474	-2.920	3.816	0.68
4	1	265	0.296	-2.529	3.133	0.64
	2	281	-0.348	-3.252	2.935	0.61
	3	192	0.035	-2.822	4.254	0.56
	4	184	0.464	-1.996	3.727	0.58
	Total	922	0.085	-3.252	4.254	0.60
5	1	303	0.655	-1.784	4.806	0.63
	2	276	0.058	-2.278	3.294	0.62
	3	163	0.477	-2.403	3.481	0.53
	4	184	0.675	-1.494	3.832	0.67
	Total	926	0.457	-2.403	4.806	0.62
6	1	241	0.983	-1.254	4.779	0.60
	2	256	0.779	-2.719	4.607	0.57
	3	161	0.808	-1.497	4.921	0.51
	4	179	1.029	-1.305	3.609	0.61
	Total	837	0.898	-2.719	4.921	0.58
7	1	234	1.136	-1.877	3.914	0.59
	2	248	1.038	-2.019	5.124	0.57
	3	170	0.809	-1.706	4.775	0.51
	4	138	1.726	-0.815	5.525	0.59
	Total	790	1.131	-2.019	5.525	0.56
8	1	225	1.414	-1.170	5.572	0.60
	2	281	1.090	-3.014	4.558	0.54
	3	186	0.910	-2.119	3.871	0.49
	4	158	1.554	-1.788	5.188	0.60
	Total	850	1.221	-3.014	5.572	0.55
11	1	859	1.784	-2.087	5.800	0.56
	2	708	1.662	-1.880	5.929	0.48
	3	560	1.313	-1.648	5.618	0.46
	4	427	1.923	-1.197	5.124	0.50
	Total	2554	1.669	-2.087	5.929	0.51

TABLE 4-12 ITEM DIFFICULTY (B-PARAMETER) AND DISCRIMINATION (A-PARAMETER), MATHEMATICS

Grade	Claim	# of Items	b-parameter			a-parameter
			Mean	Min	Max	Mean
Mathematics						
3	1	785	-1.200	-3.567	2.402	0.84
	2	127	-0.446	-2.537	1.967	0.99
	3	224	-0.288	-2.424	3.464	0.77
	4	138	-0.205	-2.677	2.298	0.81
	Total	1274	-0.857	-3.567	3.464	0.84
4	1	839	-0.348	-3.260	4.113	0.85
	2	151	-0.019	-2.248	2.574	0.89
	3	236	0.219	-2.014	3.157	0.77
	4	152	0.204	-2.148	2.219	0.70
	Total	1378	-0.154	-3.260	4.113	0.82
5	1	778	0.286	-2.791	3.606	0.77
	2	119	0.879	-1.267	3.409	0.94
	3	229	0.794	-1.903	5.278	0.70
	4	172	1.173	-1.232	4.452	0.73
	Total	1298	0.546	-2.791	5.278	0.77
6	1	745	0.759	-3.934	4.348	0.69
	2	107	1.097	-2.978	5.099	0.79
	3	185	1.418	-1.904	4.709	0.61
	4	103	1.444	-0.715	3.894	0.80
	Total	1140	0.960	-3.934	5.099	0.70
7	1	642	1.569	-1.792	5.643	0.72
	2	114	1.639	-1.085	5.071	0.84
	3	157	1.921	-1.655	6.174	0.64
	4	120	1.983	-0.881	4.373	0.76
	Total	1033	1.678	-1.792	6.174	0.72
8	1	597	1.822	-1.868	6.321	0.60
	2	72	2.632	0.046	5.751	0.77
	3	148	2.487	-0.878	6.698	0.53
	4	84	2.159	-1.364	5.354	0.66
	Total	901	2.027	-1.868	6.698	0.61
11	1	1839	2.240	-4.432	7.297	0.60
	2	184	2.921	-1.101	6.680	0.61
	3	453	2.937	-1.636	7.194	0.46
	4	208	3.098	-0.069	6.379	0.53
	Total	2684	2.471	-4.432	7.297	0.57

Although there is a wide distribution of item difficulty, pools tend to be difficult in relation to the population and to the cut score that is typically associated with proficiency (level 3 cut score). The chart below shows mean item difficulty, level 3 cut score, and mean student achievement scores (all in theta units) by grade and subject. The mean item difficulty and student achievement plotted in this figure is based on the 2016-2017 assessment.

FIGURE 4-6 COMPARISON OF ITEM DIFFICULTY, MEAN, STUDENT SCORES, CUT SCORES FOR ELA/LITERACY



Content Alignment

Content alignment addresses how well individual test items, test blueprints, and the tests themselves represent the intended construct and support appropriate inferences. With a computer adaptive test, a student's test form is a sampling of items drawn from a much larger universe of possible items and tasks. The sampling is guided by a blueprint. Alignment studies investigate how well individual tests cover the intended breadth and depth of the underlying content standards. For inferences from test results to be justifiable, the sample of items in each student's test has to be an adequate representation of the broad domain, providing strong evidence to support claims being made from the test results.

Four alignment studies have been conducted to examine the alignment between Smarter Balanced tests and the CCSS. The Human Resources Research Organization (HumRRO) conducted the first alignment study. HumRRO's comprehensive study (HumRRO, 2015), centered around the assumptions of evidence centered design (ECD), which examined the connections in the evidentiary chain underlying the development of the Smarter Balanced foundational documents (test blueprints, content specifications, and item/task specifications) and the resulting summative assessments. Among those connections were the alignment between the Smarter Balanced content specifications, the alignment between the Smarter Balanced evidence statements and content specifications, and

the alignment between the Smarter Balanced blueprint and the Smarter Balanced content specifications. Results from this study were favorable in terms of the intended breadth and depth of the alignment for each connection in the evidentiary chain.

In 2016, the Fordham Institute and HumRRO investigated the quality of the Smarter Balanced assessments relative to CCSSO criteria for evaluating high-equality assessments. In particular, the Smarter Balanced assessments were investigated to see if they place strong emphasis on the most important content for college and career readiness as called for by the CCSS and if they require that students demonstrate the range of thinking skills, including higher-order thinking skills, called for by those standards. Fordham Institute reviewed Grades 5 and 8 ELA and mathematics, and HumRRO reviewed high school ELA and mathematics.

- Fordham Institute (2017) rated Smarter Balanced Grades 5 and 8 ELA assessments an Excellent Match to the CCSSO Criteria for Content in ELA, and a Good Match for Depth in ELA
- Fordham Institute rated Smarter Balanced Grades 5 and 8 mathematics assessments as a Good Match to the CCSSO Criteria for Content in Mathematics, and a Good Match to the CCSSO Criteria for Depth in Mathematics
- HumRRO (2017) rated the Smarter Balanced high school ELA assessments an Excellent Match to the CCSSO Criteria for Content in ELA, and a Good to Excellent Match for Depth in ELA
- HumRRO (2017) rated the Smarter Balanced high school ELA assessments a Good to Excellent Match to the CCSSO Criteria for Content in ELA, and a Good to Excellent Match for Depth in ELA

An additional external alignment study, completed by WestEd (2017), employed a modified Webb alignment methodology to examine the summative assessments for grades 3, 4, 6, and 7, using sample test events built using 2015-2016 operational data. The WestEd alignment study provided evidence that the items within ELA/L and mathematics test events for grades 3, 4, 6, and 7 were well aligned to the CCSS in terms of both content and cognitive complexity.

Summary of Test Design

The intent of this chapter is to show how the assessment design supports the purposes of Smarter Balanced summative assessments. Content specifications were derived directly from the CCSS, expressing the standards as measurable elements and made explicit in Smarter Balanced claims and assessment targets structure. Building on these, test blueprints provide appropriate proportions of CCSS content coverage. Using the blueprints, item writers wrote items and tasks in quantities that supported CAT and performance task delivery. Expansion of item and task types promoted student responses that provide more insight into proficiency than that provided by multiple choice items alone. The use of performance tasks addresses the need to assess application and integration of skills. Finally, the method of delivery and test scoring, combining adaptive and non-adaptive elements, provides the most precise information and an enhanced student testing experience.

References

American Institutes for Research (2013). *Cognitive Laboratories Technical Report*. Los Angeles, CA: Smarter Balanced Assessment Consortium.

American Institutes for Research (2014). *Smarter Balanced scoring specification: 2014–2015 Administration*. Los Angeles, CA: Smarter Balanced Assessment Consortium.

Cohen, J. & Albright, L. (2014). *Smarter Balanced adaptive item selection algorithm design report*. Retrieved from http://www.smarterapp.org/specs/AdaptiveAlgorithm.html

Dana, T. M., & Tippins, D. J. (1993). Considering alternative assessment for middle level learners. *Middle School Journal, 25*, 3-5.

Fadel, C., Honey, M., & Pasnik, S. (2007, May). Assessment in the age of innovation. *Education Week*. May 18, 2007. Retrieved on July 2, 2012 from http://www.edweek.org/ew/articles/2007/05/23/38fadel.h26.html?print=1

Hansen, E. G. and Mislevy, R. J. (2008), Design patterns for improving accessibility for test takers with disabilities. *ETS Research Report Series*, 2008: i–32. doi: 10.1002/j.2333-8504.2008.tb02135.x

HumRRO. (2015). *Smarter Balanced Assessment Consortium Alignment Study Report*. Los Angeles, CA. Smarter Balanced Assessment Consortium.

Mislevy R., & Haertel G., (2006). Implications of evidence-centered design for educational testing. Menlo Park, CA: SRI International.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.

Rose, D., & Meyer, A. (2000). Universal design for learning, associate editor column. *Journal of Special Education Technology* 15 (1): 1-12.

Schmeiser, C. B., & Welch, C. J. (2006). Test Development. In R. L. Brennan (Ed.) *Educational Measurement*, 4th Edition (307-353). Washington, DC: American Council on Education.

Smarter Balanced Assessment Consortium (2010, June 23). *Race to the Top Assessment Program Application for New grants: Comprehensive Assessment Systems*, CFDA Number: 84.395B. OMB Control Number 1814-0699.

Smarter Balanced Assessment Consortium. (2014). *State procedures manual.* Los Angeles, CA: Smarter Balanced Assessment Consortium.

Smarter Balanced Assessment Consortium. (2015a). *ELA/Literacy summative assessment blueprint*. Retrieved from https://portal.smarterbalanced.org/library/en/english-language-artsliteracy-content-specifications.pdf

Smarter Balanced Assessment Consortium. (2015b). *Mathematics summative assessment blueprint*. Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

Smarter Balanced Assessment Consortium. (2015c). *Content specifications for the summative assessment of the common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA_Content_Specs.pdf

Smarter Balanced Assessment Consortium. (2015d). *Content specifications for the summative assessment of the common core state standards for mathematics*. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/Mathematics-Content-Specifications.pdf

Smarter Balanced Assessment Consortium. (2015e). *Item and task specifications*. Retrieved from http://www.smarterbalanced.org/assessments/development/

Smarter Balanced Assessment Consortium. (2015f). *Achievement level setting final report.* Retrieved from http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

Smarter Balanced Asssessment Consortium. (2016). 2013-2014 Technical Report. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/2013-14_Technical_Report.pdf

Zhang, T., Haertel, G., Javitz, H., Mislevy, R., Murray, E., & Wasson, J. (2009). A Design Pattern for a Spelling Bee Assessment for Students with Disabilities. A paper presented at the annual conference of the American Psychological Association, Montreal, Canada.

Chapter 5: Scores, Scales and Norms

Introduction

A test score is provided to stakeholders as a metric of student performance on a test. On the Smarter Balanced assessments, this test score along with the achievement levels helps students understand their progress towards career and college readiness. This chapter summarizes the processes that Smarter Balanced undertook to construct a psychometrically-sound test scale so that stakeholders would receive meaningful test scores. This chapter begins with an overview of the work done in the pilot and field tests to select a psychometric model, to construct the Smarter Balanced test scale, and to establish cut scores. Normative information from the 2017 operational administration is shared later in the chapter.

Item Response Theory

Unidimensional Item Response Theory (IRT) models were used to calibrate items and create Smarter Balanced test scale. The specific models were chosen following studies completed during the pilot and field test phases of the assessment. The IRT models for constructing and maintaining the measurement scale were selected on the basis of analyses completed during the pilot and field test stages of test development. Table 5-1 identifies and provides summaries of these analyses.

TABLE 5-1 ANALYSES COMPLETED DURING PILOT AND FIELD TEST TO SELECT CALIBRATION AND SCALING MODEL [4]

Phase	Analysis	Summary
Pilot	Dimensionality	Multidimensional IRT was used as a factor analytic approach to examine the dimensional structure of the assessments. The purpose of the study was to examine 1) the degree to which essential unidimensionality is met within a single grade and content area, and 2) the degree of invariance in the construct across two adjacent grades that contain unique grade specific items and common "vertical" linking items. Based on the results of the study, it was concluded that the data support the use of a unidimensional IRT model and a vertical scale.
	IRT Model Choice	Various unidimensional models were investigated: a Rasch one-parameter/partial credit model (Rasch/PC) combination; a two-parameter logistic/generalized partial credit model (2PL/GPC) combination; or a three-parameter logistic/generalized partial credit (3PL/GPC) combination. The choice of model was based on model simplicity, model fit, model stability, and reasonableness. Special investigations of guessing and discrimination were completed. In addition, estimates of student ability were compared across the three models. Based on the results of these studies and the considerations outlined above, Smarter Balanced chose to employ the 2PL/GPC model.

[4] Detailed information about the calibration and scaling process may be found in Chapters 6 through 9 in the *2013-2014 Technical Report* (Smarter Balanced, 2016).

Phase	Analysis	Summary
Field Test	Application of IRT Model	The IRT models selected during the pilot test phase were used for calibrating the items in the field test phase. The usefulness of IRT models is dependent on the extent to which they effectively reflect the data. Assessing fit in item response models usually involves validating assumptions underlying the models and evaluating goodness-of-fit, which specifically refers to how effectively the model describes the outcome data. IRT fit evaluation was conducted for calibrations using the 2PL/GPC combination. Item fit was evaluated in conjunction with other psychometric criteria. No items were excluded based solely on fit. The results of the evaluation supported the use of the 2PL/GPC model.
	Final Scale	Scaling for the vertical scaling sample was completed in two steps, one that linked tests horizontally within a grade level and content area, and a second that linked tests vertically to adjacent grade levels within a content area. The horizontal scale was created using a hybrid approach using both common items and randomly equivalent groups (implemented using LOFT administration). Smarter Balanced selected test characteristic curve transformation methods to construct the vertical scale, using grade 6 as the baseline and successively linking each grade level onto the scale. Items were calibrated using the IRT program PARSCALE. Following the construction of the vertical scale through the two-step procedure, the remainder of the item pool was calibrated to the vertical scale using the STUIRT software program.

Calibration and Scaling

Smarter Balanced utilizes the two-parameter logistic (2PL) model and the generalized partial credit model (GPCM) to calibrate selected-response and polytomous items, respectively. The 2PL model (Birnbaum, 1968) is given by

$$P_i(\theta_j) = \frac{\exp\left[Da_i(\theta_j - b_i)\right]}{\{1+\exp\left[Da_i(\theta_j - b_i)\right]\}},$$

Where $P_i(\theta_j)$ is the probability of a correct response to item *i* by a test taker with ability θ_j; a_i is the discrimination parameter; b_i is the difficulty parameter for item *i*; and *D* is a constant that puts the ability scale into the same metric as the normal ogive model (D=1.7).

For constructed-response items, the generalized partial credit model (GPCM; Muraki, 1992) or partial credit model (PCM; Masters, 1982) is employed. The generalized partial credit model is given by

$$P_{ih}(\theta_j) = \frac{\exp\sum_{v=1}^{h}\left[Da_i(\theta_j - b_i + d_{iv})\right]}{\sum_{c=1}^{n_i}\exp\left[\sum_{v=1}^{c}Da_i(\theta_j - b_i + d_{iv})\right]},$$

Where $P_{ih}(\theta_j)$ is the probability of examinee *j* obtaining a score of *h* on item *i*; n_i is the number of item score categories; b_i is the item location parameter; d_{iv} is the category parameter for item *i*, category *v*; and *D* is a scaling constant given previously.

PARSCALE (Muraki & Bock, 2003) was used for the IRT calibrations. PARSCALE is a multipurpose program that implements a variety of IRT models associated with mixed-item formats and associated statistics. The psychometric properties of PARSCALE are well known, and it can efficiently and

accurately calibrate large data sets such as those of Smarter Balanced assessments. The program implements marginal maximum likelihood (MML) estimation techniques for item parameters and MLE estimation of student proficiency (theta).

Vertical Scale

The IRT scaling for Smarter Balanced was performed in two steps. The first step was used to establish the horizontal and vertical scales that were used to set achievement levels. In the first step, items were initially scaled horizontally, where items in a single grade and content area were concurrently (i.e., simultaneously) calibrated. The vertical linking was accomplished using common items administered across grades (e.g., the same items given in 3rd and 4th grades) and then by placing consecutive grades onto the vertical scale. In the second step, the remaining, and much larger, item pool (containing non-common items, each administered only to one grade) were calibrated onto the vertical scale in an equating procedure that used the items from the first phase as linking/common items. For detailed description of the methods used in constructing the vertical scale, see Chapter 9 of the *2013-2014 Technical Report* (Smarter Balanced, 2016).

Transforming the Theta Metric to the Scale Score

The estimates of student proficiency and item difficulty from the calibration program, PARSCALE, are on a scale where student ability has a mean of zero and a standard deviation of approximately 1. This scale is called the theta-scale and a student's proficiency on this scale is referred to as the student's theta. Estimates of student proficiency are transformed onto a four-digit scale that is more meaningful to stakeholders. The equation for this transformation is:

$$Scale\ score = (theta * slope) + intercept$$

TABLE 5-2 SLOPE AND INTERCEPT FOR ELA/LITERACY AND MATHEMATICS

Subject	Grade	Slope	Intercept
ELA/literacy	3-8, HS	85.8	2508.2
Mathematics	3-8, HS	79.3	2514.9

Minimum and Maximum Scale Scores

A maximum likelihood procedure will not produce estimates of proficiency for students with extreme raw scores. An extreme raw score occurs when a student either gets full credit for all items taken (a perfect score) or gets no credit on any items taken (zero). Scale scores were established for these extreme cases following a non-maximum likelihood but logical procedure. These scale scores are called the Lowest Obtainable Scale Score (LOSS) and the Highest Obtainable Scale Score (HOSS).

The guidelines for establishing the LOSS and HOSS values were as follows.

1. The HOSS should be high enough so that it does not cause a disproportionate number of scale scores at the top of the scale. Likewise, the LOSS should be low enough so that it does not cause a disproportionate number of scale scores at the bottom part of the scale.
2. The HOSS should be low enough so that CSEM*HOSS < 10*Minimum(CSEMs for all scale scores), where CSEM is the conditional standard error of measurement. The LOSS should be high enough so that CSEM*LOSS < 15*Minimum(CSEMs for all scale scores).
3. The HOSS and LOSS values should increase and transition smoothly with increasing grade level.

Table 5-3 provides the Smarter Balanced LOSS and HOSS values by grade and content area. The LOSS and HOSS values give the effective range of the ELA/literacy and mathematics scales. The ELA/literacy scale ranges from a value of 2114, which is the LOSS for grade 3, to the HOSS of 2795 for high school. In mathematics, the range was from 2189 to 2862.

TABLE 5-3 LOWEST OBTAINABLE SCALE SCORES (LOSS) AND HIGHEST OBTAINABLE SCALE SCORES (HOSS) BY GRADE AND CONTENT AREA

Grade	LOSS	HOSS
ELA/literacy		
3	2114	2623
4	2131	2663
5	2201	2701
6	2210	2724
7	2258	2745
8	2288	2769
HS	2299	2795
Mathematics		
3	2189	2621
4	2204	2659
5	2219	2700
6	2235	2748
7	2250	2778
8	2265	2802
HS	2280	2862

Achievement-Level Setting

The Consortium used a multi-step achievement level setting process to establish the cut scores that separate students into achievement levels in ELA/literacy and mathematics across grades 3 through 8 and 11 (Smarter Balanced, 2015). Achievement level setting is also referred to as standard setting.

Pre-Step: Development of the Achievement Level Descriptors

Smarter Balanced developed an interconnected system of achievement levels for English language arts/literacy (ELA/literacy) and mathematics (CTB/McGraw Hill, 2014). The achievement levels were developed through a series of workshops and review cycles that allowed participation from a variety of Smarter Balanced stakeholders. At the outset of the achievement level setting process, Smarter Balanced established threshold achievement level descriptions. A threshold achievement level description includes knowledge and skills that students at the lower borderline of a given achievement level are expected to have. These descriptions guided the work of the achievement level setting process. The threshold descriptions are aligned to the Smarter Balanced content specifications and the Common Core State Standards.

Step 1: Distributed Standard Setting (Online Panel)

Smarter Balanced selected the bookmark standard setting procedure (Lewis, Mitzel, Mercado, & Schulz, 2012) to set achievement levels. The bookmark method is appropriate for assessments with a mixture of item types. The Smarter Balanced assessments are calibrated and scaled using item-response theory models. The bookmark method utilizes IRT models to create the item maps that underlie the procedure. The psychometric foundation of the bookmark procedure is well documented (e.g., Cizek & Bunch, 2007). Its usefulness has been well established through adoption of cut scores produced by bookmark-based standard-setting activities for many statewide educational achievement assessments.

The bookmark method relies on presenting panelists with sets of test items sorted by difficulty and representing test content, called an ordered item booklet (OIB). The OIBs were constructed to match Smarter Balanced guidelines with respect to targets and claims used to inform item and test development. In addition, some of the items in the OIBs for grades 4, 8, and 11 were from other tests such as the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA). These items were embedded in the spring 2014 field test to provide panelists with an external reference range for comparison to the performance of students on other tests.

In order to maximize participation, the Consortium invited educators, parents, and other concerned citizens from member states to participate in an online achievement level setting using the bookmark standard setting procedure. During the online session, thousands of teachers and other interested parties independently reviewed test questions and recommended the level of performance required for students to be considered on-track toward college and career readiness. In other words, the online panel only made recommendations in regard to Level 3.

The concept of an online panel is an innovation introduced to address the scale of the Smarter Balanced project and its number and variety of stakeholders. In addition to allowing wider achievement level setting participation, the online panel approach promotes deeper understanding of the content standards and Smarter Balanced assessments. The cut scores recommended by the online panels were presented during the in-person workshop.

Step 2: In-Person Panel

The in-person panel allowed teams of educators and other stakeholders nominated by member states to deliberate and recommend cut scores for all four achievement levels: Level 1, Level 2, Level 3, and Level 4. The in-person panel engaged in the bookmark standard setting procedure using the same ordered item booklets (OIBs) reviewed by the online panel. Separate grade-level panels for each

content area, consisting of approximately 30 members each, were convened to recommend cut scores for ELA/literacy and mathematics. Member states nominated all panelists, which included teachers and administrators, higher education faculty, business and community leaders, and parents

The in-person panels participated in three rounds of discussion and bookmark placements. In Round 1, panelists studied the items in the OIB and made recommendations. In Round 2, panelists were shown the cut scores from the online standard setting, engaged in small group discussions, and made recommendations. In Round 3, panelists were shown various forms of impact data (percentage of students in each achievement level), engaged in large group discussions, and made recommendations

For the in-person achievement level setting, the process was field tested and revised based on field test evidence. In addition, panelists were asked to provide feedback on their experience with the activities used to set cut scores. The vast majority of panelists (over 90%) agreed that the activities of the workshop (e.g., training and practice exercises, taking the practice test, engaging in discussions) were useful to their understanding of the process.

Step 3: Cross-Grade Review (Vertical Articulation Committee)

Following the in-person achievement level setting, a subset consisting primarily of educators from the in-person panels met to review the achievement levels recommended during the in-person achievement level setting (Step 2). Separate cross-grade panels were convened for ELA/literacy and for mathematics. The purpose of the cross-grade review was to ensure that the achievement levels were appropriately aligned across grades and would accurately reflect student progress from year to year. The panelists at the cross-grade review examined the earlier recommendations and suggested changes that would improve cross-grade alignment of the achievement level. For the vertical articulation panel, the process was field tested and revised based on field test evidence.

Step 4: Member Approval

The final recommendations were reviewed, adjusted, and then endorsed by the member states. Member states were not required to adopt the achievement levels. Higher education leaders participated in the decisions regarding grade 11 achievement levels to ensure they reflect the expectations of colleges and universities. The Consortium's Technical Advisory Committee, a special advisory committee on achievement level setting, and an expert auditor (Dr. Gregory Cizek) certified that the multi-step process was appropriately implemented. The achievement levels were then subject to existing approval processes within individual states and territories. The final cut scores are reported in Table 5-4.

TABLE 5-4 CUT SCORES FOR ENGLISH LANGUAGE ARTS/LITERACY AND MATHEMATICS

	Grade	SS Cut between Levels 1 and 2	SS Cut between Levels 2 and 3	SS Cut between Levels 3 and 4
ELA	3	2367	2432	2490
	4	2416	2473	2533
	5	2442	2502	2582
	6	2457	2531	2618
	7	2479	2552	2649
	8	2487	2567	2668
	HS	2493	2583	2682
Math	3	2381	2436	2501
	4	2411	2485	2549
	5	2455	2528	2579
	6	2473	2552	2610
	7	2484	2567	2635
	8	2504	2586	2653
	HS	2543	2628	2718

Step 5: Interpolating High School Cut Points

Smarter Balanced held a Cut Score Validation workshop in November 2017 to establish cut scores for grades 9 and 10 in ELA/literacy and mathematics. Smarter Balanced established preliminary cut scores for grade 9 and 10 by interpolating from the existing Grade 8 and 11 cut scores.

These interpolated cut scores were reviewed by 11 English language arts/literacy educators and 10 mathematics educators nominated by Smarter Balanced states. Panelists were instructed that the grade 8 and grade 11 cut scores could not be changed, and that the cut scores across grade levels must reflect a logical and defensible vertical articulation.

The workshop methodology and materials were based on the Bookmark standard setting procedure that had been implemented in the earlier in-person panel meetings (step 2). Workshop panelists reviewed practice tests and performance tasks, OIBs, and grade 11 achievement level descriptors. The OIBs included the same Smarter Balanced items used in 2014, with the exception of items in the Mathematics OIB that were determined to be outside the scope of grade 9 and 10 coursework. (The Mathematics educator panel later reviewed these items, as well.) After reviewing the OIBs, panelists verified or adjusted the interpolated grade 9 and 10 cut scores.

As a final step in the process, panelists drafted achievement level descriptions based on the grade 9 and 10 cut scores. Starting with the grade 11 ALDs, panelists suggested specific words and/or phrases to describe the performance levels for grades 9 and 10.

Final validated cut scores for grades 9 and 10 are reported in Table 5-5.

TABLE 5-5 VALIDATED CUT SCORES FOR GRADES 9 AND 10 ENGLISH LANGUAGE ARTS/LITERACY AND MATHEMATICS

	Grade	SS Cut between Levels 1 and 2	SS Cut between Levels 2 and 3	SS Cut between Levels 3 and 4
ELA	9	2489	2571	2672
	10	2491	2577	2678
Math	9	2517	2601	2676
	10	2533	2614	2697

Results for the 2016-2017 Assessments

Results presented below are aggregated across only the Smarter Balanced members that submitted de-identified student results data for the 2016-2017 assessment[5]. The results in this section are intended only to inform and provide context for interpreting local test results and trends and should not be used for accountability and evaluation.

Overall Results

Student results are reported in two primary ways: scale scores and achievement levels. Students are provided with results for the overall test and for subtests based on claims. The scale score represents student achievement numerically in terms of more or less. The achievement levels and level-descriptions help to convey the meaning of the scores to stakeholders. Together, scale scores, claim scores, and achievement levels provide a comprehensive description of student achievement by content area, claim, and grade level. Table 5-6 describes the goals associated with overall student performance by subject area and grade (grade 11) or grade band (grades 3-8).

TABLE 5-6 SMARTER BALANCED OVERALL ASSESSMENT CLAIMS

	ELA/Literacy	Mathematics
Overall, Grades 3-8	Students can demonstrate progress toward college and career readiness in English language arts and literacy.	Students can demonstrate progress toward college and career readiness in mathematics.
Overall, Grade 11	Students can demonstrate college and career readiness in English language arts and literacy.	Students can demonstrate college and career readiness in mathematics.

Table 5-7 through table 5-20 present student results in terms of the average scale score and the percentage of students in each achievement level. Results are presented for the reporting members' students overall (total) and by demographic group.

[5] Data for aggregated results was provided by the following Consortium members: California, Delaware, Hawaii, Idaho, Michigan, Montana, Nevada, North Dakota, Oregon, South Dakota, Vermont, Virgin Islands and Washington.

TABLE 5-7 GRADE 3 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	810768	2418.3	90.4	30.3	24.3	21.9	23.5	45.4
Female	395810	2427.4	89.5	26.6	24.1	22.8	26.4	49.2
Male	414710	2409.6	90.4	33.8	24.5	21.1	20.7	41.7
American Indian or Alaska Native	9265	2372.5	81.4	49.7	26.1	15.4	8.8	24.2
Asian	55241	2470.2	89.5	13.9	17.1	23.5	45.5	68.9
Black/African American	57794	2376.5	83.1	48.7	25.2	15.6	10.5	26.0
Native Hawaiian or Pacific Islander	7761	2391.7	85.4	40.9	26.5	18.4	14.2	32.6
Hispanic/Latino Ethnicity	303884	2392.8	84.5	39.9	27.0	19.3	13.8	33.1
White	306695	2441.7	86.0	20.3	22.9	25.6	31.3	56.9
Two or More Races	54123	2423.7	91.7	28.4	23.7	22.1	25.8	47.9
Unidentified Race	16005	2445.4	88.1	19.6	21.5	25.1	33.7	58.9
LEP Status	174557	2367.0	76.5	52.2	27.5	13.8	6.4	20.2
IDEA Indicator	77684	2351.2	86.9	62.4	18.9	10.7	8.0	18.7
Section 504 Status	6700	2418.7	88.1	28.8	26.2	22.4	22.6	45.0
Economic Disadvantage Status	460971	2390.4	83.8	41.0	26.9	19.0	13.0	32.0

TABLE 5-8 GRADE 4 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	819115	2459.5	95.1	33.1	20.3	22.8	23.8	46.6
Female	400282	2469.7	93.3	28.9	20.4	23.9	26.8	50.7
Male	418580	2449.7	95.7	37.2	20.2	21.7	21.0	42.6
American Indian or Alaska Native	9375	2411.7	85.9	53.6	21.3	16.2	8.8	25.0
Asian	55684	2515.9	93.0	14.9	14.0	24.1	47.0	71.1
Black/African American	58209	2412.8	87.7	53.9	20.0	16.0	10.1	26.1
Native Hawaiian or Pacific Islander	7897	2429.4	90.9	44.8	21.4	19.7	14.2	33.8
Hispanic/Latino Ethnicity	306486	2432.5	89.2	43.1	22.6	20.5	13.8	34.3
White	312484	2483.9	89.8	22.8	19.4	26.0	31.7	57.8
Two or More Races	52990	2466.4	95.6	30.8	19.6	23.2	26.3	49.6
Unidentified Race	15990	2491.7	93.2	21.4	17.4	25.0	36.2	61.2
LEP Status	152909	2393.1	77.3	62.1	22.0	12.1	3.8	15.9
IDEA Indicator	82673	2379.5	92.1	68.5	14.6	10.0	6.9	16.9
Section 504 Status	8677	2459.8	91.5	32.4	21.8	23.7	22.1	45.8
Economic Disadvantage Status	463134	2429.3	88.5	44.8	22.4	20.0	12.9	32.9

TABLE 5-9 GRADE 5 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	829666	2496.3	98.5	29.7	20.8	28.7	20.8	49.5
Female	406518	2509.4	96.2	24.8	20.5	30.4	24.4	54.7
Male	422883	2483.6	99.1	34.4	21.1	27.1	17.4	44.5
American Indian or Alaska Native	9455	2443.1	92.4	50.9	21.5	20.3	7.4	27.6
Asian	57094	2554.4	97.8	13.7	13.4	28.9	43.9	72.8
Black/African American	58639	2450.5	91.4	47.9	22.6	21.1	8.5	29.6
Native Hawaiian or Pacific Islander	8098	2469.3	95.4	38.8	23.1	25.8	12.3	38.1
Hispanic/Latino Ethnicity	314001	2466.1	92.5	40.1	23.7	25.1	11.1	36.2
White	314343	2523.4	91.6	19.1	19.1	33.7	28.2	61.9
Two or More Races	51120	2505.6	98.3	26.6	20.0	30.0	23.4	53.4
Unidentified Race	16916	2530.2	95.4	18.2	16.8	32.2	32.7	64.9
LEP Status	136283	2415.7	77.9	63.5	23.0	11.7	1.9	13.5
IDEA Indicator	85087	2403.1	93.4	69.5	15.2	10.7	4.6	15.3
Section 504 Status	10259	2496.3	94.5	29.4	21.3	30.1	19.2	49.3
Economic Disadvantage Status	467253	2464.5	91.8	40.7	23.7	25.0	10.5	35.5

TABLE 5-10 GRADE 6 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	810983	2520.5	96.0	25.9	26.4	31.3	16.5	47.7
Female	397062	2534.5	93.2	20.7	25.7	33.9	19.7	53.5
Male	413667	2507.1	96.7	30.9	27.0	28.7	13.4	42.2
American Indian or Alaska Native	9370	2471.4	90.2	45.0	28.9	20.4	5.7	26.1
Asian	57784	2581.6	93.9	10.5	16.0	34.4	39.0	73.5
Black/African American	57014	2471.2	89.9	45.3	28.0	20.9	5.8	26.7
Native Hawaiian or Pacific Islander	9088	2493.1	91.7	34.5	29.5	27.2	8.8	36.0
Hispanic/Latino Ethnicity	302554	2492.9	90.1	34.4	30.3	27.2	8.2	35.4
White	311121	2543.9	90.0	17.2	24.4	36.4	21.9	58.3
Two or More Races	47032	2529.3	95.6	22.9	25.7	32.5	18.9	51.3
Unidentified Race	17020	2557.8	90.7	14.2	20.8	37.4	27.6	65.0
LEP Status	108817	2432.6	75.0	62.7	27.5	8.9	0.9	9.8
IDEA Indicator	81346	2423.7	86.1	67.6	20.8	9.0	2.5	11.5
Section 504 Status	10888	2521.7	88.8	23.3	30.1	32.3	14.3	46.7
Economic Disadvantage Status	446999	2489.3	89.8	36.1	30.2	26.0	7.6	33.7

TABLE 5-11 GRADE 7 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	807500	2544.9	101.2	25.9	23.7	34.5	15.8	50.3
Female	394672	2560.4	97.2	20.5	22.9	37.6	18.9	56.6
Male	412570	2530.0	102.7	31.1	24.5	31.6	12.8	44.4
American Indian or Alaska Native	9403	2493.1	95.0	44.4	26.9	23.6	5.0	28.6
Asian	58999	2609.2	95.9	10.3	13.6	37.7	38.4	76.1
Black/African American	56826	2491.3	96.8	45.8	26.0	22.6	5.6	28.2
Native Hawaiian or Pacific Islander	8854	2512.5	96.8	36.0	27.2	29.6	7.2	36.8
Hispanic/Latino Ethnicity	294202	2514.7	95.7	35.1	27.5	29.9	7.6	37.4
White	316143	2569.5	93.9	17.0	22.0	40.3	20.7	61.0
Two or More Races	45043	2554.9	100.1	22.8	22.5	36.5	18.2	54.7
Unidentified Race	18030	2582.3	94.8	14.4	18.8	40.7	26.2	66.8
LEP Status	92508	2444.7	78.9	66.7	24.0	8.6	0.6	9.3
IDEA Indicator	77165	2441.6	89.9	68.1	19.9	10.0	2.0	12.0
Section 504 Status	12108	2549.7	93.6	22.4	26.3	37.1	14.3	51.3
Economic Disadvantage Status	433101	2511.2	95.7	36.6	27.5	28.8	7.2	36.0

TABLE 5-12 GRADE 8 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	800758	2561.2	100.5	24.0	26.1	34.6	15.3	49.9
Female	391184	2578.7	96.1	17.9	25.3	38.0	18.8	56.8
Male	409318	2544.4	101.8	29.9	26.8	31.3	11.9	43.3
American Indian or Alaska Native	9074	2508.7	96.2	42.4	29.5	22.8	5.3	28.1
Asian	58801	2624.4	96.6	9.7	15.3	38.2	36.8	75.0
Black/African American	56730	2512.1	95.1	41.5	29.1	23.7	5.6	29.4
Native Hawaiian or Pacific Islander	8298	2526.2	95.3	35.2	29.9	28.2	6.7	34.9
Hispanic/Latino Ethnicity	290660	2532.0	94.7	32.5	30.3	29.7	7.6	37.2
White	316244	2584.3	94.4	16.0	23.9	40.3	19.9	60.1
Two or More Races	42369	2568.3	99.7	21.4	25.6	36.2	16.9	53.0
Unidentified Race	18582	2597.7	94.9	13.3	20.5	41.6	24.7	66.3
LEP Status	78241	2457.9	76.8	65.8	25.9	7.6	0.6	8.3
IDEA Indicator	74709	2457.4	86.8	66.7	22.0	9.6	1.8	11.4
Section 504 Status	13070	2560.4	94.1	22.4	29.0	35.4	13.1	48.6
Economic Disadvantage Status	422261	2529.0	94.8	33.7	30.2	28.9	7.2	36.1

TABLE 5-13 GRADE 11 - ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	529447	2601.4	114.9	18.9	21.6	32.4	27.1	59.5
Female	258972	2617.4	108.6	14.2	20.7	34.4	30.7	65.1
Male	270268	2586.0	118.7	23.4	22.5	30.5	23.6	54.1
American Indian or Alaska Native	4995	2555.8	110.2	29.4	28.8	28.4	13.4	41.8
Asian	51389	2664.8	108.6	8.8	11.8	28.0	51.4	79.4
Black/African American	28324	2547.1	113.5	33.2	26.5	27.4	12.8	40.3
Native Hawaiian or Pacific Islander	5782	2556.0	109.5	28.9	28.5	29.7	13.0	42.7
Hispanic/Latino Ethnicity	238793	2572.8	109.7	24.4	26.1	32.4	17.1	49.5
White	158520	2630.5	108.7	12.2	17.5	34.2	36.1	70.3
Two or More Races	25562	2616.5	109.5	14.4	20.0	35.1	30.5	65.6
Unidentified Race	16082	2637.2	104.1	10.3	16.4	35.5	37.8	73.3
LEP Status	48621	2469.9	86.6	61.7	27.7	9.4	1.1	10.6
IDEA Indicator	48990	2483.1	99.8	57.2	26.0	13.3	3.5	16.8
Section 504 Status	10261	2601.9	111.5	17.7	22.9	33.3	26.1	59.4
Economic Disadvantage Status	283893	2570.6	111.2	25.4	26.0	31.6	17.0	48.6

TABLE 5-14 GRADE 3 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	729739	2428.2	82.4	27.5	25.3	28.2	19.0	47.3
Female	355761	2426.6	79.6	27.8	26.1	28.5	17.7	46.2
Male	373729	2429.6	85.0	27.2	24.5	28.0	20.3	48.3
American Indian or Alaska Native	8121	2387.7	76.6	46.2	27.2	19.5	7.0	26.5
Asian	49727	2488.2	80.3	9.7	14.7	29.3	46.3	75.6
Black/African American	54393	2383.5	78.0	47.7	26.9	19.0	6.5	25.5
Native Hawaiian or Pacific Islander	6892	2408.8	79.6	35.0	26.9	26.6	11.5	38.1
Hispanic/Latino Ethnicity	284731	2407.5	76.0	35.3	28.6	25.5	10.6	36.1
White	262929	2448.0	77.6	18.1	23.6	33.1	25.2	58.3
Two or More Races	46857	2430.2	85.1	27.6	24.3	27.4	20.8	48.2
Unidentified Race	16089	2456.9	80.5	16.5	21.0	32.2	30.3	62.5
LEP Status	163674	2391.4	74.6	44.0	29.0	20.2	6.9	27.1
IDEA Indicator	68156	2364.2	91.0	58.7	19.5	14.3	7.5	21.8
Section 504 Status	4798	2425.6	81.3	28.9	26.1	27.4	17.7	45.1
Economic Disadvantage Status	420340	2404.0	77.1	37.1	28.3	24.6	10.0	34.6

TABLE 5-15 GRADE 4 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	822609	2467.9	85.4	25.6	31.5	25.1	17.8	42.9
Female	401931	2465.8	81.5	25.6	33.1	25.2	16.0	41.2
Male	420423	2470.0	88.9	25.5	30.0	25.0	19.5	44.5
American Indian or Alaska Native	9359	2425.9	77.7	43.3	34.6	16.0	6.1	22.2
Asian	56589	2531.1	84.0	8.7	18.5	28.0	44.8	72.8
Black/African American	58279	2419.4	78.7	46.5	32.9	15.2	5.3	20.6
Native Hawaiian or Pacific Islander	7953	2445.8	80.5	32.9	34.7	22.5	9.9	32.3
Hispanic/Latino Ethnicity	308103	2441.2	77.5	35.1	35.7	20.7	8.5	29.1
White	313042	2491.1	79.6	15.3	29.8	30.8	24.0	54.9
Two or More Races	52989	2473.7	86.5	23.8	30.6	25.6	20.1	45.7
Unidentified Race	16295	2497.2	84.1	15.6	26.1	29.4	28.9	58.3
LEP Status	155366	2414.9	72.6	48.9	34.8	12.6	3.7	16.3
IDEA Indicator	82849	2397.3	89.8	59.5	23.6	10.9	6.0	16.9
Section 504 Status	8776	2472.2	84.1	23.3	32.9	25.2	18.6	43.8
Economic Disadvantage Status	465451	2440.2	78.5	35.7	35.5	20.4	8.5	28.9

TABLE 5-16 GRADE 5 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	832537	2492.1	93.1	35.9	27.7	17.2	19.2	36.4
Female	407870	2491.1	89.4	35.9	29.2	17.1	17.8	34.9
Male	424401	2493.1	96.5	35.9	26.2	17.4	20.5	37.9
American Indian or Alaska Native	9435	2447.0	84.5	55.2	26.8	11.2	6.7	17.9
Asian	57891	2561.9	92.1	13.8	19.2	19.8	47.2	67.0
Black/African American	58692	2438.7	82.0	59.9	25.2	9.4	5.5	14.9
Native Hawaiian or Pacific Islander	8104	2472.4	86.2	43.2	29.2	15.9	11.6	27.6
Hispanic/Latino Ethnicity	315442	2461.6	83.9	48.4	29.2	13.2	9.1	22.3
White	314654	2518.6	87.0	23.4	28.3	22.1	26.2	48.4
Two or More Races	51136	2499.8	93.5	32.9	27.3	17.9	21.8	39.7
Unidentified Race	17183	2525.9	91.0	22.2	25.4	21.4	31.0	52.4
LEP Status	138459	2426.3	74.9	67.2	23.5	6.4	3.0	9.4
IDEA Indicator	85162	2412.1	89.5	72.1	16.5	6.4	5.1	11.5
Section 504 Status	10393	2496.7	90.5	33.9	28.9	18.0	19.2	37.2
Economic Disadvantage Status	469363	2461.0	84.5	48.7	29.0	13.2	9.1	22.3

TABLE 5-17 GRADE 6 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	813479	2512.4	107.2	33.2	29.2	19.4	18.1	37.5
Female	398248	2516.0	102.5	31.5	30.3	20.3	17.8	38.1
Male	414977	2509.0	111.5	34.9	28.2	18.5	18.4	37.0
American Indian or Alaska Native	9371	2460.2	101.8	52.1	29.4	12.0	6.5	18.5
Asian	58477	2595.3	103.7	11.9	18.7	21.6	47.8	69.4
Black/African American	57055	2449.7	99.6	57.3	27.1	10.5	5.1	15.6
Native Hawaiian or Pacific Islander	9113	2484.3	102.6	42.3	31.1	16.9	9.7	26.7
Hispanic/Latino Ethnicity	304014	2478.4	99.3	45.1	30.9	15.6	8.4	24.0
White	311208	2540.7	97.3	21.5	30.1	24.3	24.1	48.4
Two or More Races	46998	2519.1	108.8	31.1	28.8	20.0	20.1	40.2
Unidentified Race	17243	2554.8	101.5	19.2	26.0	23.3	31.5	54.8
LEP Status	111268	2423.2	91.7	70.1	22.2	5.5	2.2	7.7
IDEA Indicator	81225	2403.6	105.1	74.7	16.4	5.5	3.5	8.9
Section 504 Status	11026	2517.6	100.8	30.5	32.2	19.8	17.5	37.3
Economic Disadvantage Status	449148	2476.5	100.1	45.8	30.8	15.2	8.2	23.4

TABLE 5-18 GRADE 7 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	808946	2529.3	112.9	34.0	27.5	20.0	18.4	38.5
Female	395314	2531.7	108.8	32.8	28.9	20.3	18.0	38.3
Male	413375	2526.9	116.7	35.3	26.2	19.7	18.9	38.6
American Indian or Alaska Native	9367	2474.5	101.4	53.5	28.0	12.6	5.9	18.5
Asian	59509	2619.9	110.8	12.0	16.8	22.1	49.1	71.2
Black/African American	56759	2463.4	100.7	58.5	25.7	10.8	5.0	15.8
Native Hawaiian or Pacific Islander	8870	2493.9	104.2	45.0	30.0	16.6	8.4	25.0
Hispanic/Latino Ethnicity	295538	2490.5	102.6	46.7	29.6	15.7	8.0	23.7
White	315678	2559.2	103.3	22.3	28.0	25.2	24.5	49.7
Two or More Races	44929	2537.6	113.3	31.1	27.5	20.8	20.6	41.4
Unidentified Race	18296	2573.4	109.3	20.3	24.0	24.7	31.0	55.7
LEP Status	95173	2429.0	92.9	73.5	18.9	5.3	2.2	7.5
IDEA Indicator	76894	2416.1	101.0	76.6	15.0	5.5	2.9	8.4
Section 504 Status	12178	2539.1	106.1	29.4	30.4	21.7	18.6	40.3
Economic Disadvantage Status	434618	2489.8	103.6	47.2	29.2	15.4	8.2	23.6

TABLE 5-19 GRADE 8 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	794913	2543.0	121.4	38.7	24.5	17.0	19.8	36.8
Female	387966	2549.4	116.8	35.8	25.8	18.3	20.2	38.5
Male	406691	2536.9	125.3	41.5	23.3	15.8	19.5	35.3
American Indian or Alaska Native	8972	2485.3	108.0	58.8	23.2	10.9	7.0	18.0
Asian	58514	2643.8	122.6	14.4	15.3	18.1	52.2	70.3
Black/African American	56284	2474.0	105.0	63.3	21.5	9.5	5.7	15.2
Native Hawaiian or Pacific Islander	8213	2505.6	110.3	49.7	26.7	14.0	9.6	23.6
Hispanic/Latino Ethnicity	289630	2503.2	109.1	51.5	25.3	13.8	9.4	23.2
White	312818	2572.4	113.1	27.3	26.0	21.1	25.6	46.7
Two or More Races	41722	2548.9	121.1	36.4	25.1	17.6	21.0	38.5
Unidentified Race	18760	2592.5	120.4	23.6	21.9	20.3	34.3	54.6
LEP Status	80633	2436.0	97.5	78.3	14.3	4.5	2.8	7.4
IDEA Indicator	74211	2423.6	101.6	80.8	11.9	4.3	3.0	7.3
Section 504 Status	13144	2545.9	113.5	37.0	27.3	17.0	18.6	35.6
Economic Disadvantage Status	420800	2501.9	110.0	52.2	25.0	13.4	9.5	22.9

TABLE 5-20 GRADE 11 - MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Achievement Levels (%)				
	N	M	SD	1	2	3	4	3 & 4
Total	550267	2565.8	127.0	43.3	24.7	19.6	12.4	32.0
Female	269894	2570.0	120.0	41.5	26.2	20.9	11.4	32.4
Male	280174	2561.8	133.2	45.1	23.2	18.3	13.3	31.6
American Indian or Alaska Native	5263	2513.7	115.1	60.7	22.9	12.3	4.2	16.4
Asian	52869	2670.5	126.4	16.3	17.6	26.5	39.6	66.1
Black/African American	28810	2501.1	113.1	64.6	21.1	11.2	3.2	14.4
Native Hawaiian or Pacific Islander	5972	2525.2	113.4	55.5	26.1	13.9	4.5	18.4
Hispanic/Latino Ethnicity	242719	2526.8	112.7	55.8	24.9	14.6	4.7	19.3
White	171613	2598.0	120.6	31.3	26.9	25.5	16.4	41.9
Two or More Races	26829	2574.0	124.1	40.1	26.0	20.7	13.1	33.8
Unidentified Race	16192	2601.2	119.1	30.2	25.4	27.6	16.8	44.4
LEP Status	49387	2449.9	102.2	84.1	10.1	3.9	1.9	5.8
IDEA Indicator	49159	2439.6	100.6	85.7	9.6	3.4	1.3	4.7
Section 504 Status	11339	2566.8	119.7	42.7	26.9	19.3	11.0	30.3
Economic Disadvantage Status	291777	2528.7	116.2	55.3	24.3	14.7	5.7	20.3

Claim-Level Results

Students receive test results at the claim level. For each claim, students are classified into one of three reporting categories: 'Above Standard', 'At or Near Standard,' or "Below Standard'. The standard is the Level 3 cut score. Table 5-21 describes the Smarter Balanced assessment claims. Claims 2 and 4 of mathematics are reported as a single claim. Following table 5-21 are seven grade-level tables for each claim score within each content area.

For ELA/literacy:

- Table 5-22 through Table 5-28 present results for Claim 1 by grade.
- Table 5-29 through Table 5-35 present results for Claim 2 by grade.
- Table 5-36 through Table 5-42 present results for Claim 3 by grade.
- Table 5-43 through Table 5-49 present results for Claim 4 by grade.

For Mathematics:

- Table 5-50 through Table 5-56 present results for Claim 1 by grade.
- Table 5-57 through Table 5-63 present results for Claims 2/4 by grade.
- Table 5-64 through Table 5-70 present results for Claim 3 by grade.

Results in these tables are presented in terms of the average scale score and the percentage of students in each reporting category. Results are presented for the reporting members' students overall (total) and by demographic group.

TABLE 5-21 SMARTER BALANCED ASSESSMENT CLAIMS

	ELA/ Literacy	Mathematics
Claim 1	Reading: Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.	**Concepts & Procedures**: Students can explain and apply mathematics concepts and interpret and carry out mathematics procedures with precision and fluency.
Claim 2	Writing: Students can produce effective and well-grounded writing for a range of purposes and audiences.	**Problem Solving:*** Students can solve a range of complex well-posed problems in pure and applied mathematics, making productive use of knowledge and problem solving strategies.
Claim 3	Speaking and Listening: Students can employ effective speaking and listening skills for a range of purposes and audiences.	**Communicating Reasoning:** Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.
Claim 4	Research/Inquiry: Students can engage in research and inquiry to investigate topics, and to analyze, integrate, and present information.	**Modeling and Data Analysis:*** Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems.

**Claims 2 and 4 in mathematics are reported jointly*

TABLE 5-22 GRADE 3 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809357	2412.9	101.7	36.2	40.3	23.5
Female	395177	2421.7	101.3	32.8	41.1	26.1
Male	413933	2404.5	101.3	39.5	39.5	21.0
American Indian or Alaska Native	9238	2368.4	92.3	54.5	35.5	10.0
Asian	55175	2462.0	99.9	18.2	41.1	40.7
Black/African American	57684	2372.9	91.5	54.7	34.2	11.1
Native Hawaiian or Pacific Islander	7750	2382.8	95.9	48.1	38.4	13.5
Hispanic/Latino Ethnicity	303225	2384.2	95.5	46.4	39.7	13.9
White	306284	2439.6	98.4	26.0	41.9	32.1
Two or More Races	54027	2419.3	103.1	34.1	40.3	25.7
Unidentified Race	15974	2438.3	96.3	25.5	43.1	31.4
LEP Status	174143	2358.5	87.0	58.4	34.7	6.9
IDEA Indicator	77370	2350.6	96.7	63.2	28.0	8.8
Section 504 Status	6681	2414.0	102.9	34.7	41.9	23.3
Economic Disadvantage Status	460030	2383.7	94.6	47.4	39.0	13.6

TABLE 5-23 GRADE 4 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	817905	2456.5	108.0	29.8	45.8	24.3
Female	399738	2465.2	105.9	26.5	46.9	26.6
Male	417914	2448.3	109.2	32.9	44.8	22.2
American Indian or Alaska Native	9349	2410.0	104.7	45.2	44.0	10.9
Asian	55643	2508.2	102.1	14.0	43.5	42.5
Black/African American	58117	2411.0	102.0	49.5	38.8	11.7
Native Hawaiian or Pacific Islander	7886	2423.6	105.3	39.6	46.5	13.9
Hispanic/Latino Ethnicity	305880	2426.6	103.7	38.3	47.4	14.3
White	312129	2484.4	102.0	21.0	46.0	33.0
Two or More Races	52923	2465.4	107.9	26.7	46.4	26.9
Unidentified Race	15978	2484.6	103.0	20.6	45.4	33.9
LEP Status	152599	2386.3	94.5	54.7	40.6	4.7
IDEA Indicator	82354	2381.5	108.3	58.0	34.0	8.1
Section 504 Status	8658	2459.5	107.7	26.4	49.3	24.3
Economic Disadvantage Status	462297	2425.4	103.3	39.9	45.9	14.2

TABLE 5-24 GRADE 5 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	828637	2490.8	109.4	29.5	45.6	24.8
Female	406079	2503.0	108.0	25.6	46.1	28.3
Male	422293	2479.1	109.5	33.4	45.2	21.5
American Indian or Alaska Native	9443	2442.0	105.6	46.9	41.7	11.4
Asian	57061	2542.9	106.2	14.6	42.1	43.3
Black/African American	58556	2446.1	101.9	47.8	39.9	12.3
Native Hawaiian or Pacific Islander	8090	2459.6	110.9	39.1	45.0	15.9
Hispanic/Latino Ethnicity	313504	2458.2	103.1	38.5	47.6	13.9
White	314043	2521.3	103.9	20.2	45.4	34.4
Two or More Races	51050	2501.1	110.3	26.5	45.7	27.8
Unidentified Race	16890	2518.9	103.4	20.0	46.4	33.5
LEP Status	136020	2408.4	89.9	58.5	38.1	3.5
IDEA Indicator	84806	2406.6	103.8	60.2	33.0	6.8
Section 504 Status	10237	2492.7	109.6	27.8	47.9	24.2
Economic Disadvantage Status	466548	2458.5	103.4	39.6	46.1	14.3

TABLE 5-25 GRADE 6 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809128	2511.5	110.5	29.7	48.6	21.7
Female	396323	2522.0	107.8	25.8	50.2	23.9
Male	412552	2501.5	112.0	33.5	46.9	19.6
American Indian or Alaska Native	9345	2466.3	106.3	46.0	43.7	10.3
Asian	57730	2568.1	107.2	14.6	44.3	41.1
Black/African American	56788	2464.5	104.2	46.5	43.7	9.8
Native Hawaiian or Pacific Islander	9069	2479.6	107.9	40.7	46.8	12.5
Hispanic/Latino Ethnicity	301644	2479.5	104.3	39.3	48.6	12.1
White	310623	2539.9	104.8	20.3	50.3	29.3
Two or More Races	46934	2521.0	111.4	27.1	48.5	24.4
Unidentified Race	16995	2544.4	104.1	19.1	49.4	31.5
LEP Status	108387	2418.3	89.9	64.1	33.9	2.0
IDEA Indicator	80791	2420.4	101.5	63.8	31.7	4.6
Section 504 Status	10852	2515.3	107.5	27.0	51.5	21.5
Economic Disadvantage Status	445662	2478.6	104.6	40.0	47.9	12.0

TABLE 5-26 GRADE 7 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	805305	2541.3	112.6	28.7	45.6	25.7
Female	393737	2554.0	109.2	24.1	47.2	28.7
Male	411313	2529.0	114.4	33.1	44.1	22.8
American Indian or Alaska Native	9365	2492.6	105.6	45.0	43.2	11.9
Asian	58919	2600.2	108.9	13.8	39.3	46.9
Black/African American	56569	2494.0	104.3	44.8	43.2	12.0
Native Hawaiian or Pacific Islander	8828	2504.5	110.6	40.6	44.4	15.1
Hispanic/Latino Ethnicity	293098	2505.4	106.4	39.5	45.8	14.7
White	315632	2571.2	105.9	18.8	47.1	34.2
Two or More Races	44893	2552.2	112.8	25.4	45.8	28.7
Unidentified Race	18001	2573.3	105.9	18.6	45.6	35.8
LEP Status	92062	2437.7	86.9	67.8	30.0	2.2
IDEA Indicator	76588	2443.9	99.0	65.4	29.6	5.0
Section 504 Status	12074	2546.4	110.1	26.1	48.1	25.8
Economic Disadvantage Status	431608	2506.1	106.4	39.4	45.8	14.7

TABLE 5-27 GRADE 8 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	798992	2556.0	112.0	29.2	45.5	25.4
Female	390378	2571.0	108.3	24.0	46.8	29.2
Male	408358	2541.5	113.5	34.1	44.2	21.8
American Indian or Alaska Native	9048	2507.3	108.5	46.5	41.2	12.3
Asian	58738	2613.9	106.4	14.1	40.2	45.7
Black/African American	56535	2508.9	107.2	44.8	42.6	12.6
Native Hawaiian or Pacific Islander	8264	2518.2	110.3	42.8	42.4	14.8
Hispanic/Latino Ethnicity	289760	2524.7	106.0	38.7	46.0	15.3
White	315834	2581.5	107.4	20.6	46.6	32.9
Two or More Races	42265	2564.8	111.5	26.4	45.7	27.9
Unidentified Race	18548	2589.2	105.5	18.4	46.0	35.6
LEP Status	77951	2451.0	87.5	68.3	29.5	2.2
IDEA Indicator	74305	2457.9	98.8	65.8	29.3	4.8
Section 504 Status	13018	2557.4	108.2	28.5	47.1	24.4
Economic Disadvantage Status	421069	2523.5	106.9	39.2	45.5	15.3

TABLE 5-28 GRADE 11 - ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	526883	2598.2	120.9	20.5	45.6	33.9
Female	257752	2610.7	115.4	16.8	46.4	36.8
Male	268924	2586.3	124.7	24.1	44.8	31.0
American Indian or Alaska Native	4972	2559.2	120.3	30.1	48.3	21.6
Asian	51252	2651.3	114.8	11.0	35.8	53.2
Black/African American	28081	2550.4	119.4	32.7	47.7	19.6
Native Hawaiian or Pacific Islander	5734	2553.6	120.2	31.8	48.7	19.5
Hispanic/Latino Ethnicity	237482	2569.5	115.9	26.2	50.1	23.6
White	157933	2629.9	116.3	13.6	41.6	44.8
Two or More Races	25407	2615.5	117.1	16.2	45.2	38.7
Unidentified Race	16022	2626.1	109.5	12.6	45.3	42.1
LEP Status	48221	2468.7	95.7	60.2	37.2	2.6
IDEA Indicator	48461	2488.8	111.4	53.5	39.1	7.4
Section 504 Status	10193	2603.2	118.7	19.4	45.0	35.6
Economic Disadvantage Status	282242	2568.1	117.7	27.1	49.3	23.6

TABLE 5-29 GRADE 3 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809357	2416.8	107.9	29.0	48.9	22.1
Female	395177	2427.4	106.4	25.4	49.3	25.2
Male	413933	2406.7	108.3	32.3	48.5	19.2
American Indian or Alaska Native	9238	2369.6	111.5	37.3	53.4	9.3
Asian	55175	2471.8	103.8	13.2	44.1	42.7
Black/African American	57684	2371.4	103.0	47.1	41.9	11.0
Native Hawaiian or Pacific Islander	7750	2387.4	114.7	31.2	55.6	13.2
Hispanic/Latino Ethnicity	303225	2390.9	99.0	38.7	46.5	14.9
White	306284	2440.6	106.1	19.9	52.7	27.5
Two or More Races	54027	2423.8	114.7	23.1	53.2	23.6
Unidentified Race	15974	2444.9	98.4	21.1	44.8	34.1
LEP Status	174143	2363.8	95.4	48.5	43.8	7.7
IDEA Indicator	77370	2342.9	111.4	55.0	37.0	8.0
Section 504 Status	6681	2421.4	110.6	24.5	54.8	20.7
Economic Disadvantage Status	460030	2388.1	102.5	38.7	47.7	13.6

TABLE 5-30 GRADE 4 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	817905	2457.5	113.8	29.8	48.8	21.4
Female	399738	2470.1	111.6	25.3	50.0	24.8
Male	417914	2445.4	114.5	34.1	47.7	18.2
American Indian or Alaska Native	9349	2402.9	117.8	42.6	48.8	8.6
Asian	55643	2519.4	112.2	13.2	43.1	43.7
Black/African American	58117	2408.3	106.7	49.7	40.4	9.9
Native Hawaiian or Pacific Islander	7886	2419.9	125.1	36.1	51.5	12.5
Hispanic/Latino Ethnicity	305880	2432.2	104.6	38.6	47.4	14.0
White	312129	2480.2	111.6	21.1	52.4	26.5
Two or More Races	52923	2463.1	122.1	25.0	51.9	23.1
Unidentified Race	15978	2495.2	104.6	20.1	44.0	35.9
LEP Status	152599	2391.3	98.1	54.2	40.9	4.9
IDEA Indicator	82354	2370.5	116.8	60.9	32.6	6.5
Section 504 Status	8658	2456.8	118.0	27.9	52.6	19.5
Economic Disadvantage Status	462297	2427.0	107.7	39.7	47.7	12.6

TABLE 5-31 GRADE 5 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	828637	2500.4	115.7	25.7	48.8	25.5
Female	406079	2515.0	112.5	21.0	49.5	29.5
Male	422293	2486.2	116.8	30.3	48.1	21.6
American Indian or Alaska Native	9443	2440.1	118.4	38.1	51.9	10.0
Asian	57061	2563.1	114.2	12.0	38.5	49.5
Black/African American	58556	2452.4	108.9	42.5	45.0	12.5
Native Hawaiian or Pacific Islander	8090	2467.3	127.6	29.6	54.4	16.0
Hispanic/Latino Ethnicity	313504	2472.2	107.5	34.6	47.8	17.6
White	314043	2525.3	111.8	17.1	51.9	31.0
Two or More Races	51050	2507.5	122.4	21.2	52.4	26.4
Unidentified Race	16890	2541.6	106.3	16.1	40.9	43.0
LEP Status	136020	2419.4	98.9	53.3	41.9	4.9
IDEA Indicator	84806	2398.7	116.6	59.8	33.6	6.5
Section 504 Status	10237	2501.7	115.3	23.7	53.3	23.0
Economic Disadvantage Status	466548	2468.2	109.9	35.0	49.4	15.7

TABLE 5-32 GRADE 6 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809128	2523.3	115.9	29.0	49.6	21.4
Female	396323	2539.3	112.7	23.7	50.8	25.6
Male	412552	2507.9	116.9	34.2	48.4	17.3
American Indian or Alaska Native	9345	2475.2	116.6	39.1	52.4	8.5
Asian	57730	2587.6	111.9	12.5	42.0	45.4
Black/African American	56788	2467.3	109.6	48.8	41.9	9.4
Native Hawaiian or Pacific Islander	9069	2502.0	124.7	30.4	56.9	12.7
Hispanic/Latino Ethnicity	301644	2491.9	101.8	39.5	47.2	13.3
White	310623	2550.0	116.5	19.7	53.9	26.4
Two or More Races	46934	2538.7	122.6	21.9	54.9	23.2
Unidentified Race	16995	2558.0	99.2	18.0	44.1	37.9
LEP Status	108387	2430.8	93.2	64.3	33.4	2.3
IDEA Indicator	80791	2420.6	109.1	65.6	30.5	3.9
Section 504 Status	10852	2532.3	115.3	24.3	57.0	18.7
Economic Disadvantage Status	445662	2489.5	107.8	39.6	48.3	12.1

TABLE 5-33 GRADE 7 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	805305	2542.7	118.4	26.3	50.5	23.2
Female	393737	2559.8	113.8	21.0	51.4	27.7
Male	411313	2526.4	120.4	31.4	49.7	18.9
American Indian or Alaska Native	9365	2487.2	120.4	40.7	50.7	8.6
Asian	58919	2611.3	109.4	10.3	40.7	49.0
Black/African American	56569	2481.0	118.2	46.6	43.1	10.3
Native Hawaiian or Pacific Islander	8828	2511.3	123.3	32.1	56.0	11.8
Hispanic/Latino Ethnicity	293098	2516.2	108.7	33.9	49.9	16.2
White	315632	2564.4	116.4	19.2	54.0	26.8
Two or More Races	44893	2552.9	122.8	22.1	53.9	24.0
Unidentified Race	18001	2584.7	103.6	14.9	43.6	41.5
LEP Status	92062	2442.9	100.7	61.3	36.3	2.4
IDEA Indicator	76588	2436.8	113.4	62.5	33.4	4.1
Section 504 Status	12074	2552.5	114.6	21.5	57.5	21.0
Economic Disadvantage Status	431608	2508.9	113.6	36.0	50.1	13.9

TABLE 5-34 GRADE 8 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	798992	2563.2	118.9	24.9	52.9	22.3
Female	390378	2581.9	113.7	19.0	54.0	27.0
Male	408358	2545.2	120.8	30.5	51.7	17.8
American Indian or Alaska Native	9048	2511.7	126.6	34.3	56.3	9.4
Asian	58738	2628.4	113.5	10.3	43.8	45.9
Black/African American	56535	2509.8	113.3	42.4	47.6	10.0
Native Hawaiian or Pacific Islander	8264	2527.7	127.6	28.6	60.0	11.4
Hispanic/Latino Ethnicity	289760	2532.8	107.6	34.5	50.9	14.6
White	315834	2587.4	117.7	16.6	56.6	26.9
Two or More Races	42265	2573.0	125.7	19.0	58.2	22.8
Unidentified Race	18548	2599.5	104.1	15.2	48.0	36.9
LEP Status	77951	2456.4	97.0	63.5	34.7	1.8
IDEA Indicator	74305	2453.2	113.0	61.7	34.8	3.5
Section 504 Status	13018	2566.8	119.6	21.8	58.3	19.9
Economic Disadvantage Status	421069	2529.3	112.5	34.6	52.4	13.1

TABLE 5-35 GRADE 11 - ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	526883	2600.4	133.6	21.2	44.6	34.1
Female	257752	2620.1	126.7	16.2	44.5	39.2
Male	268924	2581.6	137.1	26.0	44.7	29.2
American Indian or Alaska Native	4972	2548.9	135.7	29.0	53.2	17.8
Asian	51252	2673.8	124.4	9.2	30.7	60.0
Black/African American	28081	2539.1	130.2	36.7	45.1	18.2
Native Hawaiian or Pacific Islander	5734	2556.9	139.0	26.2	55.2	18.5
Hispanic/Latino Ethnicity	237482	2568.9	124.3	28.1	46.9	24.9
White	157933	2630.5	131.5	13.6	44.9	41.5
Two or More Races	25407	2618.4	137.1	14.4	50.4	35.2
Unidentified Race	16022	2646.1	118.6	11.9	36.8	51.3
LEP Status	48221	2461.7	105.7	62.2	34.6	3.2
IDEA Indicator	48461	2471.8	119.1	57.6	36.9	5.5
Section 504 Status	10193	2596.6	127.3	20.8	47.2	32.0
Economic Disadvantage Status	282242	2567.1	129.0	28.4	47.6	24.0

TABLE 5-36 GRADE 3 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809357	2418.4	116.0	18.8	61.1	20.1
Female	395177	2426.4	114.4	16.8	61.0	22.2
Male	413933	2410.7	117.0	20.6	61.1	18.2
American Indian or Alaska Native	9238	2371.2	105.5	41.6	48.8	9.5
Asian	55175	2468.1	110.0	10.2	55.4	34.4
Black/African American	57684	2372.9	118.0	23.6	66.4	9.9
Native Hawaiian or Pacific Islander	7750	2391.4	107.1	34.1	50.8	15.0
Hispanic/Latino Ethnicity	303225	2391.1	115.9	22.8	65.7	11.5
White	306284	2444.8	107.4	13.8	58.3	27.8
Two or More Races	54027	2423.6	111.6	23.7	53.7	22.6
Unidentified Race	15974	2443.4	114.9	11.3	64.1	24.6
LEP Status	174143	2362.3	111.7	31.0	62.7	6.3
IDEA Indicator	77370	2341.7	118.5	43.0	49.6	7.4
Section 504 Status	6681	2420.5	107.1	21.9	57.4	20.7
Economic Disadvantage Status	460030	2388.4	114.0	24.5	63.7	11.8

TABLE 5-37 GRADE 4 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	817905	2452.8	121.3	25.0	56.4	18.6
Female	399738	2460.2	120.0	22.4	57.3	20.4
Male	417914	2445.7	122.0	27.5	55.6	16.9
American Indian or Alaska Native	9349	2407.9	105.4	45.5	46.7	7.7
Asian	55643	2504.8	118.2	13.1	52.9	34.0
Black/African American	58117	2406.2	120.3	36.3	54.9	8.8
Native Hawaiian or Pacific Islander	7886	2423.3	112.1	39.1	48.4	12.5
Hispanic/Latino Ethnicity	305880	2420.5	120.7	32.3	57.5	10.2
White	312129	2483.2	111.5	17.0	57.2	25.7
Two or More Races	52923	2461.1	115.4	26.4	52.6	21.0
Unidentified Race	15978	2479.3	122.3	16.0	59.6	24.4
LEP Status	152599	2380.1	112.4	46.6	49.8	3.6
IDEA Indicator	82354	2365.3	120.9	57.0	37.4	5.7
Section 504 Status	8658	2454.4	114.4	26.3	55.8	17.9
Economic Disadvantage Status	462297	2420.4	118.2	33.3	56.2	10.5

TABLE 5-38 GRADE 5 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	828637	2494.1	121.0	20.7	58.5	20.8
Female	406079	2504.1	118.9	17.8	59.2	23.1
Male	422293	2484.4	122.2	23.5	57.9	18.6
American Indian or Alaska Native	9443	2443.2	113.5	42.3	47.6	10.1
Asian	57061	2547.3	118.3	10.7	52.6	36.7
Black/African American	58556	2450.1	119.2	29.7	59.8	10.5
Native Hawaiian or Pacific Islander	8090	2471.7	113.9	30.2	54.0	15.8
Hispanic/Latino Ethnicity	313504	2460.2	119.3	27.1	62.0	10.9
White	314043	2525.1	111.5	13.9	56.8	29.3
Two or More Races	51050	2506.0	115.5	21.1	54.1	24.8
Unidentified Race	16890	2522.0	118.6	12.9	61.2	25.9
LEP Status	136020	2407.2	109.1	43.5	53.5	3.0
IDEA Indicator	84806	2396.6	118.9	51.8	43.0	5.2
Section 504 Status	10237	2496.4	114.1	21.2	57.8	21.0
Economic Disadvantage Status	466548	2460.7	117.7	27.9	60.2	11.8

TABLE 5-39 GRADE 6 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809128	2527.6	117.9	20.1	63.4	16.5
Female	396323	2538.6	115.2	16.9	64.3	18.8
Male	412552	2517.1	119.4	23.1	62.5	14.4
American Indian or Alaska Native	9345	2473.2	111.1	44.1	47.4	8.6
Asian	57730	2584.0	112.5	9.8	62.5	27.7
Black/African American	56788	2478.9	117.4	30.0	62.3	7.7
Native Hawaiian or Pacific Islander	9069	2493.6	108.3	35.7	52.8	11.5
Hispanic/Latino Ethnicity	301644	2501.1	118.0	24.4	66.8	8.8
White	310623	2551.5	109.4	14.9	61.8	23.3
Two or More Races	46934	2534.3	113.0	22.9	57.5	19.6
Unidentified Race	16995	2564.1	114.1	10.2	69.5	20.3
LEP Status	108387	2433.8	107.9	45.2	52.8	2.0
IDEA Indicator	80791	2423.8	114.2	53.6	43.2	3.2
Section 504 Status	10852	2529.6	107.1	21.4	62.7	16.0
Economic Disadvantage Status	445662	2495.8	115.7	26.9	63.8	9.3

TABLE 5-40 GRADE 7 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	805305	2541.8	121.7	22.7	59.6	17.7
Female	393737	2552.8	118.8	19.4	60.4	20.2
Male	411313	2531.3	123.4	25.8	58.8	15.3
American Indian or Alaska Native	9365	2490.5	117.0	41.8	48.6	9.6
Asian	58919	2596.4	112.9	10.6	60.2	29.2
Black/African American	56569	2491.0	122.4	37.2	54.5	8.3
Native Hawaiian or Pacific Islander	8828	2509.5	118.4	34.1	52.6	13.3
Hispanic/Latino Ethnicity	293098	2508.2	119.8	29.5	62.0	8.5
White	315632	2571.1	112.9	15.7	58.8	25.5
Two or More Races	44893	2553.0	117.5	22.0	56.0	22.0
Unidentified Race	18001	2571.2	115.8	13.8	66.6	19.6
LEP Status	92062	2435.3	108.0	54.5	43.9	1.7
IDEA Indicator	76588	2433.7	115.8	57.1	39.8	3.1
Section 504 Status	12074	2548.5	112.4	20.5	60.9	18.5
Economic Disadvantage Status	431608	2506.7	119.5	31.3	58.9	9.8

TABLE 5-41 GRADE 8 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	798992	2562.1	122.9	19.2	63.5	17.3
Female	390378	2575.5	118.7	15.2	64.8	19.9
Male	408358	2549.2	125.4	22.9	62.2	14.8
American Indian or Alaska Native	9048	2505.6	117.7	41.6	49.6	8.8
Asian	58738	2618.3	115.7	8.9	60.0	31.1
Black/African American	56535	2512.0	122.4	30.7	61.3	8.0
Native Hawaiian or Pacific Islander	8264	2526.8	115.0	32.8	55.7	11.4
Hispanic/Latino Ethnicity	289760	2534.3	120.6	22.7	68.6	8.8
White	315834	2586.2	117.0	15.0	60.9	24.1
Two or More Races	42265	2566.5	119.9	21.9	57.8	20.3
Unidentified Race	18548	2592.8	118.4	10.5	69.2	20.3
LEP Status	77951	2458.3	109.7	46.8	51.7	1.5
IDEA Indicator	74305	2457.4	116.1	51.4	45.7	2.8
Section 504 Status	13018	2564.5	115.2	19.3	63.5	17.2
Economic Disadvantage Status	421069	2529.5	119.9	25.6	65.0	9.4

TABLE 5-42 GRADE 11 - ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	526883	2596.1	133.6	15.5	59.9	24.6
Female	257752	2606.8	129.3	12.7	60.6	26.8
Male	268924	2585.8	136.8	18.1	59.4	22.5
American Indian or Alaska Native	4972	2554.1	130.2	27.1	56.4	16.5
Asian	51252	2648.7	127.5	8.5	52.2	39.3
Black/African American	28081	2549.6	135.8	24.8	61.0	14.2
Native Hawaiian or Pacific Islander	5734	2551.2	128.1	28.6	55.9	15.5
Hispanic/Latino Ethnicity	237482	2568.1	130.9	18.7	65.6	15.7
White	157933	2627.4	127.4	11.1	54.7	34.2
Two or More Races	25407	2611.6	127.7	14.8	54.0	31.3
Unidentified Race	16022	2619.8	125.2	9.2	63.3	27.4
LEP Status	48221	2467.0	111.7	45.6	52.6	1.8
IDEA Indicator	48461	2484.4	123.6	43.3	51.7	5.1
Section 504 Status	10193	2603.7	132.6	14.2	60.3	25.5
Economic Disadvantage Status	282242	2565.6	131.4	20.2	63.4	16.4

TABLE 5-43 GRADE 3 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809357	2411.1	121.5	25.0	50.4	24.7
Female	395177	2420.0	119.4	22.3	50.9	26.7
Male	413933	2402.7	122.9	27.5	49.9	22.7
American Indian or Alaska Native	9238	2355.3	116.4	42.5	47.1	10.4
Asian	55175	2468.8	115.7	11.8	44.4	43.7
Black/African American	57684	2361.2	118.3	42.1	45.4	12.5
Native Hawaiian or Pacific Islander	7750	2381.1	120.9	35.5	47.8	16.6
Hispanic/Latino Ethnicity	303225	2386.5	119.3	30.1	53.1	16.8
White	306284	2434.5	115.6	18.8	50.1	31.1
Two or More Races	54027	2416.5	122.3	24.2	49.5	26.3
Unidentified Race	15974	2442.6	117.0	17.1	47.7	35.3
LEP Status	174143	2359.2	113.6	38.8	51.3	9.9
IDEA Indicator	77370	2340.1	121.7	46.8	43.5	9.7
Section 504 Status	6681	2409.5	121.7	25.0	51.9	23.1
Economic Disadvantage Status	460030	2381.8	118.7	32.7	51.4	15.9

TABLE 5-44 GRADE 4 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	817905	2453.4	123.9	26.2	49.8	24.0
Female	399738	2463.2	123.0	23.4	50.0	26.6
Male	417914	2444.0	124.1	28.9	49.6	21.6
American Indian or Alaska Native	9349	2398.6	117.3	41.9	48.1	10.1
Asian	55643	2515.8	115.2	11.6	44.4	44.0
Black/African American	58117	2399.9	118.2	44.0	45.0	10.9
Native Hawaiian or Pacific Islander	7886	2423.8	122.6	34.6	49.0	16.4
Hispanic/Latino Ethnicity	305880	2427.4	121.0	32.7	51.1	16.3
White	312129	2477.0	118.9	19.2	50.5	30.2
Two or More Races	52923	2460.7	124.3	24.0	50.0	26.0
Unidentified Race	15978	2489.9	120.2	17.5	46.8	35.7
LEP Status	152599	2385.2	111.7	46.4	47.0	6.5
IDEA Indicator	82354	2371.8	120.8	53.3	38.7	8.0
Section 504 Status	8658	2452.1	121.6	24.6	53.2	22.2
Economic Disadvantage Status	462297	2421.7	120.3	34.8	50.2	15.0

TABLE 5-45 GRADE 5 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	828637	2487.5	125.0	23.8	49.8	26.4
Female	406079	2501.1	122.0	20.4	49.8	29.8
Male	422293	2474.3	126.4	27.1	49.8	23.1
American Indian or Alaska Native	9443	2426.6	120.9	43.9	44.6	11.6
Asian	57061	2551.5	117.1	12.2	39.5	48.3
Black/African American	58556	2434.7	123.5	31.7	55.3	12.9
Native Hawaiian or Pacific Islander	8090	2458.3	124.4	35.3	45.9	18.8
Hispanic/Latino Ethnicity	313504	2457.4	120.6	33.0	49.3	17.7
White	314043	2514.6	118.7	15.1	51.8	33.1
Two or More Races	51050	2497.1	124.3	22.6	48.2	29.2
Unidentified Race	16890	2526.3	119.4	14.7	46.6	38.7
LEP Status	136020	2403.3	108.2	49.8	44.8	5.4
IDEA Indicator	84806	2389.1	118.9	56.9	35.8	7.4
Section 504 Status	10237	2484.2	121.9	26.0	48.8	25.2
Economic Disadvantage Status	466548	2454.1	121.0	32.0	51.3	16.7

TABLE 5-46 GRADE 6 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	809128	2516.2	125.1	22.0	50.8	27.2
Female	396323	2532.7	121.1	18.0	50.5	31.5
Male	412552	2500.3	126.8	25.9	51.0	23.1
American Indian or Alaska Native	9345	2457.0	124.1	39.5	48.5	12.0
Asian	57730	2585.8	112.8	9.0	39.7	51.3
Black/African American	56788	2460.9	122.7	37.3	49.7	13.0
Native Hawaiian or Pacific Islander	9069	2484.7	128.3	30.8	50.6	18.6
Hispanic/Latino Ethnicity	301644	2490.5	123.8	27.9	52.9	19.2
White	310623	2537.8	118.0	15.8	51.5	32.7
Two or More Races	46934	2522.5	125.9	20.8	50.2	29.1
Unidentified Race	16995	2560.3	114.2	12.0	46.2	41.8
LEP Status	108387	2424.6	112.2	49.0	45.9	5.1
IDEA Indicator	80791	2410.5	119.1	54.2	40.0	5.8
Section 504 Status	10852	2510.8	122.9	21.6	54.4	24.0
Economic Disadvantage Status	445662	2483.9	123.5	29.9	52.6	17.5

TABLE 5-47 GRADE 7 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	805305	2541.0	128.3	21.3	51.3	27.4
Female	393737	2560.0	123.1	16.9	51.2	32.0
Male	411313	2522.8	130.5	25.5	51.4	23.0
American Indian or Alaska Native	9365	2481.2	126.2	38.0	49.3	12.7
Asian	58919	2614.9	114.4	8.4	38.9	52.7
Black/African American	56569	2480.7	125.7	36.7	50.8	12.4
Native Hawaiian or Pacific Islander	8828	2503.3	131.3	31.0	51.3	17.7
Hispanic/Latino Ethnicity	293098	2513.4	126.0	27.0	53.7	19.3
White	315632	2562.5	121.5	15.7	51.9	32.5
Two or More Races	44893	2550.3	128.4	19.5	50.5	30.0
Unidentified Race	18001	2586.2	117.2	12.3	46.6	41.1
LEP Status	92062	2438.1	112.2	49.5	46.2	4.2
IDEA Indicator	76588	2429.2	118.3	52.5	42.1	5.4
Section 504 Status	12074	2541.2	122.9	19.5	55.4	25.1
Economic Disadvantage Status	431608	2506.7	126.3	29.1	53.3	17.6

TABLE 5-48 GRADE 8 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	798992	2557.4	126.9	20.3	52.6	27.2
Female	390378	2578.7	121.3	15.4	52.1	32.5
Male	408358	2537.0	128.7	24.9	53.0	22.1
American Indian or Alaska Native	9048	2498.1	125.8	36.9	50.4	12.7
Asian	58738	2630.4	115.3	9.0	39.0	52.0
Black/African American	56535	2506.0	124.1	28.2	58.1	13.7
Native Hawaiian or Pacific Islander	8264	2518.0	129.2	32.3	51.1	16.6
Hispanic/Latino Ethnicity	289760	2528.6	123.8	27.9	53.2	18.9
White	315834	2578.6	120.8	13.7	54.1	32.2
Two or More Races	42265	2563.8	128.1	19.5	51.5	29.0
Unidentified Race	18548	2601.9	117.1	11.7	46.9	41.5
LEP Status	77951	2451.3	108.8	50.3	45.9	3.8
IDEA Indicator	74305	2447.4	113.5	53.8	41.4	4.8
Section 504 Status	13018	2549.5	123.9	22.6	53.6	23.8
Economic Disadvantage Status	421069	2524.6	124.7	27.5	54.8	17.7

TABLE 5-49 GRADE 11 - ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	526883	2600.6	137.9	16.5	47.6	35.9
Female	257752	2619.5	130.8	12.7	46.6	40.7
Male	268924	2582.4	142.1	20.1	48.5	31.4
American Indian or Alaska Native	4972	2544.5	139.4	27.6	51.8	20.5
Asian	51252	2666.4	122.0	7.6	34.6	57.8
Black/African American	28081	2541.8	139.9	28.2	51.3	20.5
Native Hawaiian or Pacific Islander	5734	2545.9	143.9	27.1	52.3	20.6
Hispanic/Latino Ethnicity	237482	2576.1	136.3	20.4	51.5	28.1
White	157933	2624.0	132.2	11.8	45.5	42.7
Two or More Races	25407	2613.2	135.4	13.8	47.7	38.6
Unidentified Race	16022	2642.4	122.6	8.9	42.4	48.7
LEP Status	48221	2470.6	119.0	47.1	47.6	5.3
IDEA Indicator	48461	2476.4	126.5	45.4	47.0	7.6
Section 504 Status	10193	2594.3	136.1	16.2	50.8	33.0
Economic Disadvantage Status	282242	2572.3	138.6	21.5	51.1	27.4

TABLE 5-50 GRADE 3 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	586531	2431.6	88.7	33.8	34.7	31.6
Female	285892	2429.2	85.0	34.5	35.7	29.8
Male	300391	2433.8	92.1	33.1	33.6	33.3
American Indian or Alaska Native	7133	2386.3	84.3	54.6	31.2	14.1
Asian	47845	2496.1	85.0	12.2	26.5	61.3
Black/African American	31009	2390.9	86.8	51.3	32.6	16.2
Native Hawaiian or Pacific Islander	6225	2410.8	87.0	41.8	35.7	22.5
Hispanic/Latino Ethnicity	259812	2410.6	81.8	42.2	36.4	21.5
White	182247	2452.1	84.6	24.1	35.2	40.7
Two or More Races	39748	2432.8	92.5	34.2	32.6	33.2
Unidentified Race	12512	2453.9	84.8	23.0	34.8	42.2
LEP Status	143137	2393.2	81.2	51.6	33.6	14.8
IDEA Indicator	63805	2362.8	99.5	63.8	22.8	13.4
Section 504 Status	4790	2427.5	86.9	35.6	34.7	29.7
Economic Disadvantage Status	340896	2407.4	83.5	43.8	35.5	20.7

TABLE 5-51 GRADE 4 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	675686	2470.1	92.0	39.3	32.0	28.7
Female	329957	2467.3	86.9	40.6	33.0	26.4
Male	345474	2472.7	96.5	38.1	31.0	30.9
American Indian or Alaska Native	8300	2423.0	85.3	62.1	26.4	11.5
Asian	54649	2536.8	88.4	15.1	25.9	59.0
Black/African American	34958	2425.1	90.3	59.0	27.2	13.7
Native Hawaiian or Pacific Islander	7263	2447.3	87.6	48.9	32.4	18.7
Hispanic/Latino Ethnicity	282579	2442.1	83.7	51.7	31.7	16.6
White	229289	2495.4	86.0	26.8	34.7	38.6
Two or More Races	46043	2476.2	93.3	36.9	31.7	31.4
Unidentified Race	12605	2491.3	87.7	28.9	33.3	37.8
LEP Status	140080	2415.5	80.4	66.1	25.3	8.7
IDEA Indicator	78268	2394.9	99.8	70.7	18.8	10.5
Section 504 Status	8767	2474.0	88.8	38.7	32.4	29.0
Economic Disadvantage Status	385197	2441.7	85.5	51.8	31.3	16.9

TABLE 5-52 GRADE 5 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	686807	2495.4	100.4	45.6	29.3	25.1
Female	336030	2493.8	95.8	46.5	30.3	23.2
Male	350511	2496.9	104.6	44.7	28.3	27.0
American Indian or Alaska Native	8369	2445.5	91.2	67.0	23.2	9.8
Asian	55939	2569.5	98.0	19.2	26.7	54.1
Black/African American	35520	2446.1	94.5	65.6	23.7	10.7
Native Hawaiian or Pacific Islander	7455	2474.1	92.6	54.1	29.5	16.4
Hispanic/Latino Ethnicity	290890	2464.2	90.5	58.9	27.7	13.5
White	230620	2523.8	95.0	32.6	32.7	34.7
Two or More Races	44482	2503.7	100.4	42.3	29.5	28.2
Unidentified Race	13532	2521.4	95.5	33.1	32.5	34.4
LEP Status	125007	2429.1	83.5	75.3	19.4	5.4
IDEA Indicator	80905	2412.1	97.8	78.0	14.7	7.3
Section 504 Status	10381	2498.6	96.1	44.4	30.7	24.9
Economic Disadvantage Status	391382	2464.2	92.0	58.7	27.4	13.8

TABLE 5-53 GRADE 6 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	669092	2517.1	116.5	42.5	32.6	24.9
Female	327519	2522.4	111.1	40.2	34.6	25.2
Male	341319	2512.0	121.3	44.7	30.8	24.6
American Indian or Alaska Native	8274	2458.5	112.3	63.6	27.0	9.4
Asian	56448	2603.3	109.4	17.0	28.1	54.9
Black/African American	34562	2460.1	115.1	62.2	27.1	10.7
Native Hawaiian or Pacific Islander	8469	2486.3	113.9	52.5	32.1	15.5
Hispanic/Latino Ethnicity	279463	2481.0	108.4	55.9	30.6	13.6
White	227422	2548.7	105.9	29.4	37.2	33.4
Two or More Races	40857	2523.8	118.1	39.9	32.9	27.2
Unidentified Race	13597	2551.1	106.9	29.5	35.1	35.4
LEP Status	99520	2421.9	102.4	79.2	16.9	4.0
IDEA Indicator	77035	2402.8	115.2	80.8	14.0	5.2
Section 504 Status	11007	2520.4	108.4	41.2	35.5	23.3
Economic Disadvantage Status	375605	2480.3	110.3	55.9	30.5	13.6

TABLE 5-54 GRADE 7 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	662872	2534.0	120.3	42.6	30.8	26.6
Female	323947	2536.5	115.7	41.8	31.9	26.3
Male	338669	2531.6	124.6	43.3	29.7	27.0
American Indian or Alaska Native	8287	2474.1	109.5	63.9	26.0	10.1
Asian	57553	2627.7	116.9	16.6	23.8	59.7
Black/African American	34620	2473.5	113.2	63.8	25.3	10.9
Native Hawaiian or Pacific Islander	8211	2493.3	113.1	55.5	30.6	13.8
Hispanic/Latino Ethnicity	271849	2493.0	109.4	57.1	29.0	13.9
White	228928	2568.1	109.4	28.8	35.5	35.7
Two or More Races	38959	2542.9	120.9	39.3	31.3	29.4
Unidentified Race	14465	2568.7	112.7	29.7	32.6	37.7
LEP Status	83312	2426.1	101.9	82.0	13.9	4.0
IDEA Indicator	72946	2414.8	110.9	82.2	13.0	4.8
Section 504 Status	12142	2542.0	110.9	39.8	34.0	26.2
Economic Disadvantage Status	362852	2494.0	111.5	56.5	29.1	14.4

TABLE 5-55 GRADE 8 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	656012	2548.2	130.3	43.2	31.0	25.8
Female	320225	2554.2	125.2	40.8	32.7	26.5
Male	335533	2542.5	134.7	45.4	29.4	25.2
American Indian or Alaska Native	7966	2483.3	116.5	64.7	25.5	9.8
Asian	57186	2651.1	127.7	17.0	24.1	58.9
Black/African American	34725	2483.0	120.4	64.1	25.0	10.9
Native Hawaiian or Pacific Islander	7650	2505.9	120.7	55.8	30.2	13.9
Hispanic/Latino Ethnicity	268359	2506.5	117.7	56.5	29.4	14.1
White	228629	2581.7	121.4	30.9	35.4	33.7
Two or More Races	36545	2553.7	129.8	40.9	32.4	26.7
Unidentified Race	14952	2585.7	124.5	30.7	32.5	36.8
LEP Status	68919	2433.7	108.5	81.4	14.3	4.4
IDEA Indicator	70483	2423.8	112.4	82.5	13.0	4.4
Section 504 Status	13116	2547.9	120.0	43.4	33.4	23.2
Economic Disadvantage Status	353959	2506.6	119.9	56.4	29.1	14.4

TABLE 5-56 GRADE 11 - MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	548789	2568.6	131.2	51.0	27.6	21.4
Female	269234	2574.2	124.6	49.0	29.4	21.5
Male	279357	2563.1	137.0	52.9	25.7	21.3
American Indian or Alaska Native	5250	2512.8	120.1	69.8	20.9	9.3
Asian	52751	2677.6	130.6	20.6	24.6	54.8
Black/African American	28643	2505.6	118.8	70.7	21.4	8.0
Native Hawaiian or Pacific Islander	5937	2528.6	118.9	64.0	26.0	10.1
Hispanic/Latino Ethnicity	242043	2530.0	117.6	63.4	25.5	11.2
White	171293	2598.6	124.3	40.2	32.2	27.6
Two or More Races	26719	2574.9	128.3	49.3	28.4	22.3
Unidentified Race	16153	2606.4	123.4	36.8	32.2	30.9
LEP Status	49151	2451.5	110.4	86.9	9.2	3.9
IDEA Indicator	48890	2438.8	107.2	89.1	8.2	2.6
Section 504 Status	11298	2568.0	122.1	52.1	28.9	19.0
Economic Disadvantage Status	290840	2531.5	121.2	63.1	24.8	12.1

TABLE 5-57 GRADE 3 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	586531	2418.4	97.7	23.5	51.3	25.1
Female	285892	2416.3	96.0	24.3	51.7	23.9
Male	300391	2420.4	99.2	22.8	50.9	26.3
American Indian or Alaska Native	7133	2377.4	91.0	41.3	47.4	11.3
Asian	47845	2475.6	93.4	10.1	40.8	49.1
Black/African American	31009	2375.5	95.3	36.2	51.9	11.9
Native Hawaiian or Pacific Islander	6225	2392.8	94.2	35.5	48.4	16.1
Hispanic/Latino Ethnicity	259812	2393.7	93.6	29.2	55.2	15.6
White	182247	2445.7	90.7	15.8	49.7	34.5
Two or More Races	39748	2423.8	96.6	25.6	47.1	27.3
Unidentified Race	12512	2438.2	94.3	16.2	50.4	33.4
LEP Status	143137	2371.9	91.5	37.2	53.2	9.6
IDEA Indicator	63805	2356.5	102.8	43.0	46.0	11.0
Section 504 Status	4790	2417.9	96.3	22.9	52.9	24.2
Economic Disadvantage Status	340896	2392.0	93.7	30.8	54.0	15.2

TABLE 5-58 GRADE 4 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	675686	2461.0	100.1	29.8	48.4	21.7
Female	329957	2458.9	97.5	30.4	49.3	20.3
Male	345474	2463.0	102.5	29.3	47.6	23.1
American Indian or Alaska Native	8300	2418.0	92.0	49.1	42.3	8.6
Asian	54649	2521.7	96.4	12.7	41.8	45.5
Black/African American	34958	2415.5	94.1	45.8	45.5	8.6
Native Hawaiian or Pacific Islander	7263	2433.3	92.3	42.4	45.6	12.0
Hispanic/Latino Ethnicity	282579	2430.5	94.1	38.8	49.9	11.4
White	229289	2490.9	93.1	19.9	49.2	30.9
Two or More Races	46043	2469.8	98.7	29.0	46.7	24.3
Unidentified Race	12605	2478.9	95.4	21.4	51.4	27.2
LEP Status	140080	2399.0	90.0	51.5	43.7	4.8
IDEA Indicator	78268	2389.6	102.3	56.0	36.4	7.7
Section 504 Status	8767	2466.8	98.3	28.9	48.0	23.1
Economic Disadvantage Status	385197	2431.0	94.3	39.5	48.9	11.6

TABLE 5-59 GRADE 5 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	686807	2474.8	121.3	37.7	42.2	20.1
Female	336030	2475.0	118.5	37.6	43.1	19.3
Male	350511	2474.7	124.0	37.8	41.3	20.9
American Indian or Alaska Native	8369	2428.9	110.0	55.5	36.9	7.7
Asian	55939	2547.5	111.0	16.9	38.5	44.6
Black/African American	35520	2420.5	117.0	56.6	36.2	7.2
Native Hawaiian or Pacific Islander	7455	2455.5	104.4	45.7	42.4	11.9
Hispanic/Latino Ethnicity	290890	2434.8	119.3	50.9	39.6	9.5
White	230620	2513.9	106.7	23.8	47.1	29.1
Two or More Races	44482	2490.5	111.4	33.2	43.7	23.1
Unidentified Race	13532	2501.0	115.6	27.8	45.2	27.0
LEP Status	125007	2388.5	112.4	69.1	28.0	2.9
IDEA Indicator	80905	2386.9	120.3	69.5	24.8	5.7
Section 504 Status	10381	2483.0	115.9	34.7	44.9	20.4
Economic Disadvantage Status	391382	2437.9	117.3	50.1	40.0	9.9

TABLE 5-60 GRADE 6 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	669092	2501.7	123.5	37.7	43.1	19.1
Female	327519	2503.4	119.4	37.1	44.5	18.4
Male	341319	2500.0	127.3	38.4	41.8	19.8
American Indian or Alaska Native	8274	2441.4	119.9	58.4	34.6	7.0
Asian	56448	2584.1	113.9	15.5	39.2	45.3
Black/African American	34562	2438.6	121.5	58.6	34.8	6.7
Native Hawaiian or Pacific Islander	8469	2469.1	116.7	48.4	41.5	10.2
Hispanic/Latino Ethnicity	279463	2464.2	116.3	49.6	41.5	8.9
White	227422	2536.8	112.9	25.4	47.6	27.0
Two or More Races	40857	2510.3	123.1	35.5	42.9	21.6
Unidentified Race	13597	2533.2	113.0	26.1	48.0	25.9
LEP Status	99520	2402.8	111.6	72.8	24.9	2.4
IDEA Indicator	77035	2385.7	122.4	75.9	20.2	3.9
Section 504 Status	11007	2507.7	114.9	35.3	46.3	18.4
Economic Disadvantage Status	375605	2463.8	117.6	49.9	40.9	9.2

TABLE 5-61 GRADE 7 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	662872	2513.7	135.6	34.1	44.4	21.5
Female	323947	2514.6	133.4	33.7	45.1	21.2
Male	338669	2512.9	137.6	34.5	43.7	21.9
American Indian or Alaska Native	8287	2453.1	124.0	51.5	40.8	7.8
Asian	57553	2606.2	127.5	13.6	36.6	49.7
Black/African American	34620	2446.6	125.3	54.0	38.5	7.5
Native Hawaiian or Pacific Islander	8211	2471.6	125.7	44.7	44.6	10.7
Hispanic/Latino Ethnicity	271849	2467.9	126.0	46.7	43.3	10.0
White	228928	2554.5	124.3	21.5	48.2	30.2
Two or More Races	38959	2527.4	132.4	30.0	45.8	24.2
Unidentified Race	14465	2544.5	130.2	24.1	47.2	28.7
LEP Status	83312	2401.6	113.5	69.7	27.8	2.4
IDEA Indicator	72946	2399.6	118.6	70.8	25.3	3.8
Section 504 Status	12142	2523.7	128.5	30.5	47.7	21.8
Economic Disadvantage Status	362852	2470.2	126.8	46.1	43.3	10.6

TABLE 5-62 GRADE 8 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	656012	2533.4	139.5	38.3	38.9	22.8
Female	320225	2537.2	137.1	36.1	40.9	23.0
Male	335533	2529.9	141.6	40.4	37.0	22.6
American Indian or Alaska Native	7966	2473.0	123.2	57.0	34.4	8.6
Asian	57186	2631.8	137.8	15.5	33.0	51.4
Black/African American	34725	2467.8	124.8	59.1	32.4	8.6
Native Hawaiian or Pacific Islander	7650	2489.1	128.0	48.4	40.4	11.2
Hispanic/Latino Ethnicity	268359	2487.4	126.9	51.8	37.0	11.2
White	228629	2572.8	130.7	25.3	43.3	31.4
Two or More Races	36545	2542.8	138.4	34.3	41.0	24.7
Unidentified Race	14952	2567.1	134.2	27.9	40.8	31.4
LEP Status	68919	2409.6	110.1	78.4	19.0	2.6
IDEA Indicator	70483	2413.4	114.0	78.2	18.2	3.6
Section 504 Status	13116	2535.7	130.9	36.8	41.9	21.2
Economic Disadvantage Status	353959	2489.2	128.4	51.1	37.1	11.8

TABLE 5-63 GRADE 11 - MATHEMATICS AVERAGE CLAIM 2&4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	548789	2539.2	156.1	39.7	44.5	15.8
Female	269234	2537.9	150.7	39.3	46.3	14.4
Male	279357	2540.3	161.1	40.0	42.8	17.1
American Indian or Alaska Native	5250	2486.9	151.4	52.3	40.4	7.3
Asian	52751	2641.6	153.0	17.3	42.1	40.6
Black/African American	28643	2463.7	140.4	60.7	34.7	4.6
Native Hawaiian or Pacific Islander	5937	2483.9	154.3	51.8	41.4	6.8
Hispanic/Latino Ethnicity	242043	2496.2	141.8	50.8	42.0	7.2
White	171293	2579.5	151.4	28.2	49.8	22.0
Two or More Races	26719	2551.9	158.8	35.2	47.0	17.8
Unidentified Race	16153	2571.1	147.2	30.2	49.7	20.1
LEP Status	49151	2414.4	126.2	75.1	22.8	2.1
IDEA Indicator	48890	2409.8	128.5	76.7	21.2	2.1
Section 504 Status	11298	2544.1	152.4	38.5	46.0	15.5
Economic Disadvantage Status	290840	2498.6	146.5	50.1	41.6	8.2

TABLE 5-64 GRADE 3 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	586531	2422.1	96.4	22.8	51.2	26.0
Female	285892	2424.0	93.9	21.9	52.3	25.9
Male	300391	2420.3	98.6	23.8	50.1	26.1
American Indian or Alaska Native	7133	2377.5	90.8	40.5	48.4	11.1
Asian	47845	2482.4	92.9	9.0	39.0	51.9
Black/African American	31009	2383.2	92.1	36.6	50.5	12.9
Native Hawaiian or Pacific Islander	6225	2399.6	92.9	31.9	50.9	17.2
Hispanic/Latino Ethnicity	259812	2399.6	90.5	28.1	55.3	16.6
White	182247	2445.2	92.3	16.1	49.3	34.6
Two or More Races	39748	2425.7	99.4	22.6	49.2	28.2
Unidentified Race	12512	2444.7	92.3	15.8	49.2	35.0
LEP Status	143137	2379.7	88.6	34.4	54.9	10.7
IDEA Indicator	63805	2364.1	96.7	45.4	43.8	10.8
Section 504 Status	4790	2418.2	95.8	24.0	51.7	24.2
Economic Disadvantage Status	340896	2397.5	91.3	29.6	54.2	16.2

TABLE 5-65 GRADE 4 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	675686	2462.0	99.6	30.4	45.7	23.9
Female	329957	2461.4	96.8	30.4	46.8	22.8
Male	345474	2462.6	102.3	30.3	44.7	25.0
American Indian or Alaska Native	8300	2414.8	91.9	51.0	39.9	9.1
Asian	54649	2526.6	95.2	12.1	38.0	50.0
Black/African American	34958	2417.9	94.6	46.9	42.7	10.4
Native Hawaiian or Pacific Islander	7263	2434.9	94.8	41.2	44.7	14.1
Hispanic/Latino Ethnicity	282579	2433.4	92.2	39.4	47.3	13.2
White	229289	2488.7	94.5	20.6	46.6	32.8
Two or More Races	46043	2468.9	101.0	29.1	44.4	26.5
Unidentified Race	12605	2483.3	96.0	22.0	45.8	32.2
LEP Status	140080	2405.2	87.2	51.7	42.5	5.9
IDEA Indicator	78268	2395.1	98.6	57.8	33.8	8.4
Section 504 Status	8767	2464.1	100.1	29.3	46.6	24.1
Economic Disadvantage Status	385197	2433.2	93.1	40.1	46.6	13.4

TABLE 5-66 GRADE 5 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	686807	2482.4	112.5	36.9	44.7	18.4
Female	336030	2482.3	110.4	36.8	45.5	17.7
Male	350511	2482.5	114.5	37.0	44.0	19.0
American Indian or Alaska Native	8369	2430.5	103.4	56.0	37.9	6.1
Asian	55939	2554.2	108.5	15.7	42.2	42.1
Black/African American	35520	2433.8	103.7	55.3	37.9	6.9
Native Hawaiian or Pacific Islander	7455	2456.8	106.4	45.6	43.6	10.8
Hispanic/Latino Ethnicity	290890	2448.8	103.0	48.4	42.8	8.8
White	230620	2514.2	107.9	24.9	48.9	26.2
Two or More Races	44482	2491.6	114.2	33.7	45.1	21.3
Unidentified Race	13532	2509.4	106.0	26.6	48.2	25.2
LEP Status	125007	2410.1	94.3	64.5	32.8	2.7
IDEA Indicator	80905	2403.3	104.9	68.3	26.8	4.9
Section 504 Status	10381	2485.0	110.7	35.3	46.6	18.1
Economic Disadvantage Status	391382	2449.2	104.2	48.3	42.6	9.1

TABLE 5-67 GRADE 6 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	669092	2508.8	119.7	35.8	43.8	20.4
Female	327519	2513.4	116.7	34.1	45.0	20.9
Male	341319	2504.4	122.4	37.5	42.6	19.9
American Indian or Alaska Native	8274	2451.7	110.5	55.2	37.3	7.5
Asian	56448	2591.5	117.5	14.8	38.0	47.1
Black/African American	34562	2455.3	110.4	54.6	37.4	7.9
Native Hawaiian or Pacific Islander	8469	2477.9	111.2	45.8	42.6	11.6
Hispanic/Latino Ethnicity	279463	2473.1	109.5	47.1	42.6	10.3
White	227422	2540.2	113.7	24.4	47.7	27.9
Two or More Races	40857	2516.9	121.8	33.3	43.9	22.8
Unidentified Race	13597	2541.0	112.2	24.9	46.0	29.1
LEP Status	99520	2419.4	97.4	68.6	28.9	2.6
IDEA Indicator	77035	2410.7	106.5	72.5	23.5	4.0
Section 504 Status	11007	2509.6	114.0	35.0	46.6	18.4
Economic Disadvantage Status	375605	2472.7	110.6	47.3	42.2	10.5

TABLE 5-68 GRADE 7 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	662872	2524.7	129.5	27.4	52.1	20.4
Female	323947	2529.8	126.5	25.6	53.7	20.8
Male	338669	2519.7	132.2	29.3	50.7	20.1
American Indian or Alaska Native	8287	2459.3	120.7	45.6	47.6	6.8
Asian	57553	2616.1	123.1	10.0	41.9	48.1
Black/African American	34620	2468.2	120.5	43.7	48.4	7.9
Native Hawaiian or Pacific Islander	8211	2484.3	125.2	37.5	52.3	10.3
Hispanic/Latino Ethnicity	271849	2486.3	118.6	36.6	53.3	10.1
White	228928	2556.2	123.4	18.5	54.0	27.5
Two or More Races	38959	2531.9	132.7	25.1	52.1	22.8
Unidentified Race	14465	2559.8	121.4	17.9	53.1	28.9
LEP Status	83312	2427.3	109.2	55.5	41.8	2.7
IDEA Indicator	72946	2418.1	113.9	61.6	35.0	3.4
Section 504 Status	12142	2528.6	124.8	25.3	54.9	19.7
Economic Disadvantage Status	362852	2486.1	121.0	36.9	52.7	10.4

TABLE 5-69 GRADE 8 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	656012	2539.4	139.1	31.6	46.9	21.5
Female	320225	2550.6	133.4	27.8	49.5	22.7
Male	335533	2528.7	143.5	35.2	44.4	20.4
American Indian or Alaska Native	7966	2469.0	132.0	51.9	40.3	7.8
Asian	57186	2641.6	134.0	11.8	36.5	51.6
Black/African American	34725	2478.2	128.6	48.7	43.0	8.3
Native Hawaiian or Pacific Islander	7650	2497.2	135.7	42.1	46.3	11.7
Hispanic/Latino Ethnicity	268359	2500.3	125.2	41.0	48.2	10.8
White	228629	2569.6	135.0	22.8	48.7	28.5
Two or More Races	36545	2543.3	143.9	29.9	47.1	23.0
Unidentified Race	14952	2578.0	129.4	21.2	48.4	30.4
LEP Status	68919	2430.9	114.7	64.1	32.8	3.1
IDEA Indicator	70483	2418.1	119.5	70.1	26.6	3.3
Section 504 Status	13116	2534.0	133.8	31.9	49.1	19.0
Economic Disadvantage Status	353959	2499.2	128.9	41.6	47.2	11.2

TABLE 5-70 GRADE 11 - MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 ADMINISTRATION

Group	Scale Scores			Reporting Categories (%)		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	548789	2559.2	147.0	28.3	55.3	16.3
Female	269234	2562.8	141.6	26.4	58.0	15.5
Male	279357	2555.7	152.0	30.1	52.7	17.1
American Indian or Alaska Native	5250	2506.8	140.5	39.2	54.1	6.7
Asian	52751	2662.8	144.5	10.8	45.4	43.8
Black/African American	28643	2500.4	131.2	42.3	52.3	5.3
Native Hawaiian or Pacific Islander	5937	2515.7	141.5	36.4	56.1	7.4
Hispanic/Latino Ethnicity	242043	2521.8	132.5	35.8	56.4	7.7
White	171293	2588.9	146.4	21.3	56.9	21.8
Two or More Races	26719	2563.3	152.0	26.8	55.9	17.3
Unidentified Race	16153	2595.8	137.3	18.9	58.8	22.3
LEP Status	49151	2461.2	123.6	52.5	45.0	2.6
IDEA Indicator	48890	2452.3	123.1	57.5	40.3	2.2
Section 504 Status	11298	2557.8	145.4	28.3	56.2	15.5
Economic Disadvantage Status	290840	2523.4	136.9	35.7	55.6	8.7

Percentile Tables for Overall Scale Scores

Table 5-71 through Table 5-77 present the overall ELA/literacy scale score for the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles for grades 3 through 8 and 11.

Table 5-78 through Table 5-84 present scores for these same percentiles for the overall mathematics scale scores for grades 3 through 8 and 11. These results are presented at the aggregate level and disaggregated by gender, by race/ethnicity, and by various status flags: limited English proficiency, IDEA indicator, Section 504, and economically disadvantaged.

TABLE 5-71 GRADE 3 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2298	2335	2366	2393	2420	2445	2471	2500	2536
Female	2307	2346	2376	2404	2430	2454	2480	2508	2544
Male	2290	2325	2356	2383	2410	2436	2462	2491	2529
American Indian or Alaska Native	2272	2300	2323	2345	2367	2390	2417	2444	2483
Asian	2345	2394	2428	2455	2479	2502	2524	2548	2581
Black/African American	2275	2302	2325	2347	2369	2394	2420	2450	2491
Native Hawaiian or Pacific Islander	2283	2314	2339	2364	2388	2412	2439	2470	2507
Hispanic/Latino Ethnicity	2285	2316	2343	2367	2390	2414	2439	2468	2506
White	2323	2365	2397	2424	2447	2470	2493	2517	2549
Two or More Races	2300	2339	2371	2399	2426	2452	2478	2506	2543
Unidentified Race	2322	2368	2401	2428	2452	2475	2499	2523	2556
LEP Status	2273	2300	2322	2342	2362	2383	2405	2432	2469
IDEA Indicator	2253	2277	2296	2316	2337	2360	2389	2426	2476
Section 504 Status	2302	2339	2369	2395	2419	2444	2469	2498	2533
Economic Disadvantage Status	2283	2315	2340	2364	2387	2411	2437	2465	2503

Table 5-72 Grade 4 ELA/literacy Overall Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2332	2372	2405	2436	2464	2489	2515	2544	2582
Female	2343	2385	2419	2448	2474	2499	2524	2552	2590
Male	2322	2361	2394	2424	2452	2480	2506	2535	2574
American Indian or Alaska Native	2303	2335	2360	2382	2408	2432	2458	2488	2525
Asian	2385	2440	2476	2503	2526	2548	2571	2597	2631
Black/African American	2305	2335	2359	2381	2405	2431	2461	2492	2533
Native Hawaiian or Pacific Islander	2310	2346	2375	2401	2429	2455	2483	2513	2548
Hispanic/Latino Ethnicity	2316	2352	2381	2408	2433	2458	2484	2511	2548
White	2360	2405	2439	2467	2491	2513	2537	2562	2596
Two or More Races	2337	2379	2413	2443	2471	2497	2523	2551	2589
Unidentified Race	2360	2410	2447	2476	2501	2524	2548	2574	2608
LEP Status	2296	2326	2349	2370	2390	2411	2433	2460	2495
IDEA Indicator	2274	2301	2323	2343	2365	2390	2420	2459	2510
Section 504 Status	2339	2377	2408	2436	2462	2487	2511	2539	2579
Economic Disadvantage Status	2314	2349	2377	2403	2429	2454	2480	2508	2545

Table 5-73 Grade 5 ELA/literacy Overall Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2364	2408	2442	2472	2500	2527	2555	2584	2622
Female	2381	2424	2458	2487	2514	2540	2566	2594	2632
Male	2352	2394	2427	2458	2486	2514	2542	2573	2611
American Indian or Alaska Native	2324	2363	2390	2414	2439	2465	2494	2527	2566
Asian	2417	2473	2511	2541	2567	2590	2614	2641	2676
Black/African American	2335	2371	2398	2421	2446	2472	2500	2533	2573
Native Hawaiian or Pacific Islander	2343	2385	2418	2445	2472	2497	2524	2554	2592
Hispanic/Latino Ethnicity	2345	2385	2415	2441	2466	2492	2518	2548	2587
White	2398	2444	2478	2507	2531	2554	2577	2602	2636
Two or More Races	2374	2418	2452	2483	2511	2537	2563	2592	2630
Unidentified Race	2397	2449	2486	2516	2540	2565	2587	2612	2648
LEP Status	2316	2350	2374	2395	2415	2434	2456	2481	2516
IDEA Indicator	2293	2324	2348	2371	2392	2415	2443	2479	2533
Section 504 Status	2372	2411	2443	2472	2500	2525	2551	2579	2617
Economic Disadvantage Status	2345	2384	2413	2439	2465	2490	2517	2546	2584

Table 5-74 Grade 6 ELA/literacy Overall Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2390	2435	2470	2498	2525	2550	2577	2606	2643
Female	2408	2453	2487	2514	2539	2564	2589	2616	2652
Male	2377	2420	2453	2483	2510	2536	2564	2594	2632
American Indian or Alaska Native	2356	2390	2419	2444	2469	2494	2519	2551	2592
Asian	2453	2506	2542	2569	2594	2615	2638	2663	2697
Black/African American	2358	2390	2417	2442	2467	2494	2521	2552	2592
Native Hawaiian or Pacific Islander	2371	2413	2445	2471	2495	2520	2546	2574	2610
Hispanic/Latino Ethnicity	2372	2413	2444	2471	2495	2519	2544	2572	2609
White	2421	2466	2500	2526	2550	2574	2597	2622	2655
Two or More Races	2399	2446	2480	2508	2534	2560	2585	2614	2650
Unidentified Race	2431	2483	2516	2545	2568	2589	2612	2635	2668
LEP Status	2337	2368	2391	2411	2431	2451	2472	2496	2529
IDEA Indicator	2322	2350	2372	2393	2414	2437	2463	2494	2540
Section 504 Status	2405	2446	2476	2501	2523	2546	2571	2599	2635
Economic Disadvantage Status	2370	2409	2440	2466	2491	2515	2540	2568	2606

Table 5-75 Grade 7 ELA/literacy Overall Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2404	2455	2493	2525	2553	2578	2604	2634	2673
Female	2426	2476	2513	2543	2568	2592	2617	2645	2682
Male	2388	2436	2474	2507	2536	2564	2591	2621	2661
American Indian or Alaska Native	2370	2405	2439	2467	2494	2521	2547	2577	2618
Asian	2476	2536	2571	2599	2623	2645	2667	2692	2725
Black/African American	2366	2402	2434	2462	2489	2518	2547	2578	2620
Native Hawaiian or Pacific Islander	2381	2424	2459	2491	2519	2543	2569	2598	2635
Hispanic/Latino Ethnicity	2385	2428	2463	2493	2520	2545	2570	2599	2636
White	2439	2491	2526	2554	2579	2601	2624	2650	2685
Two or More Races	2417	2467	2505	2537	2563	2587	2613	2642	2680
Unidentified Race	2449	2506	2542	2569	2593	2615	2639	2664	2698
LEP Status	2345	2376	2399	2421	2442	2463	2486	2513	2548
IDEA Indicator	2334	2364	2388	2410	2432	2457	2484	2517	2563
Section 504 Status	2424	2470	2504	2530	2554	2578	2603	2630	2668
Economic Disadvantage Status	2382	2424	2459	2488	2515	2541	2567	2596	2633

TABLE 5-76 GRADE 8 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2426	2472	2506	2537	2566	2594	2621	2651	2689
Female	2450	2494	2528	2558	2585	2610	2636	2664	2700
Male	2409	2452	2487	2517	2547	2576	2605	2636	2676
American Indian or Alaska Native	2384	2423	2453	2480	2505	2532	2561	2595	2639
Asian	2488	2546	2584	2613	2638	2660	2682	2708	2742
Black/African American	2391	2428	2457	2483	2508	2535	2565	2597	2640
Native Hawaiian or Pacific Islander	2401	2443	2473	2500	2526	2553	2581	2612	2652
Hispanic/Latino Ethnicity	2407	2449	2480	2507	2533	2559	2586	2617	2655
White	2455	2503	2537	2567	2593	2616	2640	2667	2700
Two or More Races	2433	2481	2516	2547	2575	2601	2628	2657	2693
Unidentified Race	2467	2517	2555	2583	2609	2633	2654	2679	2713
LEP Status	2361	2393	2416	2437	2455	2474	2496	2521	2557
IDEA Indicator	2353	2384	2408	2428	2449	2471	2495	2527	2575
Section 504 Status	2437	2479	2509	2535	2562	2588	2615	2644	2681
Economic Disadvantage Status	2404	2445	2476	2503	2530	2556	2583	2614	2653

TABLE 5-77 GRADE 11 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2438	2499	2544	2581	2613	2642	2672	2705	2746
Female	2465	2524	2566	2599	2628	2656	2684	2714	2754
Male	2418	2476	2523	2562	2596	2628	2660	2694	2737
American Indian or Alaska Native	2409	2457	2495	2528	2560	2589	2620	2653	2697
Asian	2505	2579	2623	2657	2685	2711	2737	2765	2795
Black/African American	2394	2441	2480	2517	2550	2583	2615	2650	2696
Native Hawaiian or Pacific Islander	2409	2459	2496	2528	2560	2591	2620	2651	2696
Hispanic/Latino Ethnicity	2420	2473	2515	2550	2581	2610	2639	2671	2712
White	2475	2541	2584	2617	2645	2672	2698	2726	2763
Two or More Races	2461	2525	2567	2600	2629	2656	2683	2712	2752
Unidentified Race	2490	2554	2595	2627	2652	2677	2700	2728	2763
LEP Status	2362	2394	2419	2442	2464	2488	2514	2544	2585
IDEA Indicator	2362	2396	2422	2447	2473	2501	2532	2568	2619
Section 504 Status	2448	2503	2545	2580	2611	2640	2670	2701	2745
Economic Disadvantage Status	2416	2469	2511	2546	2578	2608	2638	2670	2712

TABLE 5-78 GRADE 3 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2323	2360	2386	2409	2430	2451	2473	2498	2533
Female	2325	2360	2386	2407	2428	2448	2470	2494	2528
Male	2320	2360	2387	2410	2432	2453	2476	2501	2537
American Indian or Alaska Native	2290	2326	2349	2368	2387	2407	2428	2452	2486
Asian	2382	2422	2450	2473	2493	2513	2534	2559	2592
Black/African American	2281	2319	2344	2365	2384	2404	2425	2450	2483
Native Hawaiian or Pacific Islander	2307	2343	2369	2390	2410	2431	2452	2475	2506
Hispanic/Latino Ethnicity	2311	2346	2369	2390	2409	2428	2448	2471	2503
White	2348	2386	2411	2432	2452	2471	2490	2513	2544
Two or More Races	2322	2360	2386	2410	2432	2453	2476	2503	2539
Unidentified Race	2352	2392	2419	2441	2462	2482	2501	2525	2558
LEP Status	2298	2332	2354	2373	2391	2409	2429	2453	2485
IDEA Indicator	2245	2285	2314	2338	2360	2384	2410	2442	2486
Section 504 Status	2323	2359	2383	2406	2425	2446	2469	2494	2530
Economic Disadvantage Status	2305	2341	2365	2386	2406	2425	2445	2469	2501

TABLE 5-79 GRADE 4 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2358	2394	2422	2446	2469	2491	2515	2542	2578
Female	2362	2395	2421	2445	2466	2487	2510	2536	2571
Male	2355	2394	2423	2449	2472	2495	2520	2547	2584
American Indian or Alaska Native	2331	2361	2384	2405	2424	2444	2465	2491	2526
Asian	2419	2462	2492	2517	2538	2558	2580	2604	2637
Black/African American	2322	2355	2377	2397	2418	2438	2461	2487	2522
Native Hawaiian or Pacific Islander	2346	2379	2403	2426	2446	2468	2490	2515	2548
Hispanic/Latino Ethnicity	2344	2376	2399	2421	2441	2461	2483	2507	2541
White	2387	2426	2452	2474	2494	2515	2535	2559	2590
Two or More Races	2362	2399	2428	2452	2475	2497	2521	2549	2585
Unidentified Race	2386	2427	2456	2480	2502	2524	2546	2571	2601
LEP Status	2327	2357	2377	2395	2412	2431	2451	2474	2508
IDEA Indicator	2289	2324	2348	2369	2389	2412	2439	2472	2519
Section 504 Status	2366	2401	2427	2449	2472	2493	2517	2543	2582
Economic Disadvantage Status	2342	2374	2398	2420	2440	2461	2482	2507	2541

TABLE 5-80 GRADE 5 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2373	2410	2438	2465	2492	2518	2545	2576	2614
Female	2377	2412	2439	2465	2489	2514	2541	2571	2610
Male	2368	2407	2437	2466	2494	2522	2550	2580	2618
American Indian or Alaska Native	2343	2375	2400	2421	2442	2465	2490	2520	2561
Asian	2434	2483	2519	2547	2572	2595	2618	2643	2677
Black/African American	2339	2370	2392	2412	2432	2455	2479	2509	2550
Native Hawaiian or Pacific Islander	2363	2399	2423	2447	2470	2495	2521	2549	2586
Hispanic/Latino Ethnicity	2357	2390	2414	2436	2458	2481	2506	2535	2574
White	2402	2443	2474	2500	2524	2547	2569	2595	2627
Two or More Races	2379	2416	2446	2473	2500	2527	2554	2584	2622
Unidentified Race	2403	2447	2479	2507	2533	2557	2581	2607	2639
LEP Status	2336	2366	2387	2405	2422	2440	2461	2487	2524
IDEA Indicator	2308	2339	2362	2382	2401	2422	2448	2483	2537
Section 504 Status	2382	2417	2445	2471	2496	2521	2546	2576	2615
Economic Disadvantage Status	2356	2389	2413	2435	2457	2480	2506	2535	2574

TABLE 5-81 GRADE 6 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2366	2422	2461	2492	2519	2545	2572	2603	2644
Female	2377	2430	2467	2496	2522	2547	2573	2602	2642
Male	2356	2414	2456	2488	2516	2543	2572	2604	2647
American Indian or Alaska Native	2325	2371	2406	2438	2466	2492	2519	2547	2588
Asian	2459	2514	2550	2578	2604	2629	2655	2685	2729
Black/African American	2319	2361	2395	2426	2453	2478	2505	2535	2577
Native Hawaiian or Pacific Islander	2347	2395	2435	2466	2494	2518	2543	2572	2608
Hispanic/Latino Ethnicity	2344	2392	2429	2459	2485	2510	2535	2564	2601
White	2412	2466	2500	2525	2548	2570	2594	2620	2657
Two or More Races	2372	2429	2469	2499	2526	2552	2580	2610	2652
Unidentified Race	2418	2476	2511	2538	2564	2588	2612	2639	2678
LEP Status	2305	2345	2374	2399	2424	2448	2472	2500	2538
IDEA Indicator	2271	2313	2342	2368	2394	2423	2456	2493	2544
Section 504 Status	2384	2437	2471	2499	2523	2545	2571	2600	2641
Economic Disadvantage Status	2341	2390	2426	2457	2483	2508	2533	2562	2600

TABLE 5-82 GRADE 7 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2378	2431	2469	2502	2532	2562	2593	2628	2674
Female	2388	2439	2475	2505	2533	2561	2593	2626	2672
Male	2369	2424	2464	2499	2531	2562	2594	2630	2676
American Indian or Alaska Native	2343	2386	2419	2447	2474	2501	2530	2561	2605
Asian	2468	2530	2571	2605	2632	2659	2686	2716	2760
Black/African American	2335	2376	2407	2436	2460	2487	2515	2549	2597
Native Hawaiian or Pacific Islander	2357	2404	2438	2470	2498	2524	2552	2584	2625
Hispanic/Latino Ethnicity	2355	2401	2436	2465	2492	2519	2547	2580	2623
White	2421	2474	2510	2539	2565	2592	2618	2648	2686
Two or More Races	2386	2440	2479	2512	2541	2571	2602	2637	2682
Unidentified Race	2423	2482	2522	2554	2582	2610	2637	2670	2709
LEP Status	2314	2351	2377	2401	2424	2448	2473	2504	2549
IDEA Indicator	2296	2331	2357	2380	2404	2430	2460	2497	2554
Section 504 Status	2400	2452	2485	2515	2540	2567	2596	2629	2675
Economic Disadvantage Status	2354	2399	2435	2464	2491	2518	2546	2580	2623

TABLE 5-83 GRADE 8 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2386	2437	2475	2507	2539	2574	2611	2652	2704
Female	2398	2448	2485	2516	2547	2580	2615	2653	2703
Male	2375	2425	2465	2498	2531	2567	2606	2650	2706
American Indian or Alaska Native	2351	2392	2425	2452	2480	2507	2539	2577	2630
Asian	2471	2537	2587	2627	2660	2690	2721	2757	2802
Black/African American	2344	2384	2413	2441	2468	2494	2523	2562	2618
Native Hawaiian or Pacific Islander	2362	2411	2446	2477	2504	2532	2562	2600	2650
Hispanic/Latino Ethnicity	2362	2408	2443	2473	2499	2527	2560	2599	2649
White	2424	2476	2511	2543	2575	2606	2638	2672	2717
Two or More Races	2392	2443	2482	2514	2546	2580	2616	2657	2708
Unidentified Race	2429	2488	2528	2567	2601	2633	2665	2700	2747
LEP Status	2318	2354	2381	2405	2428	2451	2477	2509	2563
IDEA Indicator	2304	2339	2365	2389	2411	2436	2464	2500	2559
Section 504 Status	2403	2451	2484	2511	2540	2570	2606	2646	2698
Economic Disadvantage Status	2361	2407	2441	2470	2497	2525	2558	2597	2649

TABLE 5-84 GRADE 11 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2401	2453	2494	2531	2565	2599	2635	2677	2734
Female	2415	2464	2503	2538	2570	2602	2636	2674	2727
Male	2388	2443	2485	2523	2560	2596	2634	2679	2741
American Indian or Alaska Native	2370	2415	2450	2480	2508	2540	2573	2611	2663
Asian	2493	2564	2612	2651	2684	2716	2750	2786	2831
Black/African American	2357	2401	2436	2466	2496	2528	2561	2600	2652
Native Hawaiian or Pacific Islander	2380	2430	2464	2496	2526	2555	2585	2620	2667
Hispanic/Latino Ethnicity	2382	2429	2463	2494	2525	2555	2588	2625	2675
White	2437	2496	2538	2572	2603	2633	2665	2702	2752
Two or More Races	2414	2467	2507	2542	2574	2606	2641	2682	2737
Unidentified Race	2440	2498	2542	2578	2610	2641	2671	2705	2749
LEP Status	2328	2367	2394	2417	2439	2462	2488	2524	2582
IDEA Indicator	2318	2357	2384	2407	2430	2452	2479	2514	2570
Section 504 Status	2415	2464	2501	2534	2565	2594	2629	2669	2725
Economic Disadvantage Status	2380	2428	2463	2494	2526	2557	2590	2629	2681

Percentile Tables for Claim-level Scale Scores

- Table 5-85 through Table 5-91 present percentile information for the Claim 1 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5-92 through table 5-98 present percentile information for the Claim 2 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5-99 through Table 5-105 present percentile information for the Claim 3 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5-106 through table 5-112 present percentile information for the Claim 4 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5-113 through Table 5-119 present percentile information for the Claim 1 mathematics scale scores for grades 3 through 8 and 11.
- Table 5-120 through Table 5-126 present percentile information for the Claim 2/4 mathematics scale scores for grades 3 through 8 and 11.
- Table 5-127 through Table 5-133 present percentile information for the Claim 3 mathematics scale scores for grades 3 through 8 and 11.

TABLE 5-85 GRADE 3 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2288	2327	2357	2384	2411	2439	2469	2501	2545
Female	2296	2336	2366	2394	2421	2449	2477	2510	2552
Male	2281	2320	2348	2375	2401	2430	2460	2493	2536
American Indian or Alaska Native	2260	2296	2321	2342	2363	2387	2412	2447	2490
Asian	2329	2379	2414	2443	2469	2493	2518	2547	2591
Black/African American	2266	2302	2325	2345	2367	2388	2415	2448	2493
Native Hawaiian or Pacific Islander	2270	2308	2332	2355	2378	2403	2431	2465	2508
Hispanic/Latino Ethnicity	2271	2308	2334	2357	2380	2405	2432	2465	2509
White	2314	2356	2388	2417	2443	2468	2494	2523	2563
Two or More Races	2292	2332	2363	2391	2419	2447	2476	2509	2552
Unidentified Race	2314	2356	2387	2416	2444	2469	2493	2521	2561
LEP Status	2257	2293	2317	2336	2355	2376	2398	2427	2470
IDEA Indicator	2242	2279	2303	2323	2342	2363	2389	2426	2481
Section 504 Status	2290	2328	2358	2385	2412	2440	2470	2502	2546
Economic Disadvantage Status	2271	2308	2334	2357	2380	2404	2431	2463	2507

TABLE 5-86 GRADE 4 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2320	2367	2402	2433	2463	2490	2518	2549	2591
Female	2330	2377	2413	2443	2471	2498	2525	2555	2597
Male	2311	2358	2392	2423	2454	2482	2511	2543	2585
American Indian or Alaska Native	2281	2330	2360	2388	2412	2438	2467	2496	2541
Asian	2372	2427	2465	2494	2519	2543	2567	2595	2634
Black/African American	2290	2332	2358	2382	2407	2435	2466	2499	2542
Native Hawaiian or Pacific Islander	2295	2342	2372	2400	2427	2453	2482	2513	2555
Hispanic/Latino Ethnicity	2299	2345	2376	2403	2430	2456	2484	2515	2556
White	2350	2400	2437	2468	2493	2517	2542	2570	2608
Two or More Races	2328	2376	2412	2444	2473	2500	2527	2557	2598
Unidentified Race	2349	2398	2436	2468	2494	2520	2545	2572	2613
LEP Status	2271	2317	2345	2368	2389	2411	2435	2463	2502
IDEA Indicator	2247	2301	2331	2356	2379	2403	2432	2469	2522
Section 504 Status	2324	2372	2408	2437	2465	2491	2519	2550	2592
Economic Disadvantage Status	2299	2343	2374	2401	2428	2455	2482	2513	2555

TABLE 5-87 GRADE 5 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2352	2398	2433	2464	2494	2523	2553	2586	2630
Female	2365	2411	2447	2478	2507	2535	2564	2597	2640
Male	2342	2387	2420	2451	2481	2511	2541	2574	2619
American Indian or Alaska Native	2315	2360	2391	2415	2440	2467	2497	2533	2577
Asian	2400	2456	2495	2526	2554	2579	2605	2635	2677
Black/African American	2325	2364	2391	2415	2441	2467	2500	2534	2580
Native Hawaiian or Pacific Islander	2326	2373	2406	2434	2460	2488	2518	2551	2598
Hispanic/Latino Ethnicity	2331	2375	2405	2432	2458	2485	2513	2546	2590
White	2386	2435	2471	2502	2529	2554	2580	2609	2650
Two or More Races	2362	2408	2445	2477	2506	2534	2563	2595	2640
Unidentified Race	2382	2431	2469	2500	2527	2552	2578	2608	2650
LEP Status	2294	2341	2368	2390	2410	2430	2453	2480	2520
IDEA Indicator	2282	2329	2358	2380	2401	2424	2451	2488	2543
Section 504 Status	2356	2402	2436	2467	2498	2525	2553	2586	2630
Economic Disadvantage Status	2332	2375	2404	2431	2458	2485	2514	2547	2590

TABLE 5-88 GRADE 6 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2371	2419	2454	2485	2515	2544	2574	2607	2652
Female	2384	2432	2467	2498	2526	2554	2583	2615	2658
Male	2359	2407	2442	2473	2503	2533	2564	2599	2646
American Indian or Alaska Native	2335	2382	2413	2440	2465	2492	2521	2556	2603
Asian	2424	2480	2520	2551	2579	2605	2631	2662	2704
Black/African American	2337	2382	2410	2438	2463	2489	2519	2553	2599
Native Hawaiian or Pacific Islander	2346	2394	2426	2453	2480	2508	2537	2570	2615
Hispanic/Latino Ethnicity	2348	2395	2426	2454	2480	2507	2536	2569	2613
White	2403	2454	2489	2519	2546	2572	2599	2629	2670
Two or More Races	2380	2429	2465	2496	2525	2554	2584	2617	2662
Unidentified Race	2406	2458	2494	2525	2553	2578	2605	2635	2676
LEP Status	2300	2350	2378	2401	2421	2442	2463	2490	2528
IDEA Indicator	2290	2343	2373	2396	2418	2440	2465	2498	2551
Section 504 Status	2379	2428	2462	2491	2518	2545	2574	2607	2651
Economic Disadvantage Status	2348	2394	2425	2453	2479	2506	2535	2568	2612

TABLE 5-89 GRADE 7 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2393	2442	2481	2515	2546	2575	2605	2640	2687
Female	2410	2460	2498	2530	2559	2587	2616	2649	2695
Male	2381	2428	2465	2499	2531	2563	2595	2630	2679
American Indian or Alaska Native	2361	2405	2436	2463	2489	2518	2550	2584	2631
Asian	2448	2510	2552	2584	2612	2639	2667	2699	2744
Black/African American	2364	2406	2436	2464	2490	2518	2549	2584	2632
Native Hawaiian or Pacific Islander	2368	2411	2444	2475	2504	2536	2566	2600	2646
Hispanic/Latino Ethnicity	2369	2413	2447	2476	2505	2534	2564	2598	2644
White	2430	2482	2521	2551	2578	2604	2631	2662	2705
Two or More Races	2405	2455	2494	2528	2559	2587	2616	2650	2696
Unidentified Race	2428	2483	2522	2555	2582	2608	2636	2668	2711
LEP Status	2327	2368	2395	2416	2436	2457	2479	2508	2548
IDEA Indicator	2324	2365	2393	2416	2437	2460	2486	2521	2574
Section 504 Status	2402	2452	2489	2523	2552	2579	2607	2640	2688
Economic Disadvantage Status	2370	2415	2448	2477	2506	2535	2565	2598	2644

TABLE 5-90 GRADE 8 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2409	2458	2496	2530	2562	2591	2621	2654	2698
Female	2428	2478	2516	2547	2577	2605	2633	2665	2708
Male	2394	2442	2478	2512	2545	2577	2608	2642	2687
American Indian or Alaska Native	2373	2417	2448	2477	2505	2534	2565	2602	2650
Asian	2466	2528	2569	2600	2626	2651	2677	2706	2747
Black/African American	2374	2419	2450	2478	2505	2535	2567	2603	2650
Native Hawaiian or Pacific Islander	2381	2426	2458	2488	2518	2549	2581	2615	2659
Hispanic/Latino Ethnicity	2388	2434	2467	2497	2525	2554	2584	2618	2661
White	2438	2491	2530	2562	2590	2617	2643	2673	2714
Two or More Races	2419	2469	2508	2542	2572	2601	2629	2661	2703
Unidentified Race	2447	2500	2540	2571	2599	2625	2650	2681	2722
LEP Status	2334	2381	2409	2430	2450	2470	2493	2520	2562
IDEA Indicator	2333	2380	2408	2431	2452	2474	2500	2536	2590
Section 504 Status	2418	2464	2500	2533	2562	2590	2620	2650	2694
Economic Disadvantage Status	2387	2433	2465	2495	2524	2553	2584	2617	2661

TABLE 5-91 GRADE 11 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2434	2495	2538	2574	2607	2638	2670	2706	2754
Female	2456	2514	2554	2588	2619	2648	2679	2713	2760
Male	2416	2477	2522	2560	2595	2628	2661	2698	2747
American Indian or Alaska Native	2402	2458	2498	2532	2564	2596	2629	2664	2714
Asian	2490	2557	2602	2637	2668	2697	2726	2759	2795
Black/African American	2393	2447	2486	2521	2553	2586	2618	2656	2705
Native Hawaiian or Pacific Islander	2401	2454	2493	2527	2557	2590	2621	2655	2705
Hispanic/Latino Ethnicity	2413	2471	2511	2544	2575	2605	2636	2671	2718
White	2471	2535	2578	2613	2643	2671	2699	2732	2777
Two or More Races	2459	2519	2561	2595	2625	2654	2683	2716	2764
Unidentified Race	2477	2537	2577	2608	2636	2663	2692	2723	2766
LEP Status	2333	2387	2420	2446	2469	2493	2518	2548	2590
IDEA Indicator	2340	2395	2429	2457	2484	2512	2543	2581	2636
Section 504 Status	2443	2501	2544	2580	2612	2643	2675	2709	2757
Economic Disadvantage Status	2411	2467	2508	2542	2574	2604	2635	2671	2718

TABLE 5-92 GRADE 3 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2280	2328	2362	2393	2420	2447	2475	2507	2551
Female	2292	2340	2374	2404	2431	2458	2485	2516	2560
Male	2271	2317	2352	2381	2409	2436	2465	2497	2542
American Indian or Alaska Native	2217	2284	2316	2347	2373	2400	2429	2461	2508
Asian	2336	2389	2425	2454	2479	2505	2531	2561	2605
Black/African American	2245	2287	2317	2344	2371	2397	2426	2460	2503
Native Hawaiian or Pacific Islander	2233	2296	2335	2365	2393	2420	2450	2483	2525
Hispanic/Latino Ethnicity	2266	2309	2341	2368	2393	2418	2444	2475	2516
White	2305	2355	2391	2419	2445	2470	2497	2527	2570
Two or More Races	2280	2332	2368	2401	2430	2457	2486	2520	2565
Unidentified Race	2314	2363	2397	2427	2452	2476	2502	2530	2569
LEP Status	2244	2288	2317	2342	2365	2389	2414	2442	2482
IDEA Indicator	2203	2254	2285	2311	2338	2365	2396	2434	2489
Section 504 Status	2283	2333	2367	2395	2424	2448	2478	2511	2557
Economic Disadvantage Status	2260	2305	2336	2364	2390	2415	2443	2474	2516

TABLE 5-93 GRADE 4 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2314	2364	2400	2431	2460	2489	2519	2553	2602
Female	2328	2378	2415	2445	2474	2502	2531	2564	2613
Male	2302	2351	2387	2418	2447	2476	2506	2541	2590
American Indian or Alaska Native	2257	2309	2345	2375	2404	2433	2465	2500	2549
Asian	2372	2429	2468	2501	2529	2557	2587	2622	2663
Black/African American	2279	2321	2353	2380	2405	2433	2464	2499	2545
Native Hawaiian or Pacific Islander	2264	2322	2360	2393	2425	2455	2487	2524	2572
Hispanic/Latino Ethnicity	2300	2347	2380	2408	2434	2461	2489	2520	2563
White	2339	2391	2426	2457	2485	2512	2540	2572	2619
Two or More Races	2308	2363	2403	2437	2468	2499	2531	2566	2616
Unidentified Race	2355	2407	2444	2475	2503	2529	2557	2588	2632
LEP Status	2270	2314	2345	2370	2394	2417	2442	2472	2512
IDEA Indicator	2229	2279	2311	2338	2365	2392	2424	2465	2523
Section 504 Status	2310	2361	2397	2429	2458	2486	2519	2554	2604
Economic Disadvantage Status	2293	2340	2373	2402	2429	2456	2484	2516	2561

TABLE 5-94 GRADE 5 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2351	2404	2442	2475	2505	2534	2564	2599	2648
Female	2370	2422	2460	2491	2520	2548	2577	2611	2660
Male	2335	2389	2427	2459	2489	2519	2550	2585	2635
American Indian or Alaska Native	2287	2341	2381	2414	2441	2472	2504	2539	2592
Asian	2408	2470	2513	2546	2576	2604	2634	2669	2701
Black/African American	2315	2363	2397	2425	2453	2480	2510	2545	2590
Native Hawaiian or Pacific Islander	2304	2365	2408	2441	2474	2505	2539	2573	2623
Hispanic/Latino Ethnicity	2332	2383	2419	2449	2477	2503	2531	2563	2607
White	2381	2434	2472	2503	2530	2558	2586	2619	2666
Two or More Races	2350	2407	2446	2481	2513	2543	2575	2612	2664
Unidentified Race	2397	2455	2494	2526	2553	2578	2604	2635	2682
LEP Status	2291	2340	2372	2398	2423	2447	2473	2501	2541
IDEA Indicator	2249	2303	2337	2366	2393	2421	2453	2493	2552
Section 504 Status	2358	2406	2441	2474	2503	2531	2563	2597	2649
Economic Disadvantage Status	2327	2378	2414	2444	2472	2499	2527	2560	2605

TABLE 5-95 GRADE 6 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2377	2430	2467	2498	2525	2552	2581	2615	2666
Female	2398	2449	2485	2514	2540	2566	2595	2629	2679
Male	2362	2413	2450	2481	2510	2537	2566	2600	2651
American Indian or Alaska Native	2323	2380	2418	2449	2477	2504	2533	2568	2619
Asian	2445	2503	2539	2568	2595	2621	2648	2681	2724
Black/African American	2330	2377	2410	2440	2468	2495	2524	2557	2603
Native Hawaiian or Pacific Islander	2348	2403	2443	2475	2503	2532	2563	2599	2652
Hispanic/Latino Ethnicity	2360	2409	2444	2472	2498	2521	2546	2574	2615
White	2407	2458	2494	2524	2551	2577	2606	2639	2692
Two or More Races	2385	2442	2480	2512	2540	2569	2599	2634	2688
Unidentified Race	2424	2480	2515	2543	2567	2590	2613	2642	2682
LEP Status	2309	2356	2386	2412	2435	2458	2481	2508	2543
IDEA Indicator	2286	2331	2365	2390	2417	2443	2471	2506	2558
Section 504 Status	2395	2442	2477	2505	2529	2553	2582	2618	2674
Economic Disadvantage Status	2353	2403	2437	2466	2493	2518	2544	2575	2620

TABLE 5-96 GRADE 7 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2383	2447	2489	2522	2551	2579	2608	2641	2687
Female	2409	2469	2509	2540	2568	2594	2622	2654	2700
Male	2363	2427	2470	2504	2534	2563	2593	2626	2673
American Indian or Alaska Native	2328	2386	2427	2460	2493	2522	2553	2588	2637
Asian	2469	2531	2570	2599	2625	2649	2674	2705	2745
Black/African American	2316	2377	2418	2455	2485	2515	2547	2583	2629
Native Hawaiian or Pacific Islander	2350	2412	2454	2488	2519	2548	2577	2612	2660
Hispanic/Latino Ethnicity	2365	2427	2468	2500	2527	2552	2577	2607	2646
White	2412	2472	2511	2543	2572	2599	2626	2658	2706
Two or More Races	2392	2454	2495	2530	2560	2590	2620	2655	2703
Unidentified Race	2446	2507	2545	2573	2597	2621	2644	2673	2711
LEP Status	2299	2354	2391	2422	2449	2476	2502	2530	2567
IDEA Indicator	2281	2337	2374	2406	2436	2465	2496	2531	2582
Section 504 Status	2406	2463	2500	2531	2557	2583	2610	2643	2692
Economic Disadvantage Status	2353	2413	2456	2489	2518	2545	2572	2603	2646

TABLE 5-97 GRADE 8 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2409	2468	2506	2538	2566	2595	2626	2662	2712
Female	2438	2492	2528	2558	2585	2613	2642	2677	2725
Male	2387	2448	2487	2519	2548	2577	2608	2645	2696
American Indian or Alaska Native	2342	2410	2454	2487	2518	2547	2578	2615	2667
Asian	2481	2540	2579	2611	2639	2666	2695	2729	2769
Black/African American	2360	2416	2455	2484	2512	2539	2567	2603	2651
Native Hawaiian or Pacific Islander	2363	2430	2471	2505	2534	2564	2592	2630	2680
Hispanic/Latino Ethnicity	2389	2445	2482	2511	2537	2563	2590	2623	2667
White	2440	2497	2533	2563	2591	2618	2648	2682	2732
Two or More Races	2413	2475	2516	2549	2579	2608	2639	2675	2726
Unidentified Race	2464	2519	2553	2583	2608	2633	2659	2689	2732
LEP Status	2320	2373	2407	2436	2462	2485	2508	2536	2575
IDEA Indicator	2299	2357	2394	2424	2452	2480	2508	2543	2596
Section 504 Status	2417	2473	2510	2540	2567	2595	2626	2663	2715
Economic Disadvantage Status	2381	2439	2477	2506	2533	2560	2588	2621	2668

TABLE 5-98 GRADE 11 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2416	2485	2533	2573	2609	2644	2681	2723	2781
Female	2448	2513	2559	2596	2629	2663	2697	2738	2795
Male	2392	2461	2510	2551	2589	2625	2663	2706	2766
American Indian or Alaska Native	2367	2435	2479	2520	2552	2586	2622	2666	2728
Asian	2495	2574	2626	2666	2701	2735	2769	2795	2795
Black/African American	2361	2422	2467	2505	2541	2577	2613	2655	2712
Native Hawaiian or Pacific Islander	2378	2444	2487	2526	2561	2596	2631	2674	2732
Hispanic/Latino Ethnicity	2397	2460	2505	2543	2576	2608	2641	2679	2730
White	2453	2523	2571	2609	2643	2676	2710	2750	2795
Two or More Races	2437	2506	2553	2594	2629	2663	2698	2740	2795
Unidentified Race	2480	2551	2595	2632	2664	2694	2724	2759	2795
LEP Status	2311	2366	2401	2431	2459	2487	2518	2552	2599
IDEA Indicator	2308	2365	2403	2436	2466	2497	2530	2569	2628
Section 504 Status	2425	2487	2531	2568	2603	2635	2671	2714	2771
Economic Disadvantage Status	2391	2455	2500	2539	2573	2606	2641	2681	2735

TABLE 5-99 GRADE 3 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2271	2327	2365	2397	2426	2454	2483	2517	2564
Female	2281	2337	2374	2406	2434	2462	2490	2523	2570
Male	2259	2317	2358	2389	2418	2447	2476	2510	2558
American Indian or Alaska Native	2237	2288	2319	2346	2372	2399	2427	2459	2503
Asian	2323	2382	2421	2452	2479	2505	2532	2565	2611
Black/African American	2210	2285	2321	2354	2381	2408	2437	2470	2517
Native Hawaiian or Pacific Islander	2255	2305	2339	2368	2395	2420	2451	2482	2526
Hispanic/Latino Ethnicity	2228	2300	2340	2369	2398	2425	2455	2488	2535
White	2307	2361	2397	2426	2453	2478	2504	2535	2579
Two or More Races	2282	2332	2369	2401	2429	2456	2485	2519	2566
Unidentified Race	2293	2356	2394	2426	2454	2480	2510	2543	2590
LEP Status	2207	2277	2311	2343	2368	2395	2422	2454	2500
IDEA Indicator	2192	2241	2284	2312	2342	2370	2402	2440	2497
Section 504 Status	2287	2335	2369	2398	2426	2451	2479	2511	2556
Economic Disadvantage Status	2234	2299	2338	2368	2395	2422	2451	2483	2529

TABLE 5-100 GRADE 4 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2295	2353	2394	2428	2460	2490	2522	2557	2607
Female	2304	2363	2403	2436	2468	2498	2528	2563	2612
Male	2287	2345	2385	2420	2452	2483	2515	2551	2602
American Indian or Alaska Native	2279	2325	2355	2381	2406	2433	2462	2495	2541
Asian	2346	2410	2453	2488	2517	2546	2576	2610	2661
Black/African American	2254	2308	2349	2380	2411	2440	2471	2509	2558
Native Hawaiian or Pacific Islander	2283	2333	2366	2397	2425	2454	2485	2520	2565
Hispanic/Latino Ethnicity	2263	2321	2361	2394	2425	2456	2488	2524	2573
White	2339	2394	2431	2463	2491	2517	2545	2578	2625
Two or More Races	2314	2365	2402	2435	2465	2495	2525	2560	2610
Unidentified Race	2314	2380	2424	2461	2492	2522	2553	2588	2639
LEP Status	2232	2290	2326	2356	2383	2410	2440	2473	2521
IDEA Indicator	2206	2268	2303	2333	2361	2390	2424	2465	2526
Section 504 Status	2312	2362	2398	2428	2457	2485	2516	2550	2598
Economic Disadvantage Status	2269	2325	2363	2395	2425	2455	2486	2521	2569

TABLE 5-101 GRADE 5 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2331	2393	2434	2470	2500	2531	2563	2599	2650
Female	2346	2407	2447	2481	2511	2541	2571	2606	2656
Male	2321	2381	2424	2459	2490	2521	2554	2591	2643
American Indian or Alaska Native	2299	2353	2387	2416	2443	2474	2503	2539	2587
Asian	2387	2453	2495	2529	2560	2589	2620	2654	2701
Black/African American	2293	2350	2391	2424	2455	2485	2516	2551	2601
Native Hawaiian or Pacific Islander	2324	2381	2419	2452	2479	2505	2534	2565	2611
Hispanic/Latino Ethnicity	2303	2360	2400	2432	2464	2494	2526	2563	2615
White	2380	2435	2474	2504	2533	2560	2587	2620	2667
Two or More Races	2357	2412	2451	2483	2512	2540	2570	2604	2654
Unidentified Race	2360	2423	2467	2502	2533	2563	2593	2630	2679
LEP Status	2254	2316	2354	2382	2411	2436	2465	2498	2545
IDEA Indicator	2227	2296	2332	2366	2394	2423	2455	2494	2552
Section 504 Status	2350	2403	2440	2472	2500	2528	2559	2593	2642
Economic Disadvantage Status	2306	2364	2403	2435	2466	2494	2526	2561	2610

TABLE 5-102 GRADE 6 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2375	2432	2471	2503	2532	2560	2592	2629	2689
Female	2389	2445	2484	2515	2543	2571	2602	2638	2698
Male	2363	2420	2459	2492	2521	2550	2581	2620	2679
American Indian or Alaska Native	2336	2382	2418	2445	2473	2501	2529	2564	2614
Asian	2436	2496	2533	2564	2593	2621	2655	2698	2724
Black/African American	2322	2386	2423	2453	2482	2509	2540	2575	2627
Native Hawaiian or Pacific Islander	2354	2407	2443	2472	2500	2526	2551	2581	2625
Hispanic/Latino Ethnicity	2349	2405	2444	2475	2504	2532	2562	2600	2657
White	2412	2464	2500	2529	2555	2582	2611	2647	2703
Two or More Races	2388	2443	2480	2510	2537	2564	2594	2631	2689
Unidentified Race	2416	2472	2509	2540	2569	2599	2632	2675	2724
LEP Status	2292	2348	2382	2412	2437	2463	2490	2520	2566
IDEA Indicator	2278	2325	2367	2394	2423	2449	2479	2515	2569
Section 504 Status	2394	2444	2478	2505	2530	2555	2583	2619	2672
Economic Disadvantage Status	2347	2402	2440	2471	2499	2527	2556	2592	2644

TABLE 5-103 GRADE 7 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2376	2440	2483	2519	2550	2580	2611	2647	2700
Female	2391	2454	2496	2531	2561	2591	2621	2656	2709
Male	2362	2427	2471	2507	2540	2570	2602	2638	2691
American Indian or Alaska Native	2334	2390	2431	2467	2496	2526	2555	2590	2637
Asian	2445	2508	2549	2580	2608	2635	2663	2699	2745
Black/African American	2321	2386	2426	2460	2494	2525	2557	2597	2650
Native Hawaiian or Pacific Islander	2350	2412	2457	2492	2522	2548	2575	2606	2651
Hispanic/Latino Ethnicity	2346	2406	2447	2481	2513	2543	2575	2611	2663
White	2420	2482	2521	2553	2580	2607	2634	2668	2719
Two or More Races	2398	2459	2501	2533	2562	2590	2618	2652	2701
Unidentified Race	2414	2476	2517	2550	2580	2609	2640	2676	2732
LEP Status	2282	2340	2376	2406	2435	2462	2492	2527	2575
IDEA Indicator	2275	2331	2370	2401	2430	2460	2492	2530	2585
Section 504 Status	2399	2460	2499	2529	2555	2581	2609	2643	2691
Economic Disadvantage Status	2345	2405	2446	2481	2512	2543	2573	2609	2659

Table 5-104 Grade 8 ELA/literacy CLAIM 3 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2396	2459	2502	2537	2568	2599	2631	2669	2725
Female	2417	2478	2518	2551	2581	2611	2642	2679	2734
Male	2377	2442	2487	2523	2555	2586	2619	2658	2715
American Indian or Alaska Native	2351	2405	2443	2477	2507	2536	2567	2604	2659
Asian	2464	2525	2565	2598	2628	2657	2690	2729	2769
Black/African American	2340	2406	2447	2484	2515	2545	2578	2616	2672
Native Hawaiian or Pacific Islander	2373	2436	2475	2505	2533	2560	2588	2621	2668
Hispanic/Latino Ethnicity	2366	2433	2475	2508	2539	2568	2600	2637	2692
White	2430	2491	2532	2565	2594	2622	2652	2689	2742
Two or More Races	2408	2469	2510	2544	2572	2602	2632	2669	2722
Unidentified Race	2434	2496	2537	2569	2601	2630	2662	2703	2761
LEP Status	2289	2356	2399	2431	2459	2487	2516	2550	2599
IDEA Indicator	2288	2352	2393	2425	2455	2484	2515	2553	2609
Section 504 Status	2413	2472	2510	2541	2568	2598	2626	2663	2715
Economic Disadvantage Status	2363	2429	2470	2504	2534	2564	2595	2631	2684

Table 5-105 Grade 11 ELA/literacy CLAIM 3 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2409	2480	2529	2569	2605	2641	2677	2720	2781
Female	2428	2496	2543	2582	2616	2650	2686	2727	2786
Male	2395	2465	2515	2557	2594	2631	2669	2713	2776
American Indian or Alaska Native	2379	2440	2488	2529	2560	2593	2626	2666	2724
Asian	2468	2543	2593	2632	2668	2701	2738	2780	2795
Black/African American	2359	2427	2475	2516	2553	2590	2629	2671	2733
Native Hawaiian or Pacific Islander	2381	2441	2487	2526	2559	2591	2623	2658	2712
Hispanic/Latino Ethnicity	2388	2454	2499	2537	2573	2607	2644	2687	2747
White	2452	2523	2571	2608	2641	2673	2706	2746	2795
Two or More Races	2438	2509	2556	2592	2623	2653	2685	2723	2783
Unidentified Race	2449	2514	2557	2596	2630	2663	2699	2740	2795
LEP Status	2302	2364	2404	2435	2465	2494	2524	2561	2613
IDEA Indicator	2310	2375	2414	2448	2479	2511	2545	2587	2650
Section 504 Status	2421	2489	2537	2577	2613	2649	2685	2732	2793
Economic Disadvantage Status	2385	2451	2497	2535	2571	2605	2642	2683	2742

Table 5-106 Grade 3 ELA/literacy Claim 4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2249	2309	2351	2388	2422	2453	2484	2518	2562
Female	2263	2321	2363	2398	2431	2461	2491	2524	2567
Male	2239	2300	2341	2378	2412	2445	2477	2511	2556
American Indian or Alaska Native	2205	2258	2296	2328	2359	2390	2421	2457	2504
Asian	2310	2377	2423	2458	2487	2514	2540	2569	2607
Black/African American	2215	2266	2303	2331	2361	2392	2427	2465	2513
Native Hawaiian or Pacific Islander	2221	2279	2317	2352	2384	2417	2452	2489	2535
Hispanic/Latino Ethnicity	2228	2288	2327	2360	2392	2424	2457	2492	2537
White	2279	2341	2384	2419	2448	2475	2502	2533	2572
Two or More Races	2252	2314	2357	2394	2428	2459	2490	2524	2567
Unidentified Race	2283	2345	2391	2427	2458	2486	2514	2546	2587
LEP Status	2214	2268	2304	2334	2361	2390	2421	2457	2504
IDEA Indicator	2186	2235	2277	2307	2336	2367	2403	2447	2503
Section 504 Status	2245	2305	2348	2387	2422	2452	2482	2514	2559
Economic Disadvantage Status	2226	2284	2323	2355	2387	2418	2451	2486	2531

Table 5-107 Grade 4 ELA/literacy Claim 4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2291	2348	2390	2427	2463	2497	2529	2563	2607
Female	2300	2358	2402	2440	2475	2507	2538	2571	2615
Male	2283	2339	2379	2415	2452	2486	2520	2555	2600
American Indian or Alaska Native	2249	2303	2339	2368	2398	2430	2463	2503	2549
Asian	2354	2425	2472	2506	2534	2560	2586	2615	2657
Black/African American	2252	2305	2341	2367	2397	2427	2464	2504	2554
Native Hawaiian or Pacific Islander	2263	2318	2358	2393	2427	2461	2497	2535	2579
Hispanic/Latino Ethnicity	2270	2325	2364	2399	2432	2465	2499	2535	2581
White	2318	2377	2422	2459	2491	2520	2548	2579	2620
Two or More Races	2297	2355	2398	2436	2472	2504	2536	2569	2614
Unidentified Race	2324	2388	2436	2475	2506	2537	2565	2597	2640
LEP Status	2241	2296	2330	2359	2386	2413	2445	2481	2529
IDEA Indicator	2220	2274	2309	2338	2365	2396	2431	2476	2535
Section 504 Status	2294	2348	2393	2429	2462	2493	2523	2558	2601
Economic Disadvantage Status	2266	2321	2359	2393	2425	2458	2492	2529	2575

Table 5-108 Grade 5 ELA/literacy Claim 4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2318	2381	2426	2464	2499	2531	2563	2597	2643
Female	2336	2398	2443	2480	2513	2543	2574	2607	2652
Male	2302	2365	2410	2449	2484	2518	2551	2586	2632
American Indian or Alaska Native	2263	2323	2363	2396	2428	2461	2495	2534	2584
Asian	2388	2459	2505	2540	2570	2596	2623	2654	2699
Black/African American	2252	2334	2373	2407	2438	2471	2505	2543	2592
Native Hawaiian or Pacific Islander	2291	2351	2394	2430	2465	2498	2531	2569	2614
Hispanic/Latino Ethnicity	2294	2353	2394	2429	2462	2495	2529	2565	2611
White	2356	2420	2464	2499	2529	2557	2583	2614	2657
Two or More Races	2329	2392	2437	2475	2510	2541	2571	2604	2650
Unidentified Race	2357	2428	2476	2513	2544	2573	2601	2630	2673
LEP Status	2254	2312	2347	2378	2405	2432	2461	2495	2543
IDEA Indicator	2235	2284	2322	2352	2381	2412	2447	2491	2552
Section 504 Status	2320	2378	2423	2460	2493	2525	2556	2591	2633
Economic Disadvantage Status	2290	2351	2392	2427	2460	2492	2525	2561	2607

Table 5-109 Grade 6 ELA/literacy Claim 4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2344	2408	2456	2497	2533	2565	2595	2626	2666
Female	2365	2431	2479	2518	2551	2580	2607	2637	2676
Male	2325	2389	2436	2477	2515	2549	2581	2613	2655
American Indian or Alaska Native	2291	2349	2389	2426	2460	2496	2531	2568	2615
Asian	2426	2503	2549	2582	2607	2631	2654	2681	2719
Black/African American	2292	2358	2397	2430	2463	2498	2535	2573	2617
Native Hawaiian or Pacific Islander	2310	2373	2421	2460	2498	2532	2565	2598	2640
Hispanic/Latino Ethnicity	2319	2380	2425	2465	2502	2537	2569	2602	2642
White	2376	2442	2488	2526	2556	2583	2609	2637	2674
Two or More Races	2349	2416	2465	2506	2540	2571	2600	2631	2671
Unidentified Race	2397	2470	2519	2554	2582	2606	2630	2656	2693
LEP Status	2273	2330	2365	2396	2424	2454	2486	2524	2572
IDEA Indicator	2260	2306	2346	2375	2405	2436	2472	2515	2571
Section 504 Status	2341	2406	2453	2493	2526	2558	2586	2616	2655
Economic Disadvantage Status	2314	2376	2420	2458	2493	2528	2561	2595	2636

Table 5-110 Grade 7 ELA/literacy CLAIM 4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2358	2428	2479	2521	2557	2590	2620	2654	2698
Female	2388	2456	2505	2544	2577	2606	2634	2666	2710
Male	2336	2405	2454	2498	2536	2572	2605	2639	2684
American Indian or Alaska Native	2312	2370	2412	2449	2485	2520	2557	2595	2644
Asian	2453	2532	2577	2610	2636	2660	2685	2715	2745
Black/African American	2307	2369	2411	2446	2483	2520	2554	2596	2644
Native Hawaiian or Pacific Islander	2324	2388	2438	2478	2517	2552	2583	2618	2662
Hispanic/Latino Ethnicity	2333	2399	2446	2488	2525	2560	2592	2626	2670
White	2394	2463	2511	2549	2580	2608	2635	2665	2707
Two or More Races	2372	2441	2492	2533	2568	2599	2628	2660	2704
Unidentified Race	2418	2492	2540	2578	2607	2632	2659	2689	2732
LEP Status	2270	2335	2377	2407	2436	2467	2499	2537	2587
IDEA Indicator	2258	2325	2360	2394	2423	2454	2490	2533	2589
Section 504 Status	2370	2435	2485	2523	2556	2584	2613	2646	2690
Economic Disadvantage Status	2330	2394	2439	2479	2517	2551	2585	2620	2664

Table 5-111 Grade 8 ELA/literacy CLAIM 4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2378	2447	2496	2536	2572	2604	2635	2668	2713
Female	2412	2477	2524	2562	2595	2623	2651	2682	2725
Male	2354	2424	2469	2510	2548	2583	2616	2651	2697
American Indian or Alaska Native	2315	2392	2434	2467	2503	2536	2571	2611	2658
Asian	2468	2544	2591	2623	2650	2675	2700	2730	2769
Black/African American	2306	2399	2441	2477	2511	2545	2579	2617	2664
Native Hawaiian or Pacific Islander	2350	2411	2453	2490	2525	2559	2594	2630	2676
Hispanic/Latino Ethnicity	2353	2419	2463	2501	2537	2571	2604	2640	2685
White	2414	2480	2528	2565	2596	2623	2650	2679	2721
Two or More Races	2388	2456	2505	2545	2579	2611	2640	2673	2717
Unidentified Race	2435	2506	2557	2594	2623	2648	2673	2703	2743
LEP Status	2288	2351	2392	2423	2451	2479	2509	2545	2595
IDEA Indicator	2288	2346	2384	2416	2445	2472	2504	2543	2598
Section 504 Status	2377	2443	2488	2525	2561	2592	2624	2657	2703
Economic Disadvantage Status	2347	2417	2459	2498	2532	2566	2600	2636	2681

TABLE 5-112 GRADE 11 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2397	2476	2536	2584	2623	2657	2690	2725	2772
Female	2428	2508	2566	2608	2642	2673	2702	2735	2781
Male	2373	2450	2508	2558	2602	2640	2675	2713	2762
American Indian or Alaska Native	2349	2420	2469	2511	2552	2592	2629	2672	2722
Asian	2488	2579	2630	2665	2694	2720	2747	2778	2795
Black/African American	2336	2410	2458	2504	2549	2591	2631	2672	2724
Native Hawaiian or Pacific Islander	2348	2422	2474	2514	2556	2595	2635	2673	2724
Hispanic/Latino Ethnicity	2375	2449	2504	2552	2593	2630	2664	2700	2748
White	2432	2515	2572	2615	2648	2677	2707	2739	2784
Two or More Races	2421	2502	2558	2601	2636	2667	2696	2729	2775
Unidentified Race	2461	2546	2600	2637	2667	2693	2721	2750	2794
LEP Status	2301	2357	2400	2433	2464	2496	2531	2574	2634
IDEA Indicator	2300	2357	2400	2436	2469	2503	2541	2588	2651
Section 504 Status	2393	2474	2531	2577	2615	2648	2680	2718	2765
Economic Disadvantage Status	2369	2443	2498	2546	2588	2626	2662	2699	2747

TABLE 5-113 GRADE 3 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2320	2361	2388	2410	2431	2453	2476	2503	2542
Female	2323	2361	2386	2408	2428	2449	2471	2497	2534
Male	2318	2361	2389	2413	2434	2456	2480	2509	2549
American Indian or Alaska Native	2281	2324	2350	2371	2390	2409	2429	2454	2493
Asian	2387	2426	2453	2477	2499	2521	2545	2573	2612
Black/African American	2272	2317	2344	2368	2388	2407	2428	2454	2489
Native Hawaiian or Pacific Islander	2302	2347	2374	2396	2416	2437	2458	2481	2517
Hispanic/Latino Ethnicity	2308	2347	2372	2393	2412	2432	2452	2477	2512
White	2345	2383	2409	2431	2451	2471	2492	2518	2554
Two or More Races	2319	2359	2387	2410	2432	2455	2479	2508	2548
Unidentified Race	2353	2395	2421	2443	2464	2486	2507	2534	2572
LEP Status	2293	2333	2358	2378	2396	2415	2435	2459	2495
IDEA Indicator	2223	2276	2312	2339	2362	2387	2413	2445	2492
Section 504 Status	2319	2359	2385	2407	2427	2448	2472	2500	2538
Economic Disadvantage Status	2301	2342	2368	2389	2408	2428	2449	2474	2509

TABLE 5-114 GRADE 4 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2357	2395	2423	2447	2471	2494	2518	2546	2584
Female	2361	2395	2421	2445	2467	2489	2512	2539	2576
Male	2352	2394	2424	2451	2475	2499	2524	2552	2591
American Indian or Alaska Native	2322	2359	2385	2405	2425	2445	2469	2495	2531
Asian	2421	2465	2495	2520	2542	2564	2586	2613	2652
Black/African American	2312	2352	2377	2398	2418	2439	2465	2493	2530
Native Hawaiian or Pacific Islander	2344	2381	2407	2429	2450	2472	2496	2520	2557
Hispanic/Latino Ethnicity	2341	2376	2400	2422	2443	2464	2487	2512	2548
White	2386	2424	2452	2474	2495	2515	2537	2562	2596
Two or More Races	2360	2399	2428	2453	2477	2499	2524	2552	2591
Unidentified Race	2386	2427	2458	2484	2506	2527	2551	2577	2611
LEP Status	2320	2356	2378	2397	2415	2434	2456	2481	2517
IDEA Indicator	2266	2316	2346	2370	2391	2414	2442	2477	2525
Section 504 Status	2365	2401	2428	2450	2473	2496	2519	2547	2588
Economic Disadvantage Status	2337	2374	2399	2421	2442	2463	2486	2512	2548

TABLE 5-115 GRADE 5 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2368	2409	2440	2468	2494	2520	2549	2581	2622
Female	2373	2411	2440	2466	2491	2516	2543	2574	2616
Male	2363	2406	2439	2469	2498	2526	2554	2587	2628
American Indian or Alaska Native	2335	2373	2401	2423	2444	2468	2494	2525	2567
Asian	2437	2486	2521	2551	2577	2602	2627	2656	2699
Black/African American	2328	2363	2390	2412	2435	2457	2483	2513	2556
Native Hawaiian or Pacific Islander	2361	2400	2426	2452	2475	2500	2526	2555	2597
Hispanic/Latino Ethnicity	2352	2389	2416	2440	2463	2486	2511	2541	2582
White	2397	2441	2472	2499	2523	2547	2572	2599	2634
Two or More Races	2375	2416	2447	2475	2502	2529	2557	2589	2629
Unidentified Race	2401	2447	2482	2510	2535	2560	2587	2614	2650
LEP Status	2328	2364	2388	2408	2427	2448	2470	2497	2536
IDEA Indicator	2293	2332	2360	2383	2404	2427	2454	2490	2543
Section 504 Status	2377	2415	2445	2472	2498	2524	2551	2582	2621
Economic Disadvantage Status	2350	2387	2414	2438	2461	2485	2510	2540	2582

TABLE 5-116 GRADE 6 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2360	2420	2463	2496	2525	2552	2579	2609	2652
Female	2371	2430	2469	2501	2529	2554	2580	2609	2650
Male	2348	2411	2456	2490	2520	2549	2577	2609	2655
American Indian or Alaska Native	2313	2366	2404	2439	2469	2497	2525	2556	2594
Asian	2461	2518	2556	2585	2610	2635	2664	2698	2748
Black/African American	2305	2356	2393	2426	2455	2483	2511	2545	2586
Native Hawaiian or Pacific Islander	2338	2393	2436	2470	2499	2527	2554	2581	2618
Hispanic/Latino Ethnicity	2334	2388	2428	2461	2489	2516	2543	2573	2612
White	2407	2465	2501	2528	2552	2575	2598	2625	2664
Two or More Races	2366	2427	2470	2503	2532	2559	2585	2615	2659
Unidentified Race	2415	2476	2515	2545	2570	2594	2619	2648	2691
LEP Status	2286	2337	2369	2398	2425	2451	2478	2509	2552
IDEA Indicator	2244	2301	2336	2365	2395	2427	2461	2501	2555
Section 504 Status	2377	2437	2473	2502	2528	2553	2577	2608	2650
Economic Disadvantage Status	2331	2385	2425	2458	2487	2514	2541	2571	2610

TABLE 5-117 GRADE 7 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2374	2432	2472	2504	2535	2565	2597	2633	2680
Female	2385	2439	2476	2507	2536	2564	2596	2631	2677
Male	2364	2424	2467	2502	2535	2566	2599	2635	2683
American Indian or Alaska Native	2332	2384	2420	2452	2478	2506	2535	2568	2612
Asian	2468	2532	2574	2609	2638	2665	2694	2730	2778
Black/African American	2323	2371	2407	2436	2463	2490	2520	2554	2603
Native Hawaiian or Pacific Islander	2349	2404	2441	2472	2501	2528	2556	2589	2633
Hispanic/Latino Ethnicity	2347	2400	2438	2468	2496	2523	2551	2585	2631
White	2420	2474	2511	2541	2567	2594	2622	2652	2691
Two or More Races	2382	2441	2480	2514	2544	2574	2606	2642	2688
Unidentified Race	2423	2484	2524	2556	2585	2615	2642	2675	2719
LEP Status	2295	2342	2375	2402	2427	2452	2478	2510	2557
IDEA Indicator	2271	2319	2351	2379	2406	2434	2466	2504	2560
Section 504 Status	2399	2452	2488	2517	2545	2571	2599	2633	2681
Economic Disadvantage Status	2345	2398	2436	2466	2494	2522	2550	2585	2631

TABLE 5-118 GRADE 8 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2379	2434	2473	2508	2541	2577	2614	2657	2715
Female	2391	2445	2483	2516	2549	2583	2618	2659	2712
Male	2369	2423	2464	2499	2534	2571	2610	2656	2717
American Indian or Alaska Native	2340	2387	2423	2453	2481	2511	2542	2581	2636
Asian	2470	2539	2589	2630	2665	2699	2733	2774	2802
Black/African American	2331	2378	2412	2438	2467	2494	2528	2569	2626
Native Hawaiian or Pacific Islander	2354	2408	2446	2478	2507	2535	2568	2608	2659
Hispanic/Latino Ethnicity	2354	2405	2442	2473	2502	2532	2565	2605	2659
White	2417	2472	2510	2543	2575	2607	2640	2678	2726
Two or More Races	2386	2440	2480	2515	2549	2581	2618	2661	2719
Unidentified Race	2424	2487	2530	2569	2604	2639	2673	2711	2763
LEP Status	2301	2347	2378	2406	2430	2456	2485	2520	2578
IDEA Indicator	2284	2329	2360	2387	2413	2440	2470	2509	2570
Section 504 Status	2397	2448	2484	2514	2543	2574	2610	2651	2708
Economic Disadvantage Status	2353	2403	2439	2470	2499	2529	2563	2603	2659

TABLE 5-119 GRADE 11 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2398	2455	2497	2534	2568	2602	2639	2681	2741
Female	2413	2466	2506	2542	2575	2607	2642	2681	2736
Male	2385	2444	2487	2526	2561	2596	2635	2682	2747
American Indian or Alaska Native	2361	2414	2450	2482	2510	2541	2574	2612	2669
Asian	2497	2569	2617	2656	2691	2723	2760	2800	2849
Black/African American	2351	2401	2439	2471	2503	2536	2569	2609	2661
Native Hawaiian or Pacific Islander	2378	2431	2469	2502	2532	2561	2591	2628	2674
Hispanic/Latino Ethnicity	2377	2429	2466	2498	2530	2561	2593	2631	2683
White	2437	2496	2537	2571	2602	2633	2666	2704	2757
Two or More Races	2410	2468	2508	2542	2574	2607	2643	2685	2742
Unidentified Race	2439	2501	2546	2583	2616	2647	2678	2712	2759
LEP Status	2313	2360	2392	2417	2441	2467	2496	2533	2594
IDEA Indicator	2301	2348	2381	2406	2431	2456	2484	2521	2578
Section 504 Status	2413	2465	2503	2536	2566	2595	2630	2670	2727
Economic Disadvantage Status	2375	2428	2465	2498	2530	2561	2595	2634	2689

Table 5-120 Grade 3 MATHEMATICS CLAIM 2/4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2291	2343	2376	2402	2426	2449	2472	2499	2536
Female	2293	2342	2374	2399	2423	2446	2469	2496	2532
Male	2290	2343	2377	2405	2429	2451	2475	2502	2539
American Indian or Alaska Native	2250	2302	2335	2360	2382	2405	2428	2455	2490
Asian	2355	2406	2438	2463	2485	2506	2527	2553	2591
Black/African American	2228	2300	2333	2359	2383	2406	2429	2456	2492
Native Hawaiian or Pacific Islander	2264	2315	2348	2374	2398	2421	2445	2471	2507
Hispanic/Latino Ethnicity	2271	2323	2354	2379	2402	2423	2446	2471	2505
White	2330	2378	2409	2432	2453	2473	2494	2518	2552
Two or More Races	2295	2345	2378	2406	2430	2453	2478	2505	2541
Unidentified Race	2319	2367	2400	2425	2448	2470	2492	2515	2548
LEP Status	2231	2302	2333	2357	2379	2400	2422	2447	2482
IDEA Indicator	2189	2268	2305	2332	2357	2383	2412	2445	2490
Section 504 Status	2295	2346	2378	2401	2424	2446	2470	2496	2533
Economic Disadvantage Status	2267	2320	2351	2377	2400	2421	2444	2470	2504

Table 5-121 Grade 4 MATHEMATICS CLAIM 2/4 Scale Score by Selected Demographic Groups, 2016-2017 Percentiles

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2334	2381	2414	2442	2466	2490	2515	2544	2584
Female	2336	2382	2414	2440	2464	2487	2511	2539	2578
Male	2332	2381	2415	2443	2469	2493	2519	2548	2589
American Indian or Alaska Native	2302	2344	2374	2398	2421	2443	2467	2495	2531
Asian	2396	2448	2481	2507	2530	2552	2576	2603	2643
Black/African American	2299	2342	2371	2396	2420	2443	2467	2494	2531
Native Hawaiian or Pacific Islander	2312	2355	2387	2414	2438	2461	2486	2512	2546
Hispanic/Latino Ethnicity	2314	2358	2388	2413	2436	2459	2482	2508	2544
White	2373	2419	2450	2474	2496	2518	2540	2566	2603
Two or More Races	2341	2389	2423	2450	2474	2498	2523	2552	2592
Unidentified Race	2357	2407	2439	2463	2487	2509	2531	2558	2593
LEP Status	2285	2331	2359	2383	2404	2425	2447	2473	2508
IDEA Indicator	2245	2311	2338	2363	2386	2412	2440	2475	2523
Section 504 Status	2345	2390	2420	2447	2470	2493	2518	2548	2586
Economic Disadvantage Status	2314	2357	2388	2413	2437	2459	2482	2509	2545

TABLE 5-122 GRADE 5 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2219	2389	2429	2461	2489	2517	2546	2577	2616
Female	2287	2392	2430	2461	2489	2516	2543	2574	2613
Male	2219	2386	2427	2460	2490	2519	2548	2579	2620
American Indian or Alaska Native	2219	2344	2382	2412	2438	2464	2492	2524	2563
Asian	2407	2469	2510	2540	2565	2588	2612	2637	2673
Black/African American	2219	2325	2380	2410	2436	2462	2488	2520	2561
Native Hawaiian or Pacific Islander	2316	2370	2406	2435	2463	2489	2515	2547	2585
Hispanic/Latino Ethnicity	2219	2355	2396	2426	2452	2478	2504	2534	2574
White	2379	2438	2475	2503	2528	2551	2574	2600	2634
Two or More Races	2343	2401	2440	2472	2501	2528	2556	2586	2624
Unidentified Race	2360	2424	2463	2494	2520	2545	2569	2594	2627
LEP Status	2219	2219	2352	2384	2408	2431	2455	2482	2520
IDEA Indicator	2219	2219	2331	2371	2398	2424	2452	2488	2541
Section 504 Status	2331	2402	2440	2468	2496	2521	2548	2578	2617
Economic Disadvantage Status	2219	2358	2397	2427	2453	2479	2506	2536	2576

TABLE 5-123 GRADE 6 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2332	2406	2451	2486	2515	2543	2571	2603	2648
Female	2342	2412	2455	2488	2516	2542	2569	2600	2645
Male	2318	2399	2448	2484	2514	2543	2572	2605	2651
American Indian or Alaska Native	2235	2339	2384	2423	2455	2483	2512	2543	2586
Asian	2438	2501	2540	2569	2596	2622	2649	2679	2724
Black/African American	2235	2330	2377	2420	2453	2483	2512	2544	2586
Native Hawaiian or Pacific Islander	2291	2370	2419	2454	2483	2511	2539	2568	2607
Hispanic/Latino Ethnicity	2263	2367	2417	2451	2480	2506	2532	2561	2600
White	2388	2456	2495	2523	2549	2573	2598	2627	2667
Two or More Races	2345	2414	2459	2493	2522	2550	2579	2612	2656
Unidentified Race	2383	2451	2492	2522	2547	2571	2595	2624	2662
LEP Status	2235	2277	2347	2381	2415	2442	2470	2500	2540
IDEA Indicator	2235	2235	2301	2349	2383	2420	2456	2495	2548
Section 504 Status	2358	2420	2462	2492	2518	2543	2570	2600	2643
Economic Disadvantage Status	2268	2365	2415	2450	2479	2506	2532	2562	2602

TABLE 5-124 GRADE 7 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2309	2400	2450	2490	2525	2558	2593	2631	2681
Female	2320	2403	2452	2491	2525	2558	2592	2630	2679
Male	2295	2396	2448	2489	2525	2559	2594	2633	2683
American Indian or Alaska Native	2250	2336	2388	2427	2461	2494	2526	2561	2609
Asian	2435	2510	2557	2594	2626	2654	2682	2714	2758
Black/African American	2250	2329	2382	2418	2453	2486	2519	2555	2606
Native Hawaiian or Pacific Islander	2282	2355	2407	2447	2481	2513	2547	2580	2629
Hispanic/Latino Ethnicity	2250	2363	2407	2445	2478	2509	2541	2577	2625
White	2387	2459	2504	2539	2568	2597	2625	2657	2699
Two or More Races	2337	2416	2466	2505	2538	2570	2604	2642	2691
Unidentified Race	2366	2445	2493	2529	2561	2590	2621	2654	2699
LEP Status	2250	2250	2337	2377	2405	2436	2465	2500	2547
IDEA Indicator	2250	2250	2325	2367	2396	2428	2461	2500	2557
Section 504 Status	2347	2421	2467	2503	2534	2565	2597	2633	2680
Economic Disadvantage Status	2250	2363	2408	2447	2480	2511	2544	2580	2628

TABLE 5-125 GRADE 8 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2345	2420	2459	2498	2536	2576	2616	2658	2712
Female	2348	2425	2465	2505	2543	2580	2618	2658	2710
Male	2336	2416	2455	2491	2529	2571	2613	2658	2713
American Indian or Alaska Native	2294	2360	2417	2446	2472	2502	2535	2579	2635
Asian	2442	2519	2576	2618	2654	2684	2716	2753	2802
Black/African American	2265	2355	2407	2438	2465	2497	2532	2577	2634
Native Hawaiian or Pacific Islander	2308	2369	2424	2459	2493	2524	2558	2599	2653
Hispanic/Latino Ethnicity	2265	2371	2426	2457	2487	2520	2557	2599	2653
White	2404	2464	2510	2549	2584	2617	2649	2684	2730
Two or More Races	2352	2428	2469	2509	2547	2585	2624	2665	2717
Unidentified Race	2380	2454	2502	2543	2581	2617	2650	2685	2731
LEP Status	2265	2265	2347	2373	2414	2440	2463	2497	2549
IDEA Indicator	2265	2301	2347	2371	2414	2440	2465	2501	2562
Section 504 Status	2357	2431	2468	2503	2536	2572	2610	2650	2703
Economic Disadvantage Status	2280	2372	2427	2458	2489	2522	2559	2602	2656

TABLE 5-126 GRADE 11 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2280	2405	2464	2509	2551	2591	2632	2677	2735
Female	2280	2414	2467	2510	2550	2588	2626	2670	2725
Male	2280	2395	2461	2508	2552	2594	2637	2684	2745
American Indian or Alaska Native	2280	2325	2413	2460	2496	2535	2574	2619	2677
Asian	2433	2524	2584	2628	2667	2700	2735	2773	2823
Black/African American	2280	2280	2379	2434	2468	2504	2544	2590	2649
Native Hawaiian or Pacific Islander	2280	2325	2405	2458	2498	2539	2576	2619	2672
Hispanic/Latino Ethnicity	2280	2342	2431	2468	2505	2541	2579	2622	2677
White	2338	2461	2518	2562	2599	2633	2668	2707	2758
Two or More Races	2302	2423	2482	2529	2569	2608	2646	2688	2742
Unidentified Race	2305	2454	2506	2551	2589	2625	2661	2698	2747
LEP Status	2280	2280	2280	2365	2418	2452	2482	2520	2578
IDEA Indicator	2280	2280	2280	2357	2410	2445	2477	2517	2578
Section 504 Status	2280	2418	2474	2518	2556	2594	2633	2676	2733
Economic Disadvantage Status	2280	2341	2429	2469	2507	2545	2584	2627	2685

TABLE 5-127 GRADE 3 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2298	2345	2377	2404	2428	2453	2477	2503	2540
Female	2304	2350	2380	2407	2430	2453	2476	2502	2538
Male	2293	2341	2374	2401	2427	2452	2477	2504	2542
American Indian or Alaska Native	2258	2308	2337	2361	2383	2404	2430	2456	2494
Asian	2359	2410	2443	2469	2491	2513	2535	2562	2600
Black/African American	2254	2306	2337	2359	2383	2404	2428	2456	2493
Native Hawaiian or Pacific Islander	2282	2327	2357	2383	2406	2429	2455	2481	2516
Hispanic/Latino Ethnicity	2280	2328	2358	2382	2405	2428	2451	2477	2511
White	2326	2373	2404	2430	2453	2474	2495	2519	2552
Two or More Races	2298	2347	2379	2406	2431	2456	2481	2509	2547
Unidentified Race	2329	2377	2411	2438	2461	2483	2504	2529	2565
LEP Status	2263	2312	2341	2365	2386	2407	2431	2457	2493
IDEA Indicator	2234	2284	2315	2339	2362	2386	2412	2446	2492
Section 504 Status	2293	2339	2371	2397	2423	2448	2471	2500	2537
Economic Disadvantage Status	2277	2325	2355	2380	2403	2425	2449	2475	2510

TABLE 5-128 GRADE 4 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2336	2383	2414	2442	2467	2491	2517	2545	2583
Female	2339	2386	2415	2441	2465	2490	2514	2542	2579
Male	2332	2381	2414	2442	2468	2493	2519	2549	2587
American Indian or Alaska Native	2302	2345	2374	2398	2419	2442	2465	2493	2532
Asian	2402	2452	2485	2513	2536	2557	2580	2606	2645
Black/African American	2285	2342	2374	2396	2418	2439	2461	2490	2527
Native Hawaiian or Pacific Islander	2315	2363	2392	2418	2442	2465	2488	2516	2555
Hispanic/Latino Ethnicity	2318	2363	2392	2416	2438	2460	2484	2512	2548
White	2366	2413	2445	2471	2493	2515	2537	2562	2596
Two or More Races	2339	2388	2420	2447	2473	2498	2524	2553	2590
Unidentified Race	2364	2414	2449	2477	2501	2526	2549	2575	2609
LEP Status	2291	2340	2369	2391	2411	2431	2452	2477	2513
IDEA Indicator	2262	2319	2348	2373	2394	2416	2441	2475	2524
Section 504 Status	2338	2386	2417	2443	2468	2493	2518	2547	2586
Economic Disadvantage Status	2316	2361	2391	2415	2437	2460	2483	2511	2548

TABLE 5-129 GRADE 5 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2338	2393	2428	2458	2486	2515	2545	2579	2622
Female	2342	2395	2430	2458	2486	2513	2543	2577	2620
Male	2333	2390	2427	2458	2487	2517	2547	2581	2625
American Indian or Alaska Native	2297	2350	2384	2412	2436	2462	2490	2520	2565
Asian	2409	2468	2508	2540	2567	2592	2618	2645	2686
Black/African American	2294	2351	2384	2411	2432	2455	2480	2513	2557
Native Hawaiian or Pacific Islander	2320	2372	2408	2436	2464	2491	2519	2551	2593
Hispanic/Latino Ethnicity	2313	2368	2401	2428	2453	2478	2505	2536	2580
White	2375	2428	2463	2492	2519	2545	2571	2601	2638
Two or More Races	2345	2399	2435	2466	2495	2524	2554	2589	2632
Unidentified Race	2376	2431	2469	2501	2530	2556	2584	2612	2649
LEP Status	2285	2336	2370	2395	2417	2439	2462	2489	2528
IDEA Indicator	2267	2315	2352	2379	2403	2426	2452	2486	2539
Section 504 Status	2341	2397	2432	2461	2489	2517	2544	2578	2623
Economic Disadvantage Status	2313	2368	2401	2428	2453	2478	2505	2536	2580

TABLE 5-130 GRADE 6 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2345	2406	2445	2479	2510	2542	2574	2609	2655
Female	2356	2414	2452	2484	2514	2545	2577	2611	2655
Male	2336	2399	2440	2474	2506	2538	2572	2608	2655
American Indian or Alaska Native	2306	2363	2396	2428	2454	2483	2513	2549	2597
Asian	2429	2493	2538	2573	2604	2632	2659	2691	2742
Black/African American	2273	2355	2392	2422	2449	2476	2504	2538	2585
Native Hawaiian or Pacific Islander	2331	2387	2423	2453	2480	2510	2539	2575	2619
Hispanic/Latino Ethnicity	2325	2382	2417	2447	2475	2503	2533	2567	2612
White	2381	2442	2481	2513	2542	2570	2598	2628	2669
Two or More Races	2355	2413	2453	2485	2518	2550	2583	2618	2663
Unidentified Race	2393	2452	2493	2529	2559	2588	2616	2647	2687
LEP Status	2282	2337	2374	2399	2423	2446	2470	2499	2542
IDEA Indicator	2267	2321	2355	2384	2407	2432	2459	2493	2548
Section 504 Status	2363	2415	2452	2482	2513	2542	2572	2605	2650
Economic Disadvantage Status	2319	2379	2415	2445	2473	2501	2531	2566	2610

TABLE 5-131 GRADE 7 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2343	2411	2457	2496	2531	2563	2596	2632	2680
Female	2355	2421	2466	2503	2536	2567	2598	2633	2681
Male	2330	2402	2448	2488	2525	2559	2593	2630	2680
American Indian or Alaska Native	2287	2361	2403	2435	2465	2495	2528	2565	2614
Asian	2446	2521	2566	2600	2630	2658	2687	2719	2767
Black/African American	2250	2353	2399	2436	2464	2496	2528	2561	2609
Native Hawaiian or Pacific Islander	2317	2382	2426	2460	2495	2526	2558	2590	2635
Hispanic/Latino Ethnicity	2316	2381	2425	2459	2492	2523	2555	2589	2633
White	2374	2447	2493	2530	2561	2590	2618	2650	2694
Two or More Races	2351	2420	2466	2504	2540	2572	2605	2641	2688
Unidentified Race	2390	2468	2515	2552	2583	2611	2640	2672	2716
LEP Status	2250	2332	2370	2403	2432	2459	2488	2522	2568
IDEA Indicator	2250	2316	2359	2386	2417	2443	2473	2509	2565
Section 504 Status	2363	2426	2469	2506	2538	2569	2598	2632	2680
Economic Disadvantage Status	2301	2377	2423	2457	2489	2521	2553	2587	2633

TABLE 5-132 GRADE 8 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2341	2417	2465	2505	2541	2576	2614	2657	2711
Female	2364	2434	2481	2518	2553	2586	2622	2662	2713
Male	2324	2401	2449	2491	2528	2566	2606	2652	2709
American Indian or Alaska Native	2283	2359	2405	2440	2474	2507	2544	2585	2640
Asian	2452	2529	2583	2626	2660	2693	2723	2759	2802
Black/African American	2267	2350	2404	2438	2473	2504	2536	2573	2628
Native Hawaiian or Pacific Islander	2317	2390	2433	2470	2505	2539	2571	2614	2665
Hispanic/Latino Ethnicity	2323	2389	2433	2470	2503	2535	2569	2608	2660
White	2379	2451	2499	2537	2571	2605	2640	2677	2724
Two or More Races	2345	2421	2471	2511	2548	2584	2622	2663	2716
Unidentified Race	2403	2474	2524	2565	2601	2635	2668	2704	2752
LEP Status	2265	2327	2368	2403	2434	2462	2494	2529	2581
IDEA Indicator	2265	2314	2349	2383	2413	2442	2475	2515	2572
Section 504 Status	2353	2422	2468	2504	2538	2572	2609	2652	2705
Economic Disadvantage Status	2311	2384	2430	2466	2500	2532	2567	2606	2659

TABLE 5-133 GRADE 11 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2016-2017 PERCENTILES

Group	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2350	2438	2488	2528	2565	2603	2643	2687	2745
Female	2363	2448	2496	2535	2571	2607	2644	2684	2737
Male	2338	2428	2479	2520	2559	2599	2641	2690	2752
American Indian or Alaska Native	2300	2391	2441	2482	2517	2550	2584	2625	2677
Asian	2467	2548	2604	2648	2684	2717	2751	2789	2845
Black/African American	2291	2384	2438	2475	2506	2538	2572	2612	2667
Native Hawaiian or Pacific Islander	2317	2403	2457	2494	2528	2559	2596	2635	2684
Hispanic/Latino Ethnicity	2320	2408	2458	2495	2528	2561	2596	2637	2690
White	2389	2473	2524	2565	2602	2637	2674	2713	2765
Two or More Races	2358	2444	2495	2536	2573	2610	2649	2692	2749
Unidentified Race	2402	2485	2534	2574	2609	2643	2678	2714	2760
LEP Status	2280	2340	2395	2436	2468	2497	2527	2561	2615
IDEA Indicator	2280	2335	2389	2429	2461	2488	2517	2551	2601
Section 504 Status	2351	2441	2490	2528	2563	2599	2638	2682	2741
Economic Disadvantage Status	2319	2407	2458	2496	2530	2563	2599	2641	2696

Modes of Administration

Table 5-134 through Table 5-147 present counts of summative assessment administrations by mode[6]. These counts are presented at the aggregate level and disaggregated by gender, by race/ethnicity, and by various status flags: limited English proficiency (LEP status), IDEA indicator, Section 504, and economically disadvantaged.

TABLE 5-134 COUNT OF GRADE 3 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	729739	724900	99.337	4839	0.663
Female	355761	353425	99.343	2336	0.657
Male	373729	371227	99.331	2502	0.669
American Indian or Alaska Native	8121	8065	99.310	56	0.690
Asian	49727	49707	99.960	20	0.040
Black/African American	54393	52957	97.360	1436	2.640
Native Hawaiian or Pacific Islander	6892	6888	99.942	4	0.058
Hispanic/Latino Ethnicity	284731	284415	99.889	316	0.111
White	262929	260247	98.980	2682	1.020
Two or More Races	46857	46605	99.462	252	0.538
Unidentified Race	16089	16016	99.546	73	0.454
LEP Status	163674	163212	99.718	462	0.282
IDEA Indicator	68156	67959	99.711	197	0.289
Section 504 Status	4798	4790	99.833	8	0.167
Economic Disadvantage Status	420340	417466	99.316	2874	0.684

[6] Data for mode counts was provided by the following Consortium members: California, Delaware, Hawaii, Idaho, Michigan, Montana, Nevada, North Dakota, Oregon, South Dakota, Vermont, Virgin Islands and Washington.

TABLE 5-135 COUNT OF GRADE 4 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	822609	817553	99.385	5056	0.615
Female	401931	399560	99.410	2371	0.590
Male	420423	417738	99.361	2685	0.639
American Indian or Alaska Native	9359	9304	99.412	55	0.588
Asian	56589	56567	99.961	22	0.039
Black/African American	58279	56865	97.574	1414	2.426
Native Hawaiian or Pacific Islander	7953	7947	99.925	6	0.075
Hispanic/Latino Ethnicity	308103	307714	99.874	389	0.126
White	313042	310227	99.101	2815	0.899
Two or More Races	52989	52741	99.532	248	0.468
Unidentified Race	16295	16188	99.343	107	0.657
LEP Status	155366	154912	99.708	454	0.292
IDEA Indicator	82849	82594	99.692	255	0.308
Section 504 Status	8776	8760	99.818	16	0.182
Economic Disadvantage Status	465451	462443	99.354	3008	0.646

TABLE 5-136 COUNT OF GRADE 5 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	832537	827780	99.429	4757	0.571
Female	407870	405593	99.442	2277	0.558
Male	424401	421921	99.416	2480	0.584
American Indian or Alaska Native	9435	9369	99.300	66	0.700
Asian	57891	57870	99.964	21	0.036
Black/African American	58692	57342	97.700	1350	2.300
Native Hawaiian or Pacific Islander	8104	8100	99.951	4	0.049
Hispanic/Latino Ethnicity	315442	315129	99.901	313	0.099
White	314654	311988	99.153	2666	0.847
Two or More Races	51136	50908	99.554	228	0.446
Unidentified Race	17183	17074	99.366	109	0.634
LEP Status	138459	138073	99.721	386	0.279
IDEA Indicator	85162	84934	99.732	228	0.268
Section 504 Status	10393	10371	99.788	22	0.212
Economic Disadvantage Status	469363	466643	99.420	2720	0.580

TABLE 5-137 COUNT OF GRADE 6 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	813479	808354	99.370	5125	0.630
Female	398248	395768	99.377	2480	0.623
Male	414977	412332	99.363	2645	0.637
American Indian or Alaska Native	9371	9321	99.466	50	0.534
Asian	58477	58465	99.979	12	0.021
Black/African American	57055	55697	97.620	1358	2.380
Native Hawaiian or Pacific Islander	9113	9109	99.956	4	0.044
Hispanic/Latino Ethnicity	304014	303687	99.892	327	0.108
White	311208	308177	99.026	3031	0.974
Two or More Races	46998	46747	99.466	251	0.534
Unidentified Race	17243	17151	99.466	92	0.534
LEP Status	111268	110914	99.682	354	0.318
IDEA Indicator	81225	81077	99.818	148	0.182
Section 504 Status	11026	11020	99.946	6	0.054
Economic Disadvantage Status	449148	446175	99.338	2973	0.662

TABLE 5-138 COUNT OF GRADE 7 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	808946	803780	99.361	5166	0.639
Female	395314	392750	99.351	2564	0.649
Male	413375	410773	99.371	2602	0.629
American Indian or Alaska Native	9367	9307	99.359	60	0.641
Asian	59509	59498	99.982	11	0.018
Black/African American	56759	55256	97.352	1503	2.648
Native Hawaiian or Pacific Islander	8870	8868	99.977	2	0.023
Hispanic/Latino Ethnicity	295538	295259	99.906	279	0.094
White	315678	312674	99.048	3004	0.952
Two or More Races	44929	44707	99.506	222	0.494
Unidentified Race	18296	18211	99.535	85	0.465
LEP Status	95173	94832	99.642	341	0.358
IDEA Indicator	76894	76766	99.834	128	0.166
Section 504 Status	12178	12165	99.893	13	0.107
Economic Disadvantage Status	434618	431749	99.340	2869	0.660

TABLE 5-139 COUNT OF GRADE 8 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	794913	789810	99.358	5103	0.642
Female	387966	385477	99.358	2489	0.642
Male	406691	404077	99.357	2614	0.643
American Indian or Alaska Native	8972	8918	99.398	54	0.602
Asian	58514	58504	99.983	10	0.017
Black/African American	56284	54844	97.442	1440	2.558
Native Hawaiian or Pacific Islander	8213	8212	99.988	1	0.012
Hispanic/Latino Ethnicity	289630	289365	99.909	265	0.091
White	312818	309781	99.029	3037	0.971
Two or More Races	41722	41521	99.518	201	0.482
Unidentified Race	18760	18665	99.494	95	0.506
LEP Status	80633	80300	99.587	333	0.413
IDEA Indicator	74211	74085	99.830	126	0.170
Section 504 Status	13144	13129	99.886	15	0.114
Economic Disadvantage Status	420800	418054	99.347	2746	0.653

TABLE 5-140 COUNT OF GRADE 11 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	550267	549732	99.903	535	0.097
Female	269894	269649	99.909	245	0.091
Male	280174	279884	99.896	290	0.104
American Indian or Alaska Native	5263	5143	97.720	120	2.280
Asian	52869	52856	99.975	13	0.025
Black/African American	28810	28802	99.972	8	0.028
Native Hawaiian or Pacific Islander	0	5972	100.000	0	0.000
Hispanic/Latino Ethnicity	242719	242670	99.980	49	0.020
White	171613	171285	99.809	328	0.191
Two or More Races	26829	26812	99.937	17	0.063
Unidentified Race	0	16192	100.000	0	0.000
LEP Status	49387	49373	99.972	14	0.028
IDEA Indicator	49159	49081	99.841	78	0.159
Section 504 Status	11339	11323	99.859	16	0.141
Economic Disadvantage Status	291777	291527	99.914	250	0.086

TABLE 5-141 COUNT OF GRADE 3 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	810768	805750	99.381	5018	0.619
Female	395810	393402	99.392	2408	0.608
Male	414710	412101	99.371	2609	0.629
American Indian or Alaska Native	9265	9206	99.363	59	0.637
Asian	55241	55209	99.942	32	0.058
Black/African American	57794	56353	97.507	1441	2.493
Native Hawaiian or Pacific Islander	7761	7757	99.948	4	0.052
Hispanic/Latino Ethnicity	303884	303509	99.877	375	0.123
White	306695	303911	99.092	2784	0.908
Two or More Races	54123	53861	99.516	262	0.484
Unidentified Race	16005	15944	99.619	61	0.381
LEP Status	174557	174136	99.759	421	0.241
IDEA Indicator	77684	77405	99.641	279	0.359
Section 504 Status	6700	6687	99.806	13	0.194
Economic Disadvantage Status	460971	458012	99.358	2959	0.642

TABLE 5-142 COUNT OF GRADE 4 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	819115	814143	99.393	4972	0.607
Female	400282	397950	99.417	2332	0.583
Male	418580	415940	99.369	2640	0.631
American Indian or Alaska Native	9375	9320	99.413	55	0.587
Asian	55684	55663	99.962	21	0.038
Black/African American	58209	56796	97.573	1413	2.427
Native Hawaiian or Pacific Islander	7897	7892	99.937	5	0.063
Hispanic/Latino Ethnicity	306486	306099	99.874	387	0.126
White	312484	309726	99.117	2758	0.883
Two or More Races	52990	52742	99.532	248	0.468
Unidentified Race	15990	15905	99.468	85	0.532
LEP Status	152909	152524	99.748	385	0.252
IDEA Indicator	82673	82420	99.694	253	0.306
Section 504 Status	8677	8661	99.816	16	0.184
Economic Disadvantage Status	463134	460182	99.363	2952	0.637

TABLE 5-143 COUNT OF GRADE 5 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	829666	825024	99.440	4642	0.560
Female	406518	404293	99.453	2225	0.547
Male	422883	420467	99.429	2416	0.571
American Indian or Alaska Native	9455	9389	99.302	66	0.698
Asian	57094	57074	99.965	20	0.035
Black/African American	58639	57292	97.703	1347	2.297
Native Hawaiian or Pacific Islander	8098	8092	99.926	6	0.074
Hispanic/Latino Ethnicity	314001	313690	99.901	311	0.099
White	314343	311756	99.177	2587	0.823
Two or More Races	51120	50895	99.560	225	0.440
Unidentified Race	16916	16836	99.527	80	0.473
LEP Status	136283	135985	99.781	298	0.219
IDEA Indicator	85087	84863	99.737	224	0.263
Section 504 Status	10259	10238	99.795	21	0.205
Economic Disadvantage Status	467253	464604	99.433	2649	0.567

TABLE 5-144 COUNT OF GRADE 6 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	810983	805976	99.383	5007	0.617
Female	397062	394642	99.391	2420	0.609
Male	413667	411080	99.375	2587	0.625
American Indian or Alaska Native	9370	9319	99.456	51	0.544
Asian	57784	57772	99.979	12	0.021
Black/African American	57014	55659	97.623	1355	2.377
Native Hawaiian or Pacific Islander	9088	9084	99.956	4	0.044
Hispanic/Latino Ethnicity	302554	302234	99.894	320	0.106
White	311121	308178	99.054	2943	0.946
Two or More Races	47032	46780	99.464	252	0.536
Unidentified Race	17020	16950	99.589	70	0.411
LEP Status	108817	108573	99.776	244	0.224
IDEA Indicator	81346	81203	99.824	143	0.176
Section 504 Status	10888	10881	99.936	7	0.064
Economic Disadvantage Status	446999	444116	99.355	2883	0.645

TABLE 5-145 COUNT OF GRADE 7 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	807500	802416	99.370	5084	0.630
Female	394672	392144	99.359	2528	0.641
Male	412570	410014	99.380	2556	0.620
American Indian or Alaska Native	9403	9345	99.383	58	0.617
Asian	58999	58988	99.981	11	0.019
Black/African American	56826	55324	97.357	1502	2.643
Native Hawaiian or Pacific Islander	8854	8852	99.977	2	0.023
Hispanic/Latino Ethnicity	294202	293926	99.906	276	0.094
White	316143	313199	99.069	2944	0.931
Two or More Races	45043	44822	99.509	221	0.491
Unidentified Race	18030	17960	99.612	70	0.388
LEP Status	92508	92249	99.720	259	0.280
IDEA Indicator	77165	77042	99.841	123	0.159
Section 504 Status	12108	12096	99.901	12	0.099
Economic Disadvantage Status	433101	430301	99.353	2800	0.647

TABLE 5-146 COUNT OF GRADE 8 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	800758	795769	99.377	4989	0.623
Female	391184	388747	99.377	2437	0.623
Male	409318	406766	99.377	2552	0.623
American Indian or Alaska Native	9074	9020	99.405	54	0.595
Asian	58801	58791	99.983	10	0.017
Black/African American	56730	55291	97.463	1439	2.537
Native Hawaiian or Pacific Islander	8298	8297	99.988	1	0.012
Hispanic/Latino Ethnicity	290660	290400	99.911	260	0.089
White	316244	313302	99.070	2942	0.930
Two or More Races	42369	42165	99.519	204	0.481
Unidentified Race	18582	18503	99.575	79	0.425
LEP Status	78241	78015	99.711	226	0.289
IDEA Indicator	74709	74592	99.843	117	0.157
Section 504 Status	13070	13054	99.878	16	0.122
Economic Disadvantage Status	422261	419589	99.367	2672	0.633

TABLE 5-147 COUNT OF GRADE 11 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Paper-Pencil	
	N	N	P	N	P
Total	529447	528936	99.903	511	0.097
Female	258972	258742	99.911	230	0.089
Male	270268	269987	99.896	281	0.104
American Indian or Alaska Native	4995	4874	97.578	121	2.422
Asian	51389	51376	99.975	13	0.025
Black/African American	28324	28316	99.972	8	0.028
Native Hawaiian or Pacific Islander	5782	5781	99.983	1	0.017
Hispanic/Latino Ethnicity	238793	238770	99.990	23	0.010
White	158520	158193	99.794	327	0.206
Two or More Races	25562	25545	99.933	17	0.067
Unidentified Race	16082	16081	99.994	1	0.006
LEP Status	48621	48611	99.979	10	0.021
IDEA Indicator	48990	48904	99.824	86	0.176
Section 504 Status	10261	10246	99.854	15	0.146
Economic Disadvantage Status	283893	283668	99.921	225	0.079

Evaluation of Vertical Scales

Test characteristics curves (TCCs) by administration year, subject, and grade are presented in figure 5-1 (for ELA/literacy in grades 3-6) and figure 5-2 (for ELA/literacy in grades 7, 8, and 11) and figure 5-3 (for mathematics in grades 3-6) and figure 5-4 (for mathematics in grades 7, 8, and 11). The test characteristic curves of the 2015-16 and 2016-2017 operational administrations are compared in each plot for a given grade and subject area. In general, there are only slight differences between the TCCs by year of administration. The ELA/literacy tests of the 2016-17 operational administration appear to be slightly easier than those from the previous year at all grades. In comparison to 2015-2016 administration, the mathematics tests of the 2016-17 administration appear to be slightly easier at grades 3 to 5 and slightly harder at grades 6 to 11. As noted in Chapter 4, differences in overall test difficulty have virtually no effect on the difficulty of tests received by individual students due to the adaptive administration of test items. It is reasonable to conclude that the slight differences in TCCs across the two administrations represented in these figures had no effect on differences in the measured achievement of students across these administrations.

FIGURE 5-1 TEST CHARACTERISTICS CURVES FOR LINKED TESTS (ELA/LITERACY, GRADES 3-6)

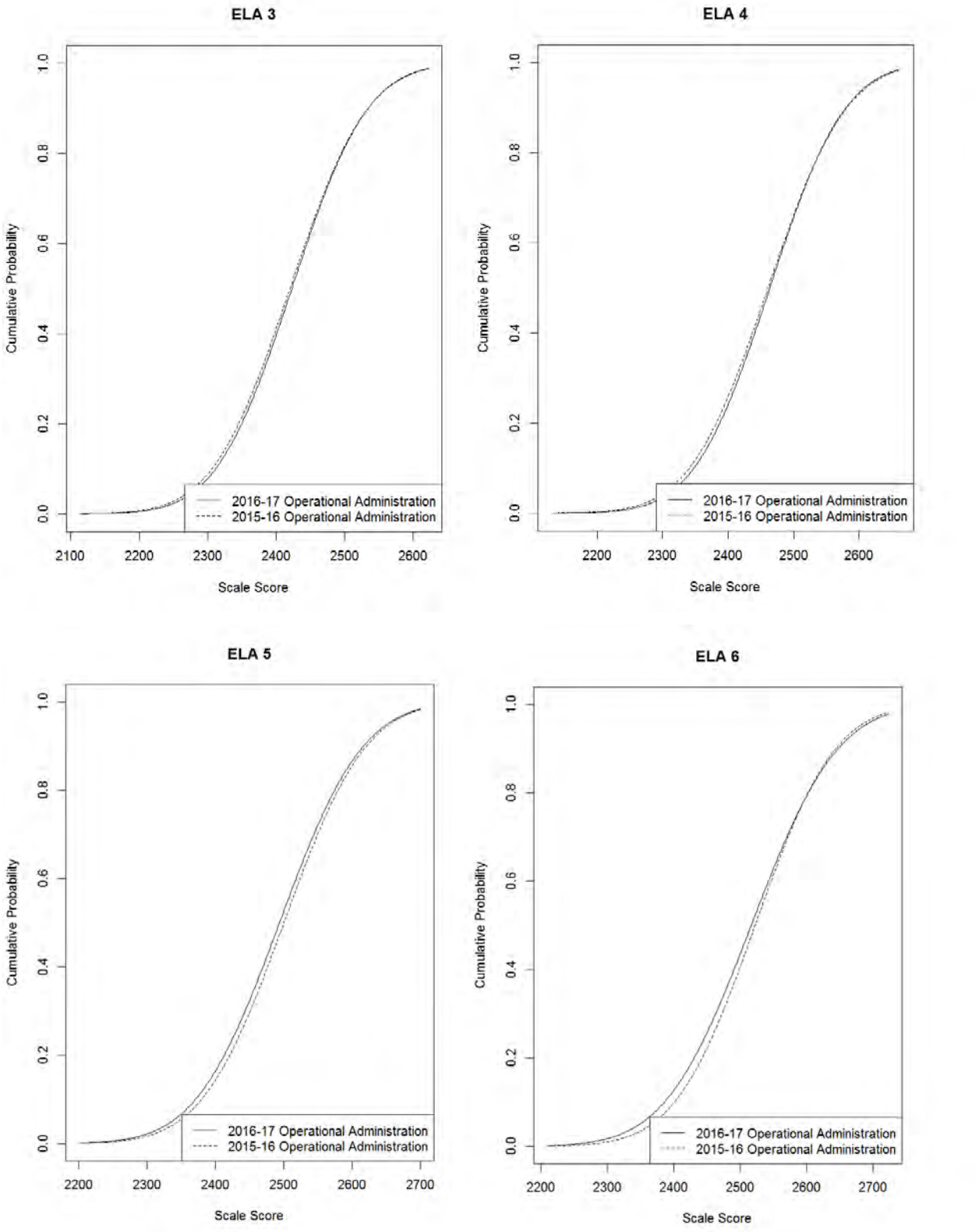


FIGURE 5-2 TEST CHARACTERISTICS CURVES FOR LINKED TESTS (ELA/LITERACY, GRADES 7, 8, AND 11)

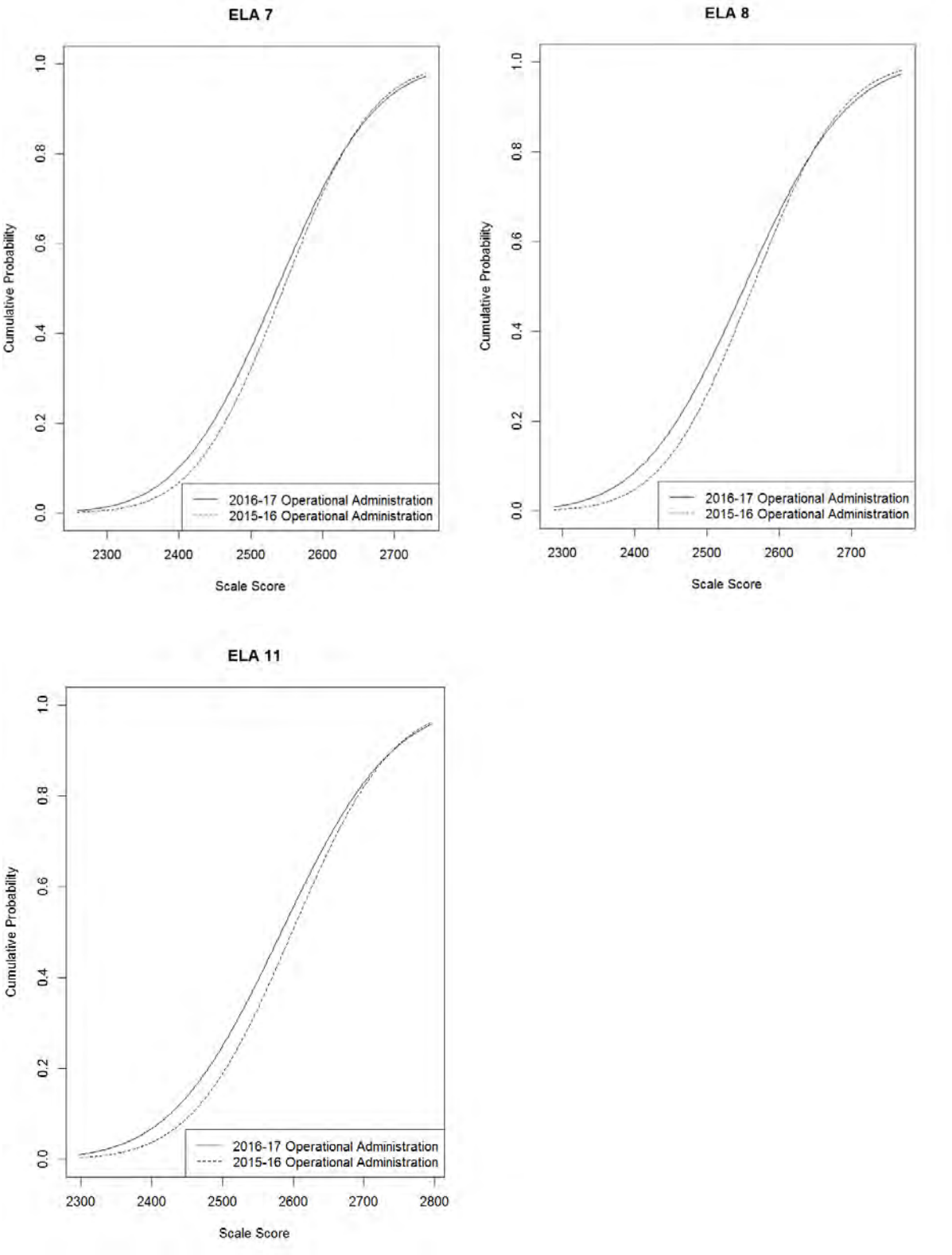


FIGURE 5-3 TEST CHARACTERISTICS CURVES FOR LINKED TESTS (MATH, GRADES 3-6)

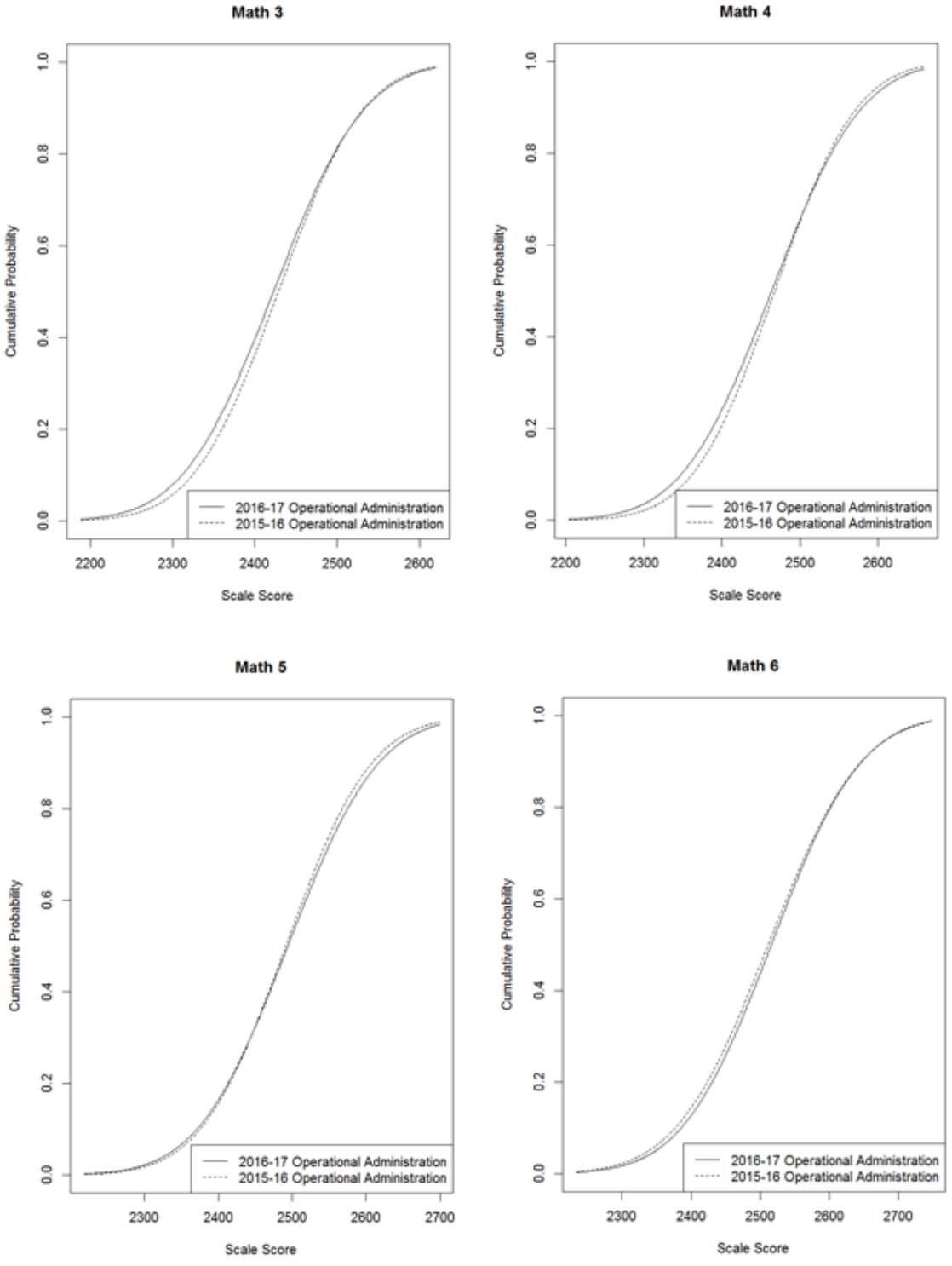
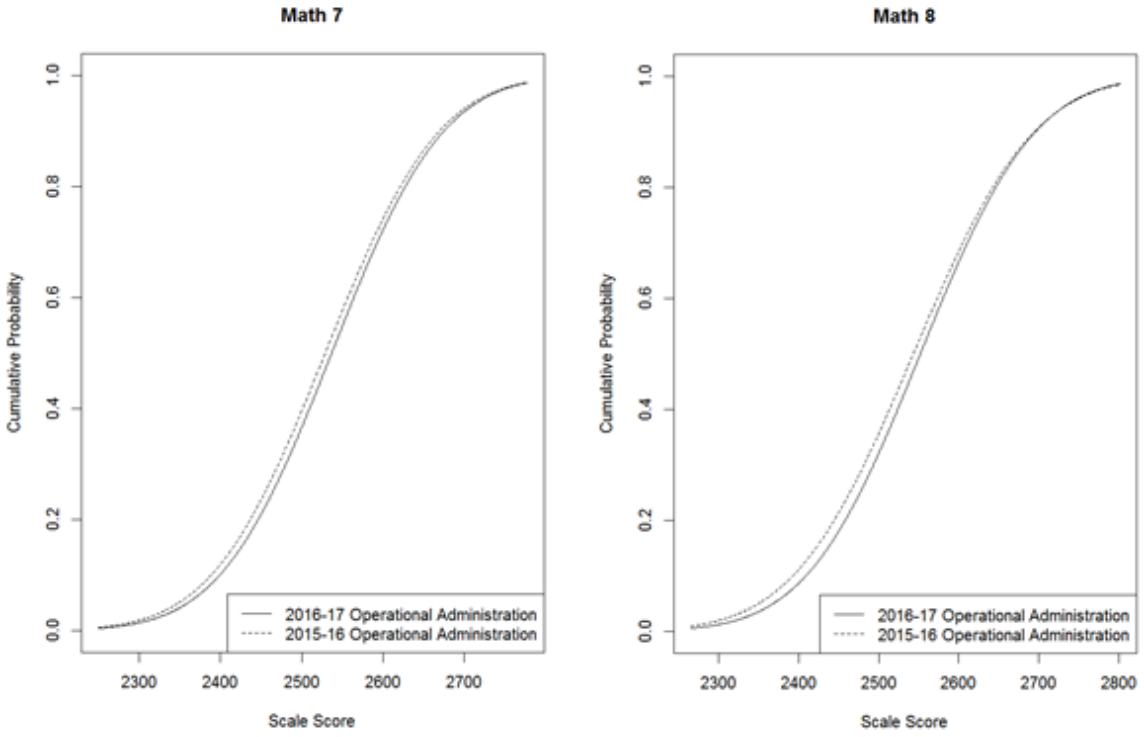
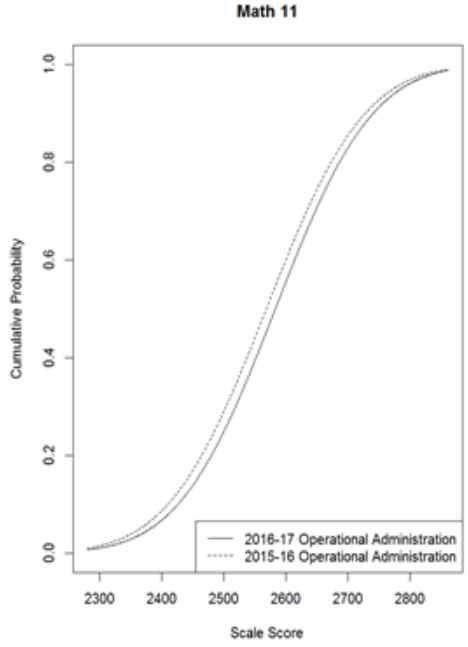


FIGURE 5-4 TEST CHARACTERISTICS CURVES FOR LINKED TESTS (MATH, GRADES 7, 8, AND 11)





Test characteristics curves (TCCs) by grade are shown for ELA/literacy and Mathematics in figure 5-5 and figure 5-6 respectively. The TCCs show that the difficulty of the tests increase by grade level for all levels of student proficiency. At any given scale score, the expected proportion correct decreases with grade level. This pattern is an expected outcome of overall test design and item writing specifications. Educational tests are expected to represent knowledge, skills, and abilities that increase in difficulty across grade levels.

FIGURE 5-5 TEST CHARACTERISTICS CURVES FOR VERTICALLY SCALED TESTS (ELA/LITERACY)

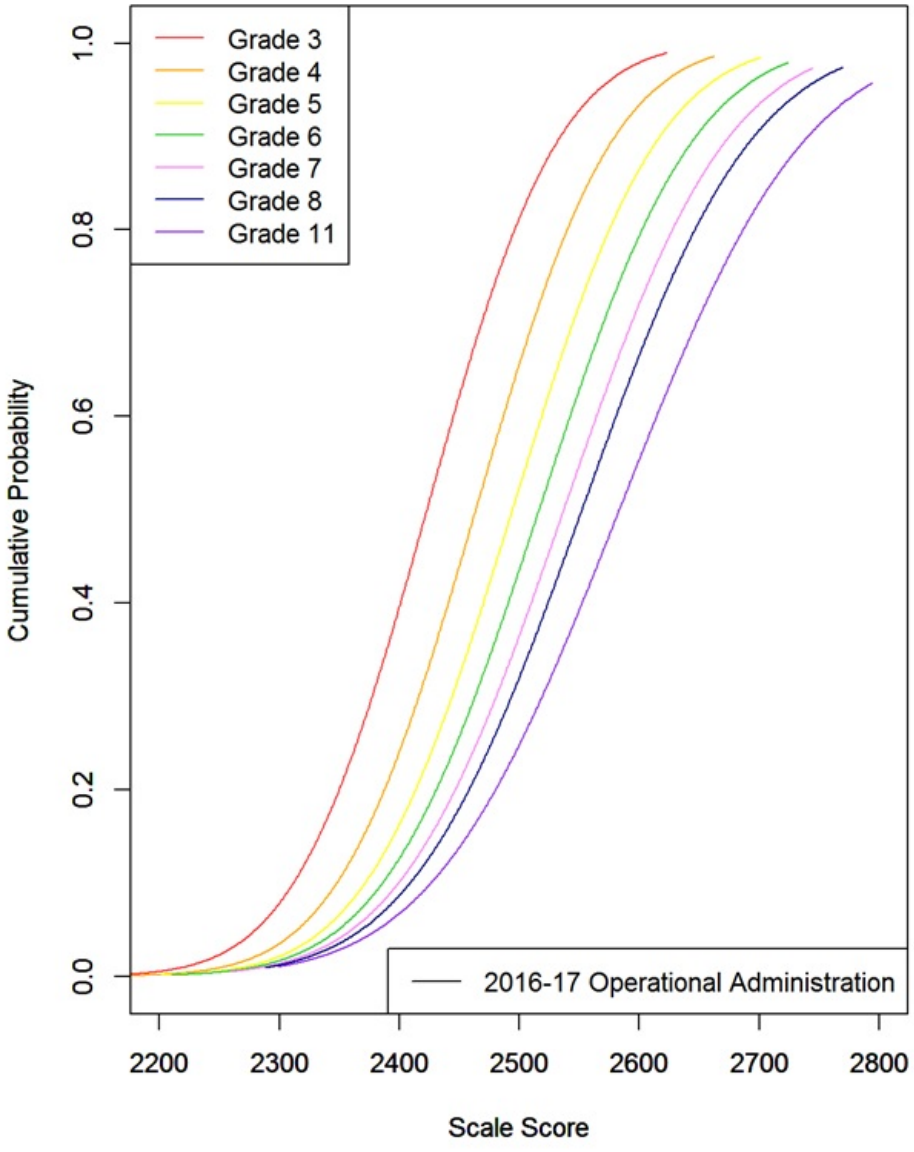
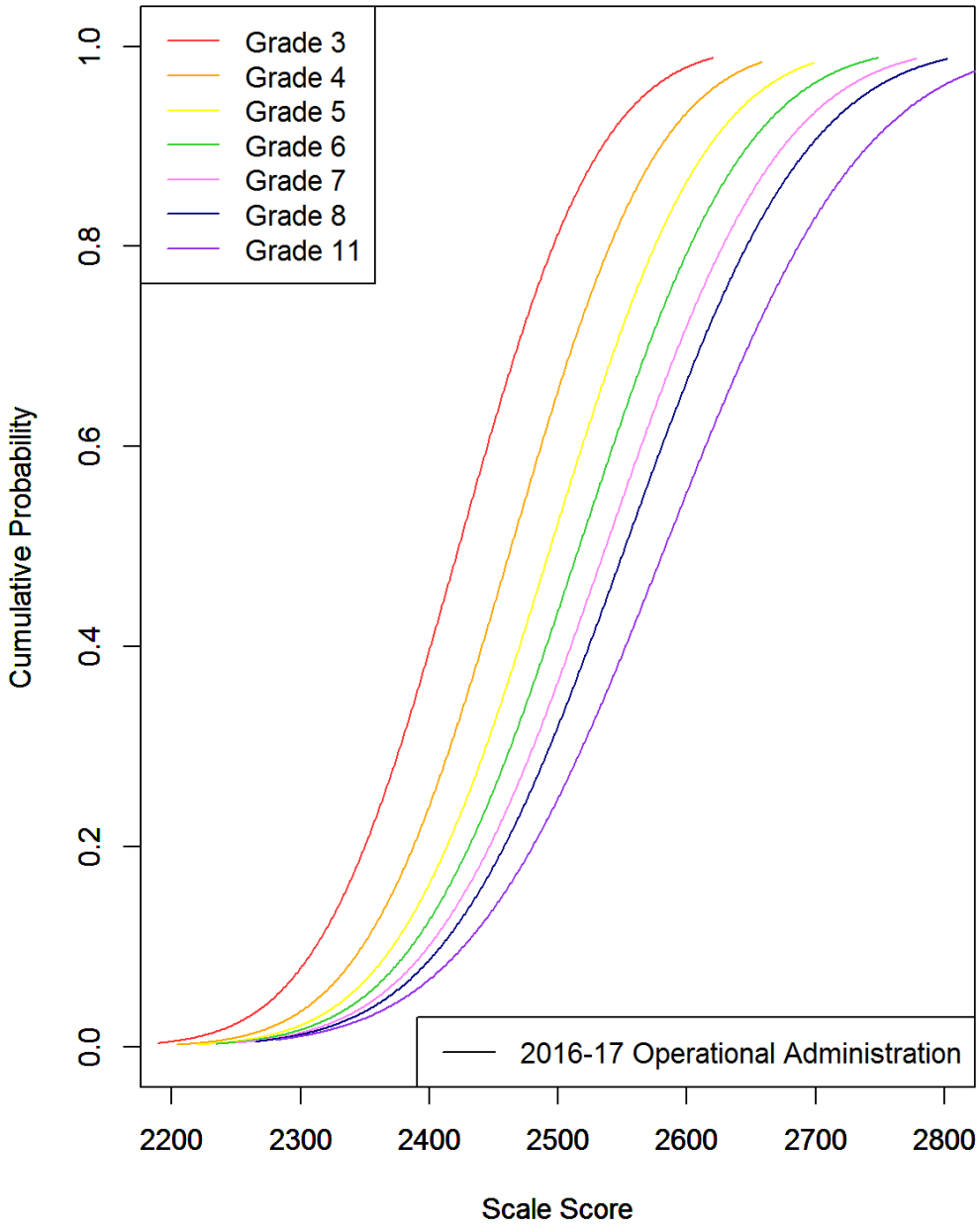


FIGURE 5-6 TEST CHARACTERISTICS CURVES FOR VERTICALLY SCALED TESTS (MATH)



References

Birnbaum, A. (1968). Some Latent trait models and their use in inferring an examinee's ability. pp. 395-479. In Lord, F. M. and Novick, M. R. *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley.

Cizek, G. J. & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

CTB/McGraw Hill. (2014). *Technical report: Initial achievement level descriptors.* Los Angeles, CA. Smarter Balanced Assessment Consortium. Retrieved from: http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/

Lewis, D. M., Mitzel, H. C., Mercado, R. L, & Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations*. New York: Routledge.

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika,* 47, 149–174.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Muraki, E., & Bock, R. D. (2003). *PARSCALE 4.1: IRT based item analysis and test scoring for rating-scale data.* Chicago, IL: Scientific Software International, Inc.

Smarter Balanced Assessment Consortium. (2015a, January 7). *Achievement level setting final report.* Los Angeles, CA: Author.

Smarter Balanced Assessment Consortium. (2016, January 14). *2013-2014 Technical Report*. Los Angeles, CA: Author. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/2013-14_Technical_Report.pdf

Chapter 6: Test Administration

Introduction

"The usefulness and interpretability of test scores require that a test be administered and scored according to the developer's instructions" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 111). Smarter Balanced created and disseminated a customizable test administration manual (2014c) to ensure standardized test administration procedures and, thus, uniform test administration conditions for all students in Smarter Balanced member states. This chapter describes the customizable Smarter Balanced Online Test Administration Manual; presents operational item exposure rates and blueprint fidelity; and shows results for the embedded field test, including item scoring processes and inter-rater reliability of field tested items.

Test Administration

Students in Smarter Balanced member states participated in the 2016-2017 test administration when a specified percentage of the school year had occurred. Each state established a schedule for the administration of the Smarter Balanced summative assessments using a testing window as defined below:

Grades 3–8

- Testing shall not begin until at least sixty-six percent (66%) of a school's annual instructional days have been completed, and
- Testing may continue up to and including the last day of school.

Grade 11

- Testing shall not begin until at least eighty percent (80%) of a school's annual instructional days have been completed, and
- Testing may continue up to and including the last day of school.

States were allowed to establish more specific windows within the constraints of the Consortium-defined windows described above.

Session Time

Table 2.1 presents the estimated testing times. These were provided within the *Online Test Administration Manual* (Smarter Balanced, 2017a). The estimated times for each session of each content area test provides sufficient time for students to attempt all items.

TABLE 6-148 ESTIMATED TESTING TIMES FOR SMARTER BALANCED ASSESSMENTS

Content Area	Grades	Computer Adaptive Test (CAT) items hrs : mins	Performance Task (PT) hrs : mins	Total hrs : mins
English Language Arts/Literacy	3–5	1: 30	2:00	3:30
	6–8	1:30	2:00	3:30
	HS	2:00	2:00	4:00
Mathematics	3–5	1:30	1:00	2:30
	6–8	2:00	1:00	3:00
	HS	2:00	1:30	3:30

Test Administration Manual

The Smarter Balanced Assessment Consortium *State Procedures Manual* provides a high-level overview of the assessment system, including expected policies and procedures for administration, required trainings, general information about the open source platform, information about the evidence states must provide to Smarter Balanced annually, procurement information, and links to resource documents. This document provides the core responsibilities that member states must assume in order to provide Smarter Balanced test results that are generalizable across states.

Specific instructions for member states to administer Smarter Balanced summative assessments are included in the *Test Administrator User Guide* (Smarter Balanced, 2016a), *the Online Test Administration Manual* (Smarter Balanced, 2017a), the *Paper Pencil Test Administration Manual for ELA* (Smarter Balanced, 2017c), and the *Pencil Paper Test Administration Manual for Mathematics* (Smarter Balanced, 2017d). Specific components of these user guides and manuals require customization to meet unique needs in each member state. These components include:

- Help Desk information
- Test expiration dates
- Administration and Registration Tools (ART) user roles
- State user roles
- Test security policy
- Links to where materials and modules are posted
- Test Security/Administration training policy
- Instructions for identifying and retrieving the Classroom Activity
- Role–specific checklists

The development of the Smarter Balanced test administration manuals were guided by the AERA, APA, and NCME 2014 *Standards*. In regard to test administration, the *Standards* provide guidance to test developers that the directions for test administration should be sufficiently clear to allow for standardized implementation in a variety of conditions (see Standard 4.15). In addition, the standards provide guidance that test developers should provide sufficient detail so that test takers can respond to items and tasks in the manner intended by the test developer (see Standard 4.16).

Clear Directions to Ensure Uniform Administration

Smarter Balanced test administration manuals include instructions that clearly articulate various aspects of the administration process. The TAM covers an extensive amount of material for events that occur before, during, and after testing. In addition, the TAM points the user to training materials that further provide detail and clarity to support reliable test administration by qualified test administrators. The details provided in the TAM describe the general rules of online testing, including; pause rules; scheduling tests; recommended order of test administration; classroom activity information; assessment duration, timing, and sequencing information; and the materials that the test administrator and students need for testing. All work together to ensure uniform test administration conditions across Smarter Balanced member states.

Section 8 of the TAM provides an overview of the universal tools, designated supports, and accommodations. All are further explicated in the Smarter Balanced *Assessment Consortium Usability, Accessibility, and Accommodations Guidelines* and the *Usability, Accessibility, and Accommodations Implementation Guide* (Smarter Balanced, 2017b).

Detailed Instructions for Test Takers

Section 10 of the TAM provides step-by-step instructions to test administrators (TA) on how to start a test session, monitor a test session, and end a test session. Throughout the steps, Smarter Balanced has embedded scripts that TAs are instructed to read to students. Test administrators are instructed to strictly adhere to scripts, use professional judgment when responding to student questions, and refrain from reading test items, suggesting answers, or evaluating student work during testing. See Section 10 of the online TAM for the script (Smarter Balanced, 2017a).

In addition, Smarter Balanced provides tutorials and practice tests[7] for each content area to familiarize students with how to navigate the online test delivery system and practice with the item types and the functionality of the testing environment. Together with the detailed instructions, the tutorials and practice tests assure that students are able to answer the items and tasks in the manner intended by Smarter Balanced.

Responsibilities of Test Administrators

The AERA, APA, and NCME *Standards* (2014) also provide guidance to test administrators and test users. Test administrators are guided to carefully follow the standardized procedures (Standard 6.1); inform test takers of available accommodations (Standard 6.2); report changes or disruptions to the

[7] http://www.smarterbalanced.org/assessments/practice-and-training-tests/

standardized test administration (Standard 6.3); furnish a comfortable environment with minimal distractions (Standard 6.4); provide appropriate instructions, practice, and other supports (Standard 6.5); and ensure the integrity of the test by eliminating opportunities for test taker malfeasance (Standard 6.6). In addition, test users are responsible for test security at all times (Standard 6.7). To align to these guidelines, the online TAM provides:

- careful direction to TAs to strictly adhere to the directions in the TAM;
- available universal tools, designated supports, and accommodations;
- requirements of the test environment, including student seating, providing a quiet environment, and access to allowable universal tools, designated supports, and accommodations;
- descriptions of testing improprieties, irregularities, and breaches.

Deviations from standardized online testing procedures (specifically testing irregularities and improprieties are handled at the local and/or state level, per the guidelines in the Online TAM. Depending on the nature and severity of the incident, a student's test may be reset, invalidated, reopened, or restored. All such incidents must be reported by authorized administrators at the local level to the state level (Smarter Balanced, 2017a).

Tools, Designated Supports, and Accommodations

To enhance student access to the assessment content during test administration, Smarter Balanced developed a conceptual model that included universal tools, designated supports, and accommodations (Smarter Balanced, 2017b). Universal Tools are access features of the Smarter Balanced assessment that are either provided as digitally-delivered components of the test delivery system (embedded) or provided separately from the test delivery system (non-embedded). Universal tools are available to all students based on student preference and selection. Embedded universal tools include (but are not limited to) such features as a "pause" feature that allows the student to take a break of 20 minutes or less during the assessment; a digital calculator that the student may access by clicking on a calculator button; and a digital notepad. Non-embedded universal tools include (but are not limited to) provision of an English dictionary for the full-write portion of the ELA/literacy performance task and the provision of physical scratch paper for all content area tests.

Designated supports for the Smarter Balanced assessments are embedded and non-embedded features that are available for use by any student for whom the need has been indicated by an educator or team of educators (along with the student and his/her parent/guardian) familiar with the student's instructional needs. Embedded designated supports include (but are not limited to) such features as color contrast, which enables students to adjust background or font color; translated test directions, translated glossaries, and stacked translations for mathematics items. Non-embedded designated supports include (but are not limited to) provision of color overlays; printing test content with different colors; use of magnification devices; and use of noise buffers.

Accommodations are changes in procedures or materials that increase equitable access during the Smarter Balanced assessments. Students receiving accommodations must have a need for those accommodations documented in an Individualized Education Program (IEP) or 504 accommodation plan. Like universal tools and designated supports, accommodations may be either embedded or non-embedded. Examples of embedded accommodations include (but are not limited to) closed

captioning and test content translated into American Sign Language (ASL) video. Non-embedded accommodations include (but are not limited to) use of an abacus, print on demand, and use of an external communication device (speech-to-text). Universal tools, designated supports, and accommodations all yield valid scores that count as participation in assessments that meet the requirements of ESEA when used in a manner consistent with the Smarter Balanced *Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2017b). A complete summary of all embedded and non-embedded universal tools, designated supports, and accommodations is included in the *Usability, Accessibility, and Accommodations Guidelines*.

Item Exposure Rates

Item exposure rates were obtained using online and adaptive test instances with valid scale scores for which item data were available from the 2016-2017 summative administration. The exposure rate for a given item is the proportion (or percentage) of test instances in the grade and content area on which the item appeared.

Table 6-2 and Table 6-3 presents a summary of the item exposure results for ELA/literacy and mathematics, respectively. Within each grade and component (CAT and PT), both tables present the number of items in the operational pool (N), along with various descriptive statistics, including the mean, standard deviation (SD), range (Min, Max), and median of the observed exposure rates. For example, table 6-2 shows that, on average, each item eligible for administration at grade 3 was seen by 7% of Grade 3 examinees. As a rule of thumb, Smarter Balanced attempts to maintain a maximum exposure rate of 25% (i.e., 25% of examinees will see the same item). Table 6-2 shows that the mean and median exposure rates for ELA/literacy items are well below 25%. Table 6-3 shows that the mean and median exposure rates for mathematics items are also well below 25%.

Table 6-4 and Table 6-5 provide further information about item exposure by showing the number of and proportion of items in the operational pool (N) with exposure rates falling into certain ranges (bins with a width of 0.1), including those that were completely unexposed (Unused). Table 6-4 shows over 80% of ELA/literacy CAT items had exposure rates between 0 and 20%. Also, nearly all of the PT items had item exposure rates between 0 and 20%. Table 6-5 shows the majority of mathematics CAT items had item exposure rates between 0 and 20%. About 70% of PT items for Grades 3, 6, 8 and 11 had item exposure rates between 0 and 20%, while the rest were unused. For Grades 4, 5, and 7, 100% of PT items had item exposure rates between 0 and 20%.

In both content areas, there were a handful of items with exposure rates over 50%. This might indicate cases in which there were few items available fulfilling specific blueprint requirements. The properties of these items needs to be further investigated. Future item development could help provide greater pool depth to reduce overexposure.

TABLE 6-149 SUMMARY OF ELA/LITERACY ITEM EXPOSURE RATES BY GRADE AND COMPONENT

Grade	Type	N	Mean	SD	Min	Max	Median
3	CAT	872	0.04	0.07	0.00	0.43	0.02
4	CAT	857	0.05	0.06	0.00	0.51	0.03
5	CAT	845	0.05	0.07	0.00	0.55	0.02
6	CAT	785	0.05	0.08	0.00	0.68	0.02
7	CAT	705	0.06	0.08	0.00	0.46	0.02
8	CAT	786	0.05	0.08	0.00	0.57	0.02
11	CAT	2397	0.02	0.03	0.00	0.43	0.01
3	PT	49	0.09	0.03	0.07	0.14	0.07
4	PT	64	0.07	0.02	0.05	0.11	0.05
5	PT	74	0.06	0.02	0.05	0.13	0.05
6	PT	48	0.09	0.03	0.07	0.14	0.07
7	PT	60	0.07	0.02	0.05	0.10	0.05
8	PT	68	0.06	0.02	0.05	0.10	0.05
11	PT	81	0.05	0.02	0.04	0.08	0.04

TABLE 6-150 SUMMARY OF MATHEMATICS ITEM EXPOSURE RATES BY GRADE AND COMPONENT

Grade	Type	N	Mean	SD	Min	Max	Median
3	CAT	1194	0.03	0.04	0.00	0.52	0.01
4	CAT	1292	0.03	0.03	0.00	0.24	0.02
5	CAT	1255	0.03	0.03	0.00	0.29	0.02
6	CAT	1100	0.03	0.05	0.00	0.45	0.02
7	CAT	961	0.04	0.06	0.00	0.46	0.01
8	CAT	859	0.04	0.08	0.00	0.75	0.01
11	CAT	2631	0.01	0.03	0.00	0.25	0.00
3	PT	81	0.07	0.00	0.07	0.07	0.07
4	PT	96	0.05	0.00	0.05	0.05	0.05
5	PT	86	0.07	0.00	0.06	0.07	0.07
6	PT	73	0.08	0.00	0.08	0.08	0.08
7	PT	88	0.06	0.00	0.06	0.06	0.06
8	PT	59	0.08	0.00	0.08	0.08	0.08
11	PT	62	0.08	0.00	0.08	0.09	0.08

TABLE 6-151 PROPORTION OF ELA/LITERACY ITEMS BY EXPOSURE RATES

Grade	Type	N	Unused	(0.0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]	(0.9, 1.0]
3	CAT	887	0.02	0.85	0.09	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00
4	CAT	866	0.01	0.89	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
5	CAT	862	0.02	0.85	0.07	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00
6	CAT	812	0.03	0.84	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	CAT	731	0.04	0.81	0.09	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00
8	CAT	800	0.02	0.84	0.07	0.05	0.01	0.01	0.00	0.00	0.00	0.00	0.00
11	CAT	2463	0.03	0.96	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	PT	51	0.06	0.67	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	PT	69	0.09	0.65	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	PT	78	0.06	0.92	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	PT	54	0.13	0.61	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	PT	72	0.17	0.57	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	PT	79	0.14	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	PT	95	0.16	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 6-152 PROPORTION OF MATHEMATICS ITEMS BY EXPOSURE RATES

Grade	Type	N	Unused	(0.0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]	(0.9, 1.0]
3	CAT	1193	0.00	0.95	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	CAT	1291	0.00	0.96	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	CAT	1254	0.00	0.96	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	CAT	1099	0.00	0.94	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
7	CAT	960	0.00	0.89	0.06	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
8	CAT	858	0.00	0.90	0.07	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00
11	CAT	2630	0.00	0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	PT	0.02	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
4	PT	0.04	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
5	PT	0.03	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
6	PT	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	PT	0.03	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
8	PT	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	PT	0.02	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02

Blueprint Fidelity

A key design document of the summative assessments is the test blueprint, which specifies the number of items by claim, target, depth of knowledge to be administered. A blueprint fidelity study is conducted after each operational administration to assess the quality of the item pool and the functioning of the CAT algorithm to meet blueprint requirements. A full report of the blueprint fidelity analysis for the 2016-2017 summative assessment is available (Smarter Balanced, 2018a). The following paragraphs is a summary of results.

Analyses using 2016-2017 operational assessment data were performed for both ELA/literacy and mathematics and in all the tested grade levels (3-8 and high school). At each grade level, three ELA/literacy item pools were evaluated: the General pool, the Braille pool, and the American Sign Language (ASL) pool. For mathematics five pools were evaluated: the General pool, the Braille pool, the ASL pool, the Spanish pool, and the Translated Glossaries pool. Only operational items from the computerized adaptive test (CAT) component were considered in this study; field test items and performance task (PT) component items were not included.

The analyses showed that the operational tests delivered in the 2016-2017 administration fulfilled the blueprint requirements very well. Virtually all tests delivered to students in the general population met blueprint requirements for the number of items per claim, target group, and depth of knowledge (DOK) within claim. No fewer than 99.35 percent of ELA tests and 99.98 percent of Math tests met requirements for the number of items per claim. Deviations from the blueprint were more often a case of one too many, rather than too few, items being delivered for a given blueprint specification. The vast majority of deviations were within one point of the specified item count.

For purposes of future item development, Smarter noted the few combinations of requirements that were met by fewer than 90% of the tests delivered. These case were more likely to occur for combinations of claims, targets and depth of knowledge requirements and within certain grades and accommodations pools. In the worst case, 34% of mathematics tests delivered in Spanish to grade 3 students met blueprint requirements for the number of items representing target group A or D within Claim 3 (3 items). However, 96% of the tests delivered in this case were within 1-point of meeting the exact requirement.

Deviations from blueprint requirements, though rare, are always investigated by Smarter for the possibility that there may be systematic shortages or surpluses of items in some areas of the blueprint that should be addressed through item development. The possibility that the CAT algorithm should be adjusted or that the blueprint is more restrictive than necessary and should be modified are also considered. Also considered are the sample sizes for some groups, such as Braille students, which can be quite small and cause blueprint fidelity percentages to fall below a certain threshold, such as 90%, by chance.

Spring 2016 and Spring 2017 Embedded Field Test Results

A total of 1,074 items were field tested in the 2016 and 2017 administrations. (These administrations correspond to, respectively, the 2015-2016 and 2016-2017 school years.) In 2016 only 243 CAT items were field tested – 100 in mathematics and 143 in ELA/literacy. In 2017, 831 items were field tested – 385 in mathematics and 446 in ELA/literacy. The 2017-field tested items

included 322 performance task (PT) items – 147 in mathematics and 175 in ELA/literacy. The PT items were organized into 4 or 5 distinct and non-overlapping sets of items per grade within subject.

The PT sets within ELA/literacy included 35 full write stimuli. Each full write stimulus was represented in scoring by two partial credit items – one two-point item representing the conventions rubric and one four-point item representing the rounded-up average of the four-point rubrics for evidence/elaboration and organization/purpose. With the 35 full write items being represented by two scored items, a total of 1,074 field test items were processed statistically.

Table 6-6 presents the number of field test items administered at each grade level by component – CAT or PT.

TABLE 6-153 NUMBER OF FIELD TEST ITEMS ADMINISTERED BY COMPONENT WITHIN GRADE

Grade	ELA/literacy		Math	
	CAT	PT	CAT	PT
3	56	25	65	20
4	40	25	41	22
5	45	25	34	20
6	52	25	48	22
7	78	25	51	20
8	77	25	61	22
HS	66	25	38	21
Total by component within subject:	414	175	338	147
Total by Subject:	589		485	

General Procedures and Results

This section describes procedures and results more or less applicable to all items including hand-scored items. Classical item analyses yielding average item scores and item-total correlations are applicable to all items. Analyses of differential item functioning are applicable to all items. All items are flagged for quality control using statistical criteria applicable to their item type. Distractor analyses are performed for machine-scored, selected-response (SR) items only but results are included in this section.

Classical Item Analyses

Classical item statistics were calculated for all field test items using the procedures described in Chapter 5. Table 6-7 and table 6-8 show for ELA/literacy and mathematics respectively the average item score and the average item-total correlation (point biserial) for the field test items by grade. The average item score is on a 0-to-1 point scale for all items with lower averages indicating harder

items. For dichotomously-scored items, the average item score is also called the item's p-value. For polytomously-scored (partial credit) items, the average item score is the item's average score divided by the maximum possible score.

Table 6-7 shows that the average item score by grade in ELA/literacy was in the range of 0.31 to 0.40, indicating that, on average, the items were difficult for the students. Table 6-7 also shows that the average item-total correlation by grade level ranged from 0.36 to 0.41, indicating that the items tend to work well together to differentiate students with higher overall performance from students with lower overall performance.

Table 6-8 presents the similar information for the mathematics items. The average item score in mathematics by grade ranged from a low of 0.24 in Grade 11 to a high of 0.43 in Grade 5, indicating that the items tended to be difficult for students. The mean item-total correlations ranged from 0.42 to 0.48.

TABLE 6-154 CLASSICAL ITEM STATISTICS FOR ELA/LITERACY FIELD TEST ITEMS

Grade	Number of Items	Average Item Score		Item-Total Correlation*	
		Mean	SD	Mean	SD
3	81	0.32	0.18	0.38	0.12
4	65	0.35	0.21	0.41	0.14
5	70	0.36	0.21	0.40	0.15
6	77	0.40	0.20	0.41	0.16
7	103	0.35	0.18	0.40	0.15
8	102	0.37	0.20	0.37	0.18
11	91	0.33	0.20	0.36	0.16
Total:	589				

*Total score was the overall scale score based on the operational test items.

TABLE 6-155 CLASSICAL ITEM STATISTICS FOR MATHEMATICS FIELD TEST ITEMS, SPRING 2015

Grade	Number of Items	Average Item Score		Item-Total Correlation*	
		Mean	SD	Mean	SD
3	85	0.42	0.22	0.43	0.15
4	63	0.41	0.20	0.48	0.12
5	54	0.43	0.22	0.47	0.10
6	70	0.33	0.18	0.44	0.11
7	71	0.34	0.21	0.42	0.13
8	83	0.36	0.20	0.43	0.11
11	59	0.24	0.14	0.42	0.15
Total:	485				

* Total score was the overall scale score based on the operational test items.

Item Flagging

Items were flagged using the criteria shown in Table 4-2 with the exception that flag R was conditional on the item-total correlation being less than 0.1. Table 6-9 presents an abbreviated description of the flagging criteria and shows the number of items flagged according to each criterion by subject. The type of item to which criteria were applied are identified in the table and the footnote. A single item may be flagged by more than one criterion, but not all criteria apply to all items. Hand-scored items (HS) are generally, but not always partial credit (PC) items. Selected response (SR) items include only items with a single correct answer and multiple distractors.

Overall, 23% of ELA/literacy items (134 of 589) and 20% of mathematics items (96 of 485) items were flagged by one or more criteria. The process and results of applying DIF and IR criteria are described in detail in the following sections.

TABLE 6-156 NUMBER OF FIELD TEST ITEMS FLAGGED BY CRITERION WITHIN SUBJECT

Flag	Definition	Item Type*	Number Flagged	
			ELA/Literacy	Mathematcs
A	Average Item score < 0.1	All	39	35
B	Score percent < 3	PC	37	9
C	Higher average scale score for students in lower score category	PC	6	2
D	High-performing students more likely to select distractor	SR	28	15
F	Students selecting a distractor have higher scale score than students selecting correct answer	SR	13	1
H	Average item score > .95	All	0	0
P	Distractor has positive point biserial	SR	14	0
R	Item-total correlation < 0.1	All	25	5
DIF	Differential item functioning	All	31	42
IR	Low inter-rater agreement	HS	9	0
		Total flags:	202	109
		Total items flagged:	134	96

*All = any item; PC=partial credit items; SR=selected response item; HS=hand scored item.

Differential Item Functioning (DIF)

DIF statistics were computed for all field test items using the procedures described in Chapter 3. DIF was evaluated for eight subgroup comparisons (focal – reference)

- Gender: Female – Male
- Race/Ethnicity: Asian – White
- Race/Ethnicity: Black – White
- Race/Ethnicity: Hispanic – White
- Race/Ethnicity: Native American – White
- IEP: yes – no
- LEP: yes – no
- Title 1: yes – no

DIF categories/grades assigned based on Mantel-Haenszel chi-square and delta-DIF statistics and standardized mean effect sizes:

- A: negligible
- B: slight to moderate
- C: moderate to large

Table 6-10 and table 6-11 summarize the number of items flagged for DIF by grade in ELA/literacy and mathematics, respectively. The third column shows the number of items flagged for any moderate to large DIF (C DIF). Approximately 5% of ELA/literacy items and 7% of mathematics items were flagged for moderate to large DIF.

TABLE 6-157 NUMBER OF ELA/LITERACY FIELD TEST ITEMS FLAGGED FOR DIF, SPRING 2015

Grade	items	Any C	DIF	M/F	A/W	B/W	H/W	NA/W	IEP/no	LEP/no	Title1/no
3	81	3	A	81	65	40	77	0	77	73	81
			B	0	7	4	4	0	4	8	0
			C	0	3	0	0	0	0	0	0
4	65	1	A	65	58	20	61	0	47	54	60
			B	0	2	7	4	0	17	5	5
			C	0	0	0	0	0	1	0	0
5	70	7	A	69	58	29	66	0	54	49	67
			B	1	5	2	4	0	9	9	3
			C	0	0	0	0	0	7	0	0
6	77	8	A	77	69	36	73	0	56	49	70
			B	0	5	6	4	0	13	18	7
			C	0	0	0	0	0	8	0	0
7	103	2	A	101	93	57	98	0	86	45	96
			B	2	4	7	5	0	17	19	7
			C	0	2	0	0	0	0	0	0
8	102	4	A	98	94	62	97	0	87	47	97
			B	3	5	5	5	0	12	20	5
			C	1	0	0	0	0	3	0	0
11	91	6	A	89	78	57	86	0	80	43	82
			B	2	8	9	5	0	9	17	9
			C	0	5	0	0	0	0	1	0

TABLE 6-158 NUMBER OF MATHEMATICS FIELD TEST ITEMS FLAGGED FOR DIF, SPRING 2015

Grade	items	Any C	DIF	M/F	A/W	B/W	H/W	NA/W	IEP/no	LEP/no	Title1/no
3	85	6	A	80	70	56	73	0	71	65	69
			B	5	10	9	12	0	14	18	16
			C	0	5	0	0	0	0	2	0
4	63	1	A	59	54	36	58	0	43	44	57
			B	4	9	5	5	0	20	18	6
			C	0	0	0	0	0	0	1	0
5	54	9	A	52	47	28	50	0	36	43	48
			B	2	4	2	4	0	13	9	6
			C	0	3	1	0	0	5	0	0
6	70	7	A	64	63	16	65	0	61	54	64
			B	6	3	3	5	0	6	5	6
			C	0	4	0	0	0	3	0	0
7	71	3	A	64	64	6	63	0	61	10	60
			B	7	6	2	8	0	9	4	10
			C	0	0	0	0	0	1	1	1
8	83	8	A	80	71	14	72	0	64	9	68
			B	3	7	0	11	0	16	2	15
			C	0	5	0	0	0	3	1	0
11	59	8	A	54	47	20	55	0	51	10	48
			B	5	8	0	4	0	7	0	10
			C	0	4	0	0	0	1	3	1

Procedures and Results for Hand-Scored Items

A total of 288 items field tested in the 2016 or 2017 administrations were hand scored – 70 mathematics items and 218 ELA/Literacy items. Procedures and results in this section are applicable to hand-scored items exclusively.

Scoring Procedures

For the purpose of ensuring standardized scoring processes and standards, Smarter Balanced developed and implemented detailed training requirements, qualification standards, and scoring quality standards for all hand scored items. For field test hand-scoring, training procedures differed based on content area and item type. For the ELA/literacy full write items, readers were trained using anchor sets for a specific trait at a specific grade level. For ELA/literacy short text items, readers were trained by grade band for a claim and target subcategory. For mathematics hand scored PTs and short text CAT items, training was based on task models. Qualification standards were determined by the number of points available within a specific item as follows:

Item Points Available	Qualification Standard
0-1	90% (no non-adjacent scores)
0-2	80 % (no non-adjacent scores)
0-3	80% (no non-adjacent scores)
0-4	70% (no non-adjacent)

For field test scoring, a minimum of ten validity papers per item was presented to each reader with the expectation that the reader would maintain the following exact agreement standards:

Item Points Available	Exact Agreement
0-1	90%
0-2	80%
0-3	80%
0-4	70%

Scoring supervisors reviewed quality data including inter-rater reliability, validity check-set results, third-read adjudication results, item-level and reader-level reports on item score-point frequencies, and item-level reports showing mean scores throughout the scoring event.

Interrater Reliability Results

At least 10% of the field test responses in ELA/literacy and mathematics were scored independently by a second reader. The statistics for the inter-rater reliability were calculated for all items at all grades. To determine the reliability of scoring, the percentage of perfect agreement and adjacent agreement between the two readers was examined. Additionally, the item-level quadratic weighted kappa statistic was calculated to reflect the level of improvement beyond the chance level in the consistency of scoring.

Items are flagged for review and possible elimination from the item pool based on criteria for exact agreement among independent raters during operational scoring. The criteria for exact agreement depends on the number of points in the scoring rubric as shown below.

Score Point Range	Exact Agreement
0-1	80%
0-2	70%
0-3	70%
0-4	60%

Table 6-12 shows the number of items flagged by subject and grade. There were 9 items flagged across all grades in the two content areas. No mathematics items were flagged.

TABLE 6-159 NUMBER OF HAND-SCORED FIELD TEST ITEMS FLAGGED BY SUBJECT AREA AND GRADE, SPRING 2015

Subject	Grade	Number of Flagged Items
ELA/literacy	6	2
ELA/literacy	7	3
ELA/literacy	8	2
ELA/literacy	11	2

Data Review

All items flagged by one or more criteria were submitted for data review. Measurement Incorporated (MI) was contracted to conduct the data review. With the participation and input of educators from SBAC member states, MI psychometricians and content experts facilitated the examination of the flagged items and made recommendations as to each item's suitability for use on an operational or interim test. The data review process is described in detail in a separate report (Smarter Balanced, 2018b).

Table 6-13 summarized the outcome of the data review process. Sixty-eight percent of mathematics items and fifty-five percent of ELA/literacy items were recommended for acceptance as operational items without revisions. The remaining items were recommended to be rejected or revised or a consensus was not reached. .

TABLE 6-160 RECOMMENDATIONS FROM DATA REVIEW

	Mathematics		ELA/Literacy	
Number of Items:	96		134	
Accept	65	*(68%)*	74	*(55%)*
Reject	21	*(22%)*	48	*(36%)*
Revise	9	*(9%)*	12	*(9%)*
No Consensus	1	*(1%)*	0	*(0%)*

Smarter Balanced content and psychometric staff reviewed the recommendations from the data review process and also reviewed the statistics and content of all items not flagged. With one exception, all items not flagged were approved for operational use without revisions. The exception was an item representing the conventions score of a full writing item for which the item representing the 4-point score (the rounded average of the organization/purpose and evidence/elaboration scores) had been deleted.

Table 6-14 shows the final disposition of the items that had been sent to data review. In ELA/literacy, Smarter Balanced staff accepted fewer items than recommended by the data review committee and decided to reject several items recommended for revision. In mathematics, Smarter Balanced staff tended to accept the items recommended for acceptance by data review, but decided to revise, rather than reject more items. Scoring key errors were discovered for a few items.

TABLE 6-161 FINAL DISPOSITION OF ITEMS SENT TO DATA REVIEW

	Mathematics		ELA/Literacy	
Number of Items:	96		134	
Accept	66	*(69%)*	60	*45%)*
Reject	14	*(15%)*	71	*(53%)*
Revise	15	*(16%)*	1	*(1%)*
Rescore	1	*(1.0%)*	2	*(1%)*

When the data review results are combined with the flagging results, the percentages of mathematics items and ELA/literacy items accepted for operational use with no revisions were 94% (455 of 485) and 87% (514 of 589) respectively. These percentages are relatively high in the industry and reflect well on Smarter Balanced's item development process.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Smarter Balanced Assessment Consortium. (2017a, September). *Online, summative, test administration manual: Test of English language arts/literacy and mathematics.* Santa Cruz, CA: Author. https://portal.smarterbalanced.org/library/en/v2.0/online-test-administration-manual.docx

Smarter Balanced Assessment Consortium. (2017b, June). *Usability, Accessibility, and Accommodations Guidelines.* Santa Cruz, CA: Author. https://portal.smarterbalanced.org/library/en/v2.0/usability-accessibility-and-accommodations-guidelines.pdf

Smarter Balanced Assessment Consortium. (2017c, November). *English language arts summative assessment: Paper-pencil test administration manual*. Santa Cruz, CA: Author. https://portal.smarterbalanced.org/library/en/p-p-tam-ela-form-3-non-secure.docx

Smarter Balanced Assessment Consortium. (2017d, November). *Mathematics summative assessment: Paper-pencil test administration manual*. Santa Cruz, CA: Author. https://portal.smarterbalanced.org/library/en/p-p-tam-math-form-3.docx

Smarter Balanced Assessment Consortium. (2016a, November). *Test administrator user guide*. Santa Cruz, CA: Author. https://portal.smarterbalanced.org/library/en/v1.0/test-administrator-user-guide.docx

Smarter Balanced Assessment Consortium. (2016b, November). Administration and Registration Tools (ART) User's Guide. Santa Cruz, CA: Author. https://portal.smarterbalanced.org/library/en/v1.3/administration-and-registration-tools-art-user-guide.pdf

Smarter Balanced Assessment Consortium. (2018a, in prep). Blueprint fidelity analysis and results for the 2016-2017 Summative Assessment. Santa Cruz, CA: Author.

Smarter Balanced Assessment Consortium. (2018b, in prep). 2018 Data Review Process and Results. Santa Cruz, CA: Author.

Chapter 7: Reporting and Interpretation

Introduction

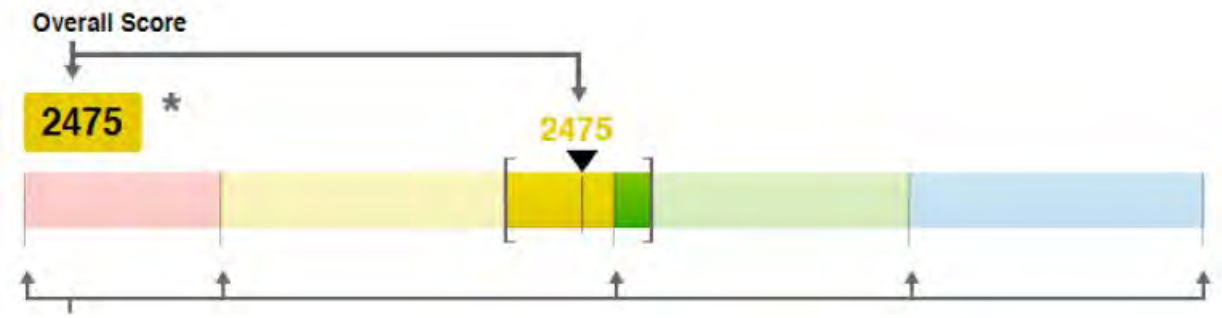
Scores from summative assessments provide information about student achievement with regard to college and career readiness. As noted in chapters on test design and scoring, summative tests provide an overall indicator of proficiency and a set of subscores corresponding to broad areas within the content area domains. The consortium provides a set of reports based on these scores and subscores that members may customize for their own use. This chapter provides an overview of the report system. For detailed information, consult the Smarter Balanced *Reporting System User Guide* (Smarter Balanced, 2016). Since use of the Smarter Balanced reporting system is optional and configurable, information about a specific member's reports should be gathered from member websites and materials. Smarter Balanced reports are based on information provided in the output of the test scoring algorithm. Overall scores and subscores each have an associated standard error of measurement (SEM) that indicates the reliability of the score. (For the definition of SEM, please refer to Chapter 2.)

Overall Test Scores

Scale scores are the basic units of overall reporting. These scores fall along a vertical scale (from approximately 2000 to 3000) that increases across grade levels and are used to describe an individual student's level of achievement, as well as to track growth over time (The growth report is undergoing research and development and has not been released to date.) When aggregated, scale scores are used to describe achievement for different groups of students. The method and process for setting achievement level standards to delineate proficiency levels is explained in Chapter 5.

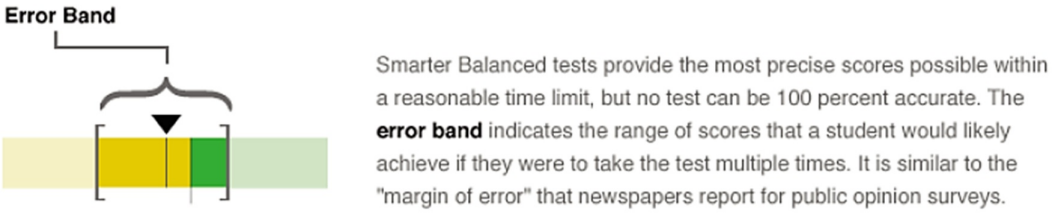
The Smarter Balanced reporting system communicates an overall scale score in relation to Achievement Levels using graphics similar to Figure 7-1. By default, the system uses generic terms for the achievement levels, Level 1, Level 2, Level 3, and Level 4. Consortium members may use terms such as "novice, developing, proficient, advanced" or others.

FIGURE 7-1 PORTRAYAL OF SCORE REPORTING LEVELS. FROM *REPORTING SYSTEM USER GUIDE*, P.13.



Scale scores are reported with an error band of ± one SEM. Smarter Balanced reporting provides information to help users understand the meaning of the error bands, as shown in Figure 7-2.

FIGURE 7-2 EXPLANATION OF ERROR BANDS DISPLAYED ON SMARTER BALANCED REPORTS. FROM *REPORTING SYSTEM USER GUIDE*, P.120.



Depicting errors and error bands in score reporting is an important measurement principle. In this portrayal, the score is represented by the vertical line and black triangle. The error band is shown by the brackets. If the test were to be given again, the score is likely to fall within this band. In Figure 7-1, the scale score of 2475 falls in Level 2, but the error band encompasses Level 3. This means there is a fair chance that the student's true score is in Level 3.

Smarter Balanced has developed a set of optional Reporting Achievement Levels for English language arts/literacy (ELA/literacy) and mathematics that are aligned with the Common Core State Standards (CCSS) and Smarter Balanced assessment claims. The intent of these descriptors is to specify, in content terms, the knowledge and skills that students may display at four levels of achievement. The full set of optional Reporting ALDs are shown in Appendix C.

Subscores

Subscores are reported for student performance on important domains within each content area. In most cases, subscores correspond to Claims. In mathematics, however, Claims 2 and 4 are so intertwined that they are reported as a single subscore. The Claims and reporting categories (subscores) are primary structural elements in the test blueprints and item development. Tables 7.1 and 7.2 provide the claims or subscore reporting categories for ELA/literacy and mathematics.

TABLE 7-1 ENGLISH LANGUAGE ARTS/LITERACY CLAIMS

Claim #1- Reading • Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.
Claim #2- Writing • Students can produce effect and well-grounded writing for a range of purposes and audiences.
Claim #3- Speaking and Listening • Students can employ effective speaking and listening skills for a range of purposes and audiences. At this time, only listening is assessed.
Claim #4- Research • Students can engage in research /inquiry to investigate topics and to analyze, integrate, and present information.

TABLE 7-2 MATHEMATICS CLAIMS AND SCORE REPORTING CATEGORIES

Claim #1- Concepts and Procedures • Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency.
Claim #2- Problem Solving/ Claim #4- Modeling and Data Analysis • Students can solve a range of complex, well-posed problems in pure and applied mathematics, making productive use of knowledge and problem solving strategies. Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems • Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems
Claim #3- Communicating Reasoning • Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.

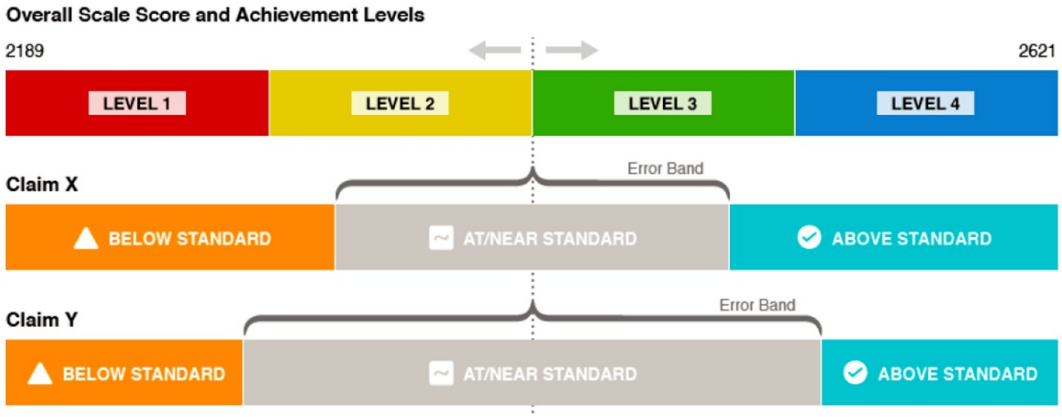
Achievement levels for claims are not established for reporting subscores as the overall test scores, partially because the SEM at the claim/subscore level is fairly large due to the limited number of items per claim. Consequently subscores are characterized by an indication of whether they are "Below Standard", "At or Near Standard", or "Above Standard". These designations are based on the SEM of the subscore and the difference of the subscore from the Level 3 cut score (the lower boundary of Level 3), which is an indicator of being on-track for college or career readiness. Table 7-3 presents the criteria for each designation.

TABLE 7-3 SUBSCORE CATEGORIES

Above Standard	The subscore is at or above the Level 3 cut score by more than 1.5 of its SEM.
At or Near Standard	The subscore does not meet the definition for above or below standard.
Below Standard	The subscore is below the Level 3 cut score by more than 1.5 of its SEM.

A practical way to understand this is portrayed in the graphic below. Instead of using error bands, it shows the reporting level area that would result from a scale score and SEM.

FIGURE 7-3 PORTRAYAL OF SUBSCORE REPORTING. FROM *REPORTING SYSTEM USER GUIDE*, PP.116-117.



Although subscores are portrayed in Smarter Balanced reports by the three-level system above (also called "traffic-light" indicators) the actual scale scores and SEMs associated with subscores are available to members in the data provided from the test scoring system. Members may use these in local reporting systems.

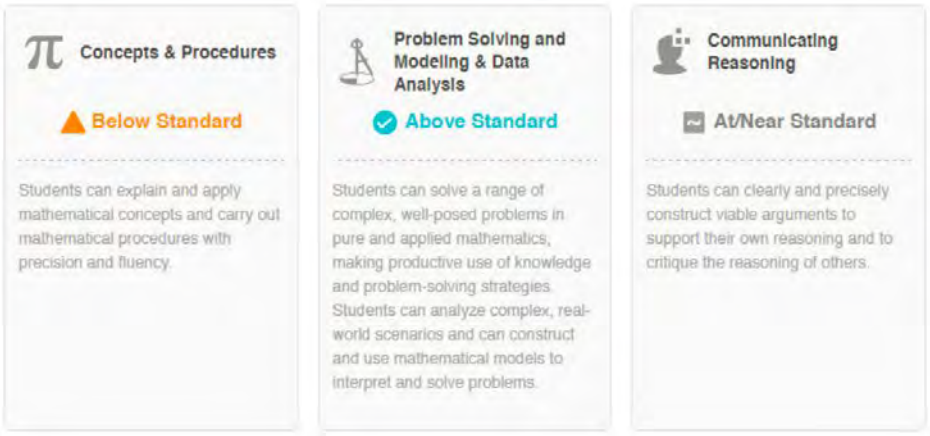
Types of Reports

The Smarter Balanced reporting system is an interactive, online reporting platform that provides a range of reports. Members can log into the system to create reports. Members can configure the system to show a state or groups logo or test name and can use their own labels for achievement levels. They can also use their own student groups. There are three basic report types: Individual student reports (ISRs), lists, and aggregate reports. These will be described briefly here, but the reader is urged to consult the Smarter Balanced *Reporting System User Guide* for more detail.

Individual Student Report (ISR)

This report presents individual student assessment scores, SEMs and achievement levels. They also display the reporting levels for claim/subscore results along with claim level ALDs (achievement level descriptions). The scores and descriptions provide context for understanding what the assessment has measured and how to interpret the scores and subscores. Teachers, students and parents use this report to understand a student's achievement and progress toward mastery of the CCSS (common core state standards). The report may be part of a larger set of information to provide context for instructional focus. In addition to the overall score displays, subscores are reported as shown in Figure 7-4 below.

FIGURE 7-4 ILLUSTRATION OF SUBSCORE REPORTING ON INDIVIDUAL STUDENT REPORTS

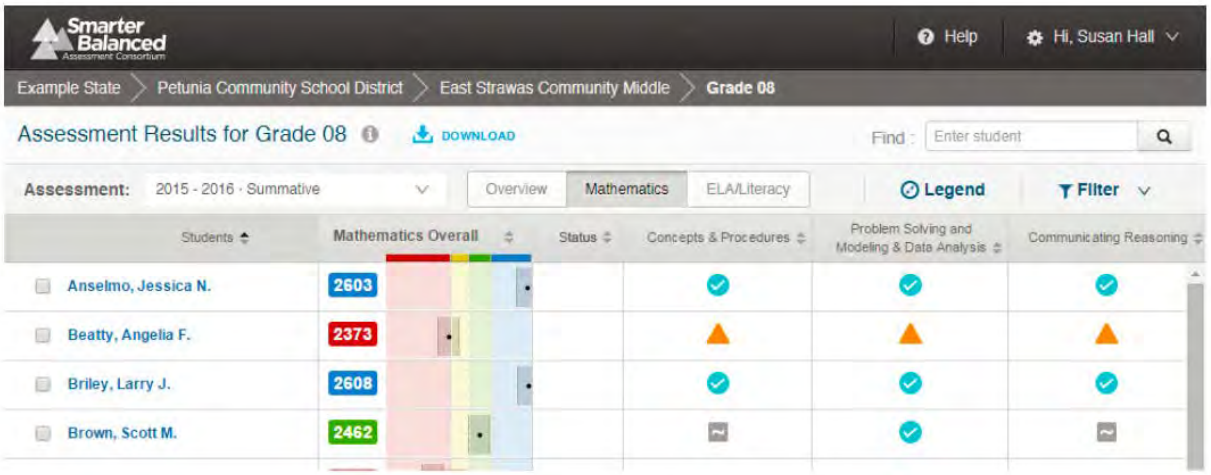


Individual Student Reports can be downloaded as PDF files for easy printing and distribution to parents.

Lists

Lists are generated for available groups. They are most commonly used at the school or district level, but may be used for other groupings if these are available to the system. Teachers and administrators commonly use lists to identify patterns across groups or to identify students most in need of assistance. Along with other information, lists can be used to provide a direction for further investigation about instructional emphasis or to aid in resource allocation. Figure 7-5 is an extract of a typical list report that presents a compact display of scores, errors, achievement categories and subscore levels. Note that lists can be filtered and sorted for different purposes. They may be filtered by gender or student demographic data (e.g. LEP, Race/Ethnicity, IEP, Gender, 504, Economic Disadvantage, or Migrant Status), or based on the completeness or validity of student test attempts.

FIGURE 7-5 EXAMPLE OF A LIST

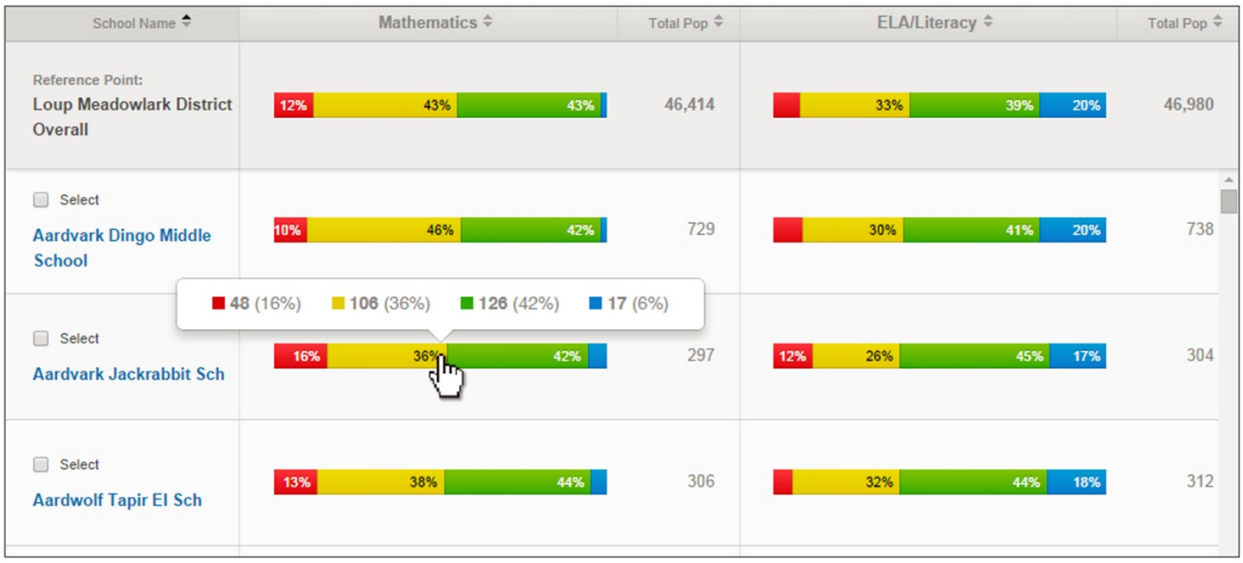


Aggregate Reports

Group aggregate reports provide score data at the state, district, school, and grade level. Educators may examine data at multiple levels, depending on their level of access, and can create custom subsets filtered in the same ways as list reports. Aggregate reports can be viewed onscreen or downloaded as CSV files that can be used in other reporting systems or combined with other data. In the Smarter Balanced reporting system aggregate reports show how groups are distributed across the four achievement levels. They are usually used to compare among groups or to identify areas of resource need. Like the reports above, aggregate reports can be filtered.

Figure 7-6 shows a district-level report with overall district results at the top and school results below. The shaded areas correspond to the achievement levels. Percentages of students in each level are shown. The figure shows the window that pops up when the cursor hovers over the display. It shows the numbers of students in each category as well as percentages.

FIGURE 7-6 EXAMPLE AGGREGATE REPORT



Data Downloads

In addition to the predesigned reports, the reporting system offers authorized users the ability to download data for distribution or further review and analysis in external systems. User authorization is closely controlled for ISRs and personally identifiable information (PII) in files. The list of available data downloads appears below. Note that these downloads assume that members have loaded data into the Smarter Balanced Data Warehouse. In practice, many members get this information directly from test delivery service providers and do not go through the Data Warehouse.

FIGURE 7-7DATA DOWNLOAD OPTIONS

Download Type	Description
Student Assessment Results	*This is a bulk download of the assessment results for the selected assessment, with one row per student. The resulting files contain all of the data for Overall and Claim scores (e.g., scale score, error band, level determination), as well as all the student data (e.g., demographics, grade/school/district/state attribution, etc.) for the specific summative or interim assessment being viewed.*
Printable Student Reports	*Printable versions of list and aggregate reports*
State Download: Student Registration Statistics	*This download shows statistics of registration records for a specified academic year and compares them to those of previous years to detect errors. This download is primarily intended for Consortium, state, and district administrators.*
State Download: Assessment Completion Statistics	*For a specified assessment administration, this download provides counts of registered and assessed students and percentages of students assessed. This enables an administrator to review how many of the registered students have been assessed.*
State Download: Audit XML	*This download ensures that all information for a given student assessment is maintained, including usage reports for Universal Tools and Designated Supports, as well as any additional data provided by a Test Delivery System*

Summary

Smarter Balanced reports tie together report categories, Achievement levels, and optionally the Reporting Achievement Level Descriptors to provide coherent information about student progress. Reporting categories are based on test structure which in turn reflects close analysis of the CCSS. In addition, the Smarter Balanced scale and Achievement levels were set by the comprehensive process described in Chapter 5.

The dynamic nature of the reports, allowing users to sort and filter to get custom information and the provision of customized download data for any kind of analysis, gives Consortium members a rich flexible set of results. By providing capability for multiple reports and downloads, the Smarter Balanced system provides members a dynamic and flexible system.

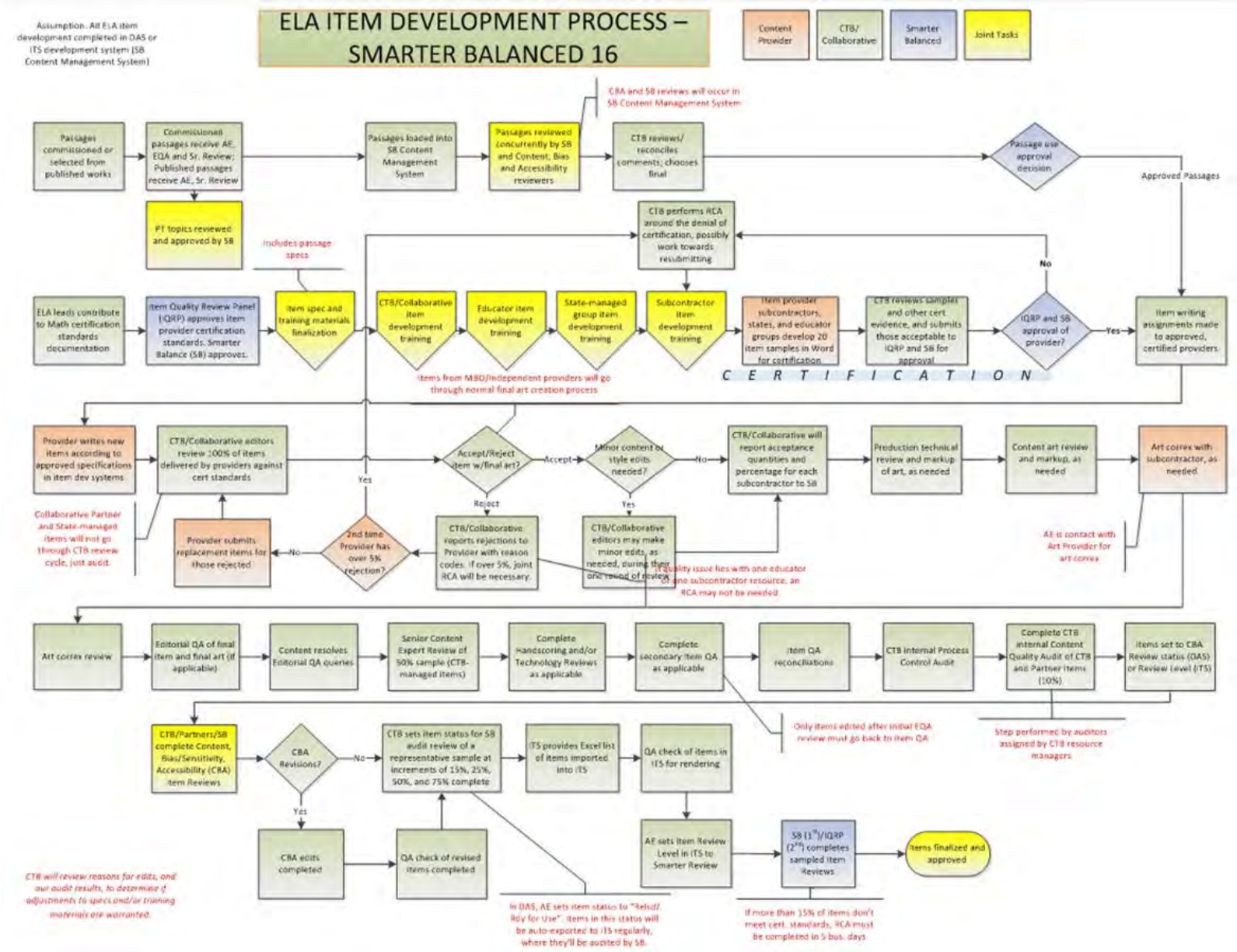
References

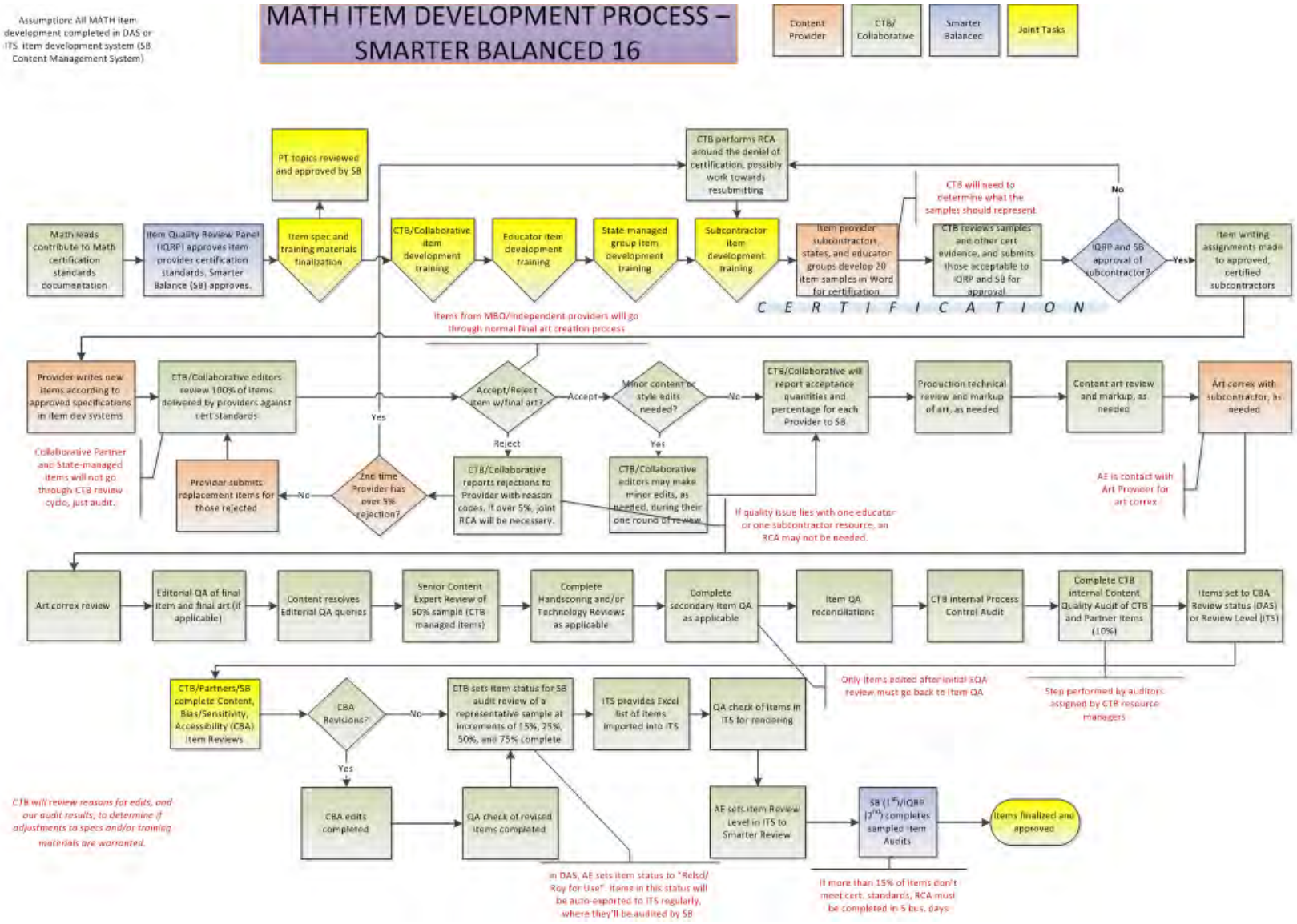
Smarter Balanced Assessment Consortium. (2014, November 14). *Interpretation and Use of Scores and Achievement Levels.* Los Angeles, CA: Author. Retrieved from https://portal.smarterbalanced.org/library/en/v1.0/interpretation-and-use-of-scores-and-achievement-levels.pdf

Smarter Balanced Assessment Consortium. (2016, March 29). *Reporting System User Guide.* Los Angeles, CA: Author. Retrieved from http://www.smarterapp.org/documents/Reporting-UserGuide.pdf

Appendix A: Item Development Process

The charts below outline the detailed process for stages of item development. They describe the many checks and reviews each item receives before it is approved for field testing. Item content, graphics, artwork, response processes and stimuli get extensive reviews. Items are also subject to reviews for possible cultural bias or material that may distract some test takers because it is in an area of sensitivity. Throughout the process there are checks to assure that items are accessible to as many students as possible.


ELA ITEM DEVELOPMENT PROCESS – SMARTER BALANCED 16
Assumption: All ELA item development completed in DAS or ITS development system (SB Content Management System)
Content Provider
CTB/ Collaborative
Smarter Balanced
Joint Tasks
CRA and SB reviews will occur in SB Content Management System
Passages commissioned or selected from published works
Commissioned passages receive AE, EQA and Sr. Review; Published passages receive AE, Sr. Review
Passages loaded into SB Content Management System
Passages reviewed concurrently by SB and Content, Bias and Accessibility reviewers
CTB reviews/ reconciles comments; chooses final
Passage use approval decision
Approved Passages
PT topics reviewed and approved by SB
Includes passage specs
CTB performs RCA around the denial of certification, possibly work towards resubmitting
No
ELA leads contribute to Math certification standards documentation
Item Quality Review Panel (IQRP) approves item provider certification standards. Smarter Balance (SB) approves.
Item spec and training materials finalization
CTB/Collaborative item development training
Educator item development training
State-managed group item development training
Subcontractor item development training
Item provider subcontractors, states, and educator groups develop 20 item samples in Word for certification
CTB reviews samples and other cert evidence, and submits those acceptable to IQRP and SB for approval
IQRP and SB approval of provider?
Yes
Item writing assignments made to approved, certified providers
C E R T I F I C A T I O N
Items from MBD/independent providers will go through normal final art creation process
Provider writes new items according to approved specifications in item dev systems
CTB/Collaborative editors review 100% of items delivered by providers against cert standards
Accept/Reject item w/final art?
Accept
Minor content or style edits needed?
No
CTB/Collaborative will report acceptance quantities and percentage for each subcontractor to SB
Production technical review and markup of art, as needed
Content art review and markup, as needed
Art correx with subcontractor, as needed
Yes
Reject
Collaborative Partner and State-managed items will not go through CTB review cycle, just audit
Provider submits replacement items for those rejected
No
2nd time Provider has over 5% rejection?
CTB/Collaborative reports rejections to Provider with reason codes. If over 5%, joint RCA will be necessary.
CTB/Collaborative editors may make minor edits, as needed, during their one round of review
If quality issue lies with one educator or one subcontractor resource, an RCA may not be needed.
AE is contact with Art Provider for art correx
Art correx review
Editorial QA of final item and final art (if applicable)
Content resolves Editorial QA queries
Senior Content Expert Review of 50% sample (CTB-managed items)
Complete Handscoring and/or Technology Reviews as applicable
Complete secondary item QA as applicable
Item QA reconciliations
CTB internal Process Control Audit
Complete CTB internal Content Quality Audit of CTB and Partner items (10%)
Items set to CRA Review status (DAS) or Review Level (ITS)
Only items edited after initial EQA review must go back to item QA
Step performed by auditors assigned by CTB resource managers
CTB/Partners/SB complete Content, Bias/Sensitivity, Accessibility (CRA) item Reviews
CRA Revisions?
No
CTB sets item status for SB audit review of a representative sample at increments of 15%, 25%, 50%, and 75% complete
ITS provides Excel list of items imported into ITS
QA check of items in ITS for rendering
Yes
CRA edits completed
QA check of revised items completed
AE sets Item Review Level in ITS to Smarter Review
SB (1st)/IQRP (2nd) completes sampled Item Reviews
Items finalized and approved
CTB will review reasons for edits, and our audit results, to determine if adjustments to specs and/or training materials are warranted.
In DAS, AE sets item status to "Reloc/ Rdy for Use". Items in this status will be auto-exported to ITS regularly, where they'll be audited by SB.
If more than 15% of items don't meet cert. standards, RCA must be completed in 5 bus. days


Assumption: All MATH item development completed in DAS or ITS item development system (SB Content Management System)
MATH ITEM DEVELOPMENT PROCESS – SMARTER BALANCED 16
Content Provider
CTB/ Collaborative
Smarter Balanced
Joint Tasks
PT topics reviewed and approved by SB
CTB performs RCA around the denial of certification, possibly work towards resubmitting
Math leads contribute to Math certification standards documentation
Item Quality Review Panel (IQRP) approves item provider certification standards. Smarter Balance (SB) approves.
Item spec and training materials finalization
CTB/Collaborative item development training
Educator item development training
State-managed group item development training
Subcontractor item development training
Item provider subcontractors, states, and educator groups develop 20 item samples in Word for certification
CTB will need to determine what the samples should represent
CTB reviews samples and other cert evidence, and submits those acceptable to IQRP and SB for approval
No
IQRP and SB approval of subcontractor?
Yes
Item writing assignments made to approved, certified subcontractors
C E R T I F I C A T I O N
Items from MBO/independent providers will go through normal final art creation process
Provider writes new items according to approved specifications in item dev systems
CTB/Collaborative editors review 100% of items delivered by providers against cert standards
Accept/Reject item w/final art?
Accept
Minor content or style edits needed?
No
Yes
Reject
CTB/Collaborative will report acceptance quantities and percentage for each Provider to SB
Production technical review and markup of art, as needed
Content art review and markup, as needed
Art correx with subcontractor, as needed
Collaborative Partner and State-managed items will not go through CTB review cycle, just audit.
Provider submits replacement items for those rejected
No
2nd time Provider has over 5% rejection?
CTB/Collaborative reports rejections to Provider with reason codes. If over 5%, joint RCA will be necessary.
CTB/Collaborative editors may make minor edits, as needed, during their one round of review
If quality issue lies with one educator or one subcontractor resource, an RCA may not be needed.
AE is contact with Art Provider for art correx
Art correx review
Editorial QA of final item and final art (if applicable)
Content resolves Editorial QA queries
Senior Content Expert Review of 50% sample (CTB-managed items)
Complete Handscoring and/or Technology Reviews as applicable
Complete secondary item QA as applicable
Item QA reconciliations
CTB internal Process Control Audit
Complete CTB internal Content Quality Audit of CTB and Partner items (10%)
Items set to CBA Review status (DAS) or Review Level (ITS)
Only items edited after initial EQA review must go back to Item QA
Step performed by auditors assigned by CTB resource managers
CTB/Partners/SB complete Content, Bias/Sensitivity, Accessibility (CBA) Item Reviews
CBA Revisions?
No
Yes
CTB sets item status for SB audit review of a representative sample at increments of 15%, 25%, 50%, and 75% complete
ITS provides Excel list of items imported into ITS
QA check of items in ITS for rendering
CBA edits completed
QA check of revised items completed
AE sets item Review Level in ITS to Smarter Review
SB (1st/IQRP 2nd) completes sampled Item Audits
Items finalized and approved
CTB will review reasons for edits, and our audit results, to determine if adjustments to specs and/or training materials are warranted.
In DAS, AE sets item status to "Relsd/Rdy for Use". Items in this status will be auto-exported to ITS regularly, where they'll be audited by SB
If more than 15% of items don't meet cert. standards, RCA must be completed in 5 bus. days

Detailed information about item writing, development, review and scoring can be obtained upon request. These documents are in the process of publication.

Topic	*Sub-topic*	*Document Name*
Item Writing	*Process Flow*	*20150512 Item Development Process Description FINAL*
		20150512 Smarter process maps FINAL
		Smarter 16 ITS Final Content Approval checklist FINAL
		Smarter 16 Final Web Approval Checklist20150512
	Models-Specifications	*20131003 Smarter 16 Item pool specification v12a Math FINALnew*
		20131006 Smarter 16 Item pool specification v12d ELA FINALnew
		ELA Archetypes
		Math_Archetype_Metadata
	Review criteria	*SB_16_ELA_Quality_Criteria_FINAL*
		SB_16_MATH_Quality_Criteria_FINAL
		CBA Item Review Business Rules 9-25
Human Scoring	*Process Description*	*20150512 Smarter Hand Scoring Process FINAL*
	Qualifications	*20150512 Smarter Hand Scoring Rater Qualifications FINAL*
	Quality Monitoring	*20150512 Smarter Hand Scoring Quality Monitoring FINAL*
	Recruitment-Training	*0150512 Smarter Hand Scoring Rater Training FINAL*
Data Review		*20150512 Smarter 2014 Field Test Data Review Summary Report FINAL*
		20150512 Smarter Data Review Results Summary

Appendix B: Test Design Development Activity and Outcomes

Major types of assessment design specifications that did not necessarily occur sequentially are summarized below that fall generally under the rubric of test design. These steps primarily relate to content validity of the Smarter Balanced assessments, particularly with respect to nonstandard administrations. Other test specifications concern the establishment of achievement levels and psychometric specifications that pertain to scaling and implications for scores. In many cases, the results were reviewed by one or more Stakeholder groups.

1) Conducted Initial Analysis of the Content and Structure of the CCSS

   An initial analysis of how each standard within the CCSS could be assessed in terms of item/task type and DOK was conducted. This was intended to support content and curriculum specialists and test- and item/task-development experts. Analysis and recommendations were made for all ELA/literacy and mathematics standards in grades 3 to 8 and high school. Multiple levels of review were conducted that included the Smarter Balanced Technical Advisory Committee, Smarter Balanced member states, and Smarter Balanced Executive Committee.

2) Developed Content Specifications for ELA/literacy and Mathematics

   Content specifications (e.g., claims, inferences, and evidence), item/task development criteria, and sample item/task sets were developed. This was intended to support the development of test blueprints and test specifications. Key constructs underlying each content area and critical standards/strands were identified in terms of demonstrating evidence of learning. Standards and bundled standards based on "bigger ideas" within the CCSS that require measurement through non-selected-response items (e.g., innovative item types) were identified. Reviews were conducted by CCSS authors, content experts, and assessment specialists.

3) Specified Accessibility and Accommodations Policy Guidelines

   Guidelines that describe the accessibility and accommodations framework and related policies for test participation and administration were created that incorporated evidence-based design (ECD) principles and outcomes from small-scale trials. State survey and review of best practices were reviewed as well as recommendations on the use of assessment technology. Input was solicited from the Smarter Balanced English Language Learners Advisory Committee and the Students with Disabilities Advisory Committee.

4) Developed Item and Task Specifications

   Smarter Balanced item/task type characteristics were defined as sufficient to ensure that content measured the intent of the CCSS and there was consistency across item/task writers and editors. This included all item types, such as selected-response, constructed-response, technology-enhanced, and performance tasks. In addition, passage/stimulus specifications (e.g., length, complexity, genre) and scoring rubric specifications for each item/task type were included. Specifications for developing items for special forms (e.g., braille) were also included.

5) Developed and Refined Test Specifications and Blueprints

   The test form components (e.g., number of items/tasks, breadth and depth of content coverage) necessary to consistently build valid and reliable test forms that reflect emphasized CCSS content were defined. These specifications included purpose, use, and validity claims of each test, item/task, test form, and CAT attribute. These were reviewed and revised based on CAT simulation studies, small-scale trials, Pilot and Field testing, and as other information was made available.

6) Developed Initial Achievement Levels

   Achievement expectations for mathematics and ELA/literacy were written in a manner that students, educators, and parents could understand. Panelists were recruited, and panels consisting of Institutes of Higher Education and a Cross-Consortia Technical Advisory Committee were convened in order to define college and career readiness. A period for public comment and various levels of review was implemented by the Smarter Balanced Technical Advisory Committee and selected focus groups with the approval of Governing Members. These activities were coordinated with the PARCC consortium.

7) Developed Item and Task Prototypes

   Prototype items and tasks using accessibility and Universal Design principles were produced that maximize fairness and minimize bias by using the principles of evidence-based design. Recommendations were made on how best to measure standards for innovative item types (per content specifications). This included prototypes for scoring guides, selected-response items, constructed-response items, and performance tasks. These prototypes were annotated, describing key features of items/tasks and scoring guides, passage/stimulus specifications (e.g., length, complexity, genre), and scoring rubric guidelines for each item/task type. Reviews, feedback, and revisions were obtained from educator-focus groups and Stakeholders, Smarter Balanced work groups, the Smarter Balanced English Language Learners Advisory Committee, and the Students with Disabilities Advisory Committee.

8) Wrote Item and Performance Task Style Guide

   The style guide specifies item/task formatting sufficient to ensure consistency of item/task formatting and display. The style guide specified the font, treatment of emphasized language/words (e.g., bold, italics), screen-display specifications, constraints on image size, resolution, colors, and passage/stimulus display configuration. Comprehensive guidelines for online and paper style requirements for all item types (e.g., selected-response, constructed-response, technology-enhanced, performance tasks) were specified.

9) Developed Accessibility Guidelines for Item and Task Development

   Guidelines were produced for item and task writing/editing that ensure accessibility of test content that addressed all item types. Interoperability standards at the item and test level were determined. Reviews, feedback, and revisions were based on educator-focus groups, Smarter Balanced work groups, the Smarter Balanced English Language Learners Advisory Committee, and the Students with Disabilities Advisory Committee.

10) Developed and Distributed Item/Task Writing Training Materials

   Training materials were created that specified consistent use of item/task specifications, style guides, accessibility guidelines, and best practices in item/task development (e.g., Universal Design, bias and sensitivity concerns) that were sufficient to ensure valid and reliable items/tasks that are free from bias and maximize accessibility to content. Training for item/task writing and editing was developed as online modules that enabled writers and editors to receive training remotely. Item writer and editor qualifications were established, and quality control procedures to ensure item writers were adequately trained were implemented.

11) Reviewed State-Submitted Items and Tasks for Inclusion in Smarter Balanced Item Pool

State-submitted items/tasks were reviewed for inclusion in the Pilot and/or Field Test item bank using the item bank/authoring system. This consisted of developing protocols for the submission and collection of state-submitted items/tasks for potential use in Pilot or Field Tests. These items were reviewed for item/task alignment, appropriateness (including access), and bias and sensitivity. Feedback was provided to states on the disposition of submitted items/tasks, and a gap analysis was conducted to determine the item/task procurement needs.

12) Planned and Conducted Small-Scale Trials of New Item and Task Types

Small-scale trials of new item/task types were used to inform potential revision of item/task specifications and style guides. Cognitive labs were conducted for new item/task types. Small-scale trials reflected an iterative development process, such that recommended revisions were evaluated as improvements became available.

13) Developed Automated-Scoring Approaches

The initial automated scoring methodology (e.g., regression, rules-based, or hybrid) was based on information from the content specifications, item/task specifications, item/task prototypes, and response data from the small-scale item/task trials. Reports documenting analysis were created, and independent review of this information with recommendations was made. Consultation, review, and approval of recommendations by the Smarter Balanced Technical Advisory Committee were made.

14) Developed Smarter Balanced Item and Task Writing Participation Policies and Guidelines

Documentation of processes for Smarter Balanced member states and Stakeholders to be involved in Smarter Balanced item/task writing activities (e.g., content and bias/sensitivity, data review, Pilot Testing, Field Testing) was developed. Criteria for selecting committee members (e.g., regional representation, expertise, experience) were also made.

15) Developed Content and Bias/Sensitivity Pilot Item and Task Review Materials

Methods for consistent training for content- and bias-review committees and for meeting logistics guidelines were provided. Review committees were recruited consistent with Smarter Balanced assessment participation policies.

16) Conducted Content and Bias/Sensitivity Reviews of Passages and Stimuli

Feedback from educators and other Stakeholders regarding passage/stimulus accuracy, alignment, appropriateness, accessibility, conformance to passage/stimulus specifications and style guides, and potential bias and sensitivity concerns was obtained. Educator feedback was documented, and procedures for feedback-reconciliation review were made.

17) Conducted Content and Bias/Sensitivity Pilot and Field Item and Task Review Meetings

Feedback from educators and other Stakeholders regarding item/task accuracy, alignment, appropriateness, accessibility, conformance to item/task specifications and style guides, and potential bias and sensitivity concerns was obtained. Reviews included all aspects of items/tasks (stem, answer choices, art, scoring rubrics) and statistical characteristics.

18) Developed Translation Framework and Specifications Languages

Definitions of item/task translation activities that ensure consistent and valid translation processes consistent with Smarter Balanced policy were produced. Review and approval of this process by the ELL Advisory Committee was made.

19) Translated Pilot and Field Test Items and Tasks into Identified Languages

Items/tasks translated into the specified languages were edited in sufficient quantity to support both Pilot- and Field-testing and operational assessments. Items/tasks included a full array of Smarter Balanced item types (selected-response, constructed-response, technology-enhanced, performance tasks). Review for content and bias/sensitivity of item/tasks and passages/stimuli was conducted.

20) Developed Content and Bias/Sensitivity Field Test Item and Task Review Materials

Supporting materials that ensure consistent training for content- and bias-review committees and meeting logistics guidelines were developed.

21) Revised Field Test Items and Tasks Based on Content and Bias/Sensitivity Committee Feedback

Fully revised items/tasks were available to be included on Field Test forms. Review panels were identified and convened, and training of state-level staff to edit and improve items/tasks that included all aspects of items/tasks (e.g., art, scoring rubrics) was conducted.

22) Developed Translation Framework and Specifications Languages

Definitions of item/task translation activities that ensured consistent and valid translation processes consistent with Smarter Balanced policy were created and approved by the ELL Advisory Committee.

23) Translated Pilot and Field Test Items and Tasks into Identified Languages

Translated items/tasks written by vendors, teachers, or provided through state submissions were edited in sufficient quantity to support Pilot and Field Tests and operational assessment.

24) Developed Content and Bias/Sensitivity Field Test Item and Task Review Materials

Review materials that ensure consistent training for content- and bias-review committees and meeting logistics guidelines were created. Feedback from educators and other Stakeholders regarding item/task accuracy, alignment, appropriateness, accessibility, conformance to item/task specifications and style guides, and potential bias and sensitivity concerns was obtained.

25) Produced a Single Composite Score Based on the CAT and Performance Tasks

A dimensionality study was conducted to determine whether a single sale and composite score could be produced or if separate scales for the CAT and performance task components should be produced. Based on the Pilot Test, a dimensionality study was conducted and the results presented to the Smarter Balanced Technical Advisory Committee. A unidimensional model was chosen for the Smarter Balanced scales and tests.

26) Investigated Test Precision for the CAT Administrations

An investigation of targets was conducted for score precision in the case in which tests are constructed dynamically from a pool of items and a set of rules must be established for the adaptive algorithm. A number of supporting simulation studies were conducted. The findings were used to inform subsequent test design for the operational CAT that was presented to the Smarter Balanced Technical Advisory Committee.

27) Selected IRT Models for Scaling

Using the Pilot Test data, the characteristics of various IRT models for selected- and constructed-response items were compared. The results of this study were presented to the Validation and Psychometrics/Test Design Work Group and the Smarter Balanced Technical Advisory Committee for comment. The two-parameter logistic (2-PL) model for selected-response and the Generalized Partial Credit (GPC) Model for constructed-response were chosen as the scaling models.

Appendix C: Reporting Achievement Levels

Descriptors for Smarter Balanced achievement levels provided to Consortium members for the 2016-2017 test administration. Please note that members may choose to alter the descriptors or name the four achievement levels.

Enclosed are the Achievement Levels for the English language arts/literacy and mathematics Smarter Balanced assessments. Please note states may choose to alter the descriptions or name descriptors.

Mathematics Reporting Achievement Levels

High School	Grades 6–8	Grades 3–5
Level 4 The student has exceeded the achievement standard and demonstrates the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 4 The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 4 The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in mathematics needed for likely success in future coursework.
Level 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after completing high school coursework.	Level 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in mathematics needed for likely success in future coursework.
Level 2 The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 2 The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 2 The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in mathematics needed for likely success in future coursework.

High School	Grades 6–8	Grades 3–5
Level 1 The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 1 The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in mathematics needed for likely success in entry-level credit-bearing college coursework after high school.	Level 1 The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in mathematics needed for likely success in future coursework.

English language arts/literacy Reporting Achievement Levels

High School	Grades 6–8	Grades 3–5
Level 4 The student has exceeded the achievement standard and demonstrates the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	Level 4 The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	Level 4 The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in future coursework.
Level 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after completing high school coursework.	Level 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	Level 3 The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in future coursework.

High School	Grades 6–8	Grades 3–5
Level 2 The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	**Level 2** The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	**Level 2** The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in future coursework.
Level 1 The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	**Level 1** The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.	**Level 1** The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in future coursework.