

Smarter Balanced Assessment Consortium: 2014-15 Technical Report

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

June 23, 2016
Updated October 5, 2016



Table of Contents

Introduction and Overview.....	vi
Overview	vi
Technical Report Approach	vii
Peer Review Guidelines and Established Standards.....	viii
Overview and Background of the Smarter Balanced Theory of Action.....	ix
Six Principles of Smarter Balanced Underlying the Theory of Action.....	ix
Purposes for the Smarter Balanced Assessment System.....	xi
Overview of Report Chapters:.....	xii
Chapter 1: Validity.....	xiii
Chapter 2: Reliability/Precision and Errors of Measurement	xiii
Chapter 3: Test Fairness	xiii
Chapter 4: Test Design	xiii
Chapter: 5 Scores, Scales and Norms.....	xiv
Chapter 6: Test Administration	xiv
Chapter 7: Reporting and Interpretation	xiv
Acknowledgments.....	xv
Outside Groups and Organizations that Collaborated with the Smarter Balanced Assessment Consortium	xv
Contributors to the Accessibility Accommodations Framework.....	xv
Other Acknowledgments.....	xvi
References	xvii
Chapter 1 : Validity.....	1-1
Introduction	1-2
Purposes of the Smarter Balanced System for Summative Assessments.....	1-2
Summary of Validity Argument.....	1-3
Validity Framework for Summative Assessments.....	1-3
Essential Validity Evidence Derived from the <i>Standards</i>	1-13
The 2014 <i>Standards</i> ' Five Primary Sources of Validity Evidence.....	1-14
Conclusion for Summative Test Validity Results.....	1-16
References	1-17
Chapter 2 : Reliability, Precision and Errors of Measurement.....	2-1

Introduction	2-2
Simulations Studies for 2014-15 Operational Summative Tests.....	2-2
Tests for Special Populations.....	2-7
Item exposure.....	2-10
Blueprint fidelity	2-10
Observed Reliability	2-12
Paper/Pencil Test Reliability	2-19
Classification Accuracy	2-19
Standard Errors of Measurement	2-35
Online Platform Effects.....	2-41
References	2-45
Chapter 3 : Test Fairness	3-1
Introduction	3-2
Definitions for Validity, Bias, Sensitivity, and Fairness.....	3-3
The Smarter Balanced Accessibility and Accommodations Framework	3-4
Meeting the Needs of Traditionally Underrepresented Populations.....	3-7
How the Framework Meets Needs of Students Who Are ELLs.	3-7
How the Framework Meets Needs of Students with Disabilities.	3-7
The Individual Student Assessment Accessibility Profile (ISAAP).	3-8
Usability, Accessibility, and Accommodations Guidelines: Intended Audience and Recommended Applications.	3-9
Guidelines for Accessibility for English Language Learners.....	3-11
Provision of Specialized Tests or Pools	3-13
Fairness as a Lack of Measurement Bias: Differential Item Functioning Analyses.....	3-14
Differential Item Functioning (DIF) Analyses for the Calibration Item Pool.....	3-14
Test Fairness and Implications for Ongoing Research	3-21
References	3-22
Chapter 4 : Test Design	4-1
Introduction	4-2
A Brief Description of Smarter Balanced Content Structure.....	4-2
Synopsis of Assessment System Components	4-4
Evidence-Centered Design in Constructing Smarter Balanced Assessments.....	4-4
Test Blueprints	4-5

Operational Summative Assessment Blueprints and Specifications.....	4-7
CAT and Performance Task Test Components.....	4-11
Adaptive Test Design and Algorithm Overview.....	4-11
Item attributes.....	4-12
Item measurement data:.....	4-12
Test Operation Walkthrough.....	4-13
Item and Task Development	4-14
Item and Task Specifications	4-14
Performance Task Design	4-16
The Item/task Pool Specification.....	4-17
Item Writing	4-18
Item Reviews	4-19
Field Testing.....	4-20
Composition of Summative Item Pools.....	4-22
Content Alignment	4-31
Fordham/HumRRO study	4-32
Pool analysis and adequacy: Background and Recommendations.....	4-33
Summary of Test Design.....	4-37
References	4-38
Chapter 5 : Scores, Scales and Norms.....	5-1
Introduction	5-2
Item Response Theory	5-2
Calibration and Scaling.....	5-3
Vertical Scale.....	5-4
Transforming the Theta Metric to the Scale Score	5-4
Minimum and Maximum Scale Scores	5-5
Achievement-Level Setting.....	5-6
Results for the 2014-2015 Assessments	5-8
Overall Results	5-8
Claim-Level Results.....	5-16
Percentile Tables for Overall Scale Scores.....	5-42
Percentile Tables for Claim-level Scale Scores	5-42

Modes of Administration	5-75
References	5-83
Chapter 5 Addendum.....	5-84
Item Drift Analysis	5-84
References	5-99
Chapter 6 : Test Administration	6-1
Introduction	6-2
Test Administration	6-2
Session Time	6-2
Test Administration Manual	6-3
Clear Directions to Ensure Uniform Administration	6-4
Detailed Instructions for Test Takers.....	6-4
Responsibilities of Test Administrators	6-5
Tools, Designated Supports, and Accommodations	6-5
Item Exposure Rates.....	6-6
Blueprint Fidelity	6-9
Spring 2015 Embedded Field Test Results.....	6-9
Machine-Scored Items.....	6-10
Hand-Scored Items	6-16
Spring 2015 Embedded Field Test Results.....	6-17
References	6-18
Chapter 7 : Reporting and Interpretation	7-1
Introduction	7-2
Overall Test Scores	7-2
Sub-scores.....	7-3
Types of Reports	7-5
Individual Student Report (ISR).....	7-5
Lists.....	7-6
Aggregate Reports.....	7-7
Data Downloads.....	7-7
Summary	7-8
References	7-9

Introduction and Overview

Overview

The Smarter Balanced Assessment Consortium's (Smarter Balanced) Assessment System includes a set of balanced components that are designed to meet a diversity of students' needs across all of the Consortium's members. This system provides valid, reliable, and fair assessments of the deep disciplinary understanding and higher-order thinking skills increasingly demanded by a knowledge-based global economy. The system is based on the belief that assessment must support ongoing improvements in instruction and learning experiences for students that lead to outcomes valued by all stakeholders. Smarter Balanced supports the goals of its members who seek to ensure that all students leave high school prepared for postsecondary success in college or a career through a planned sequence of educational experiences and opportunities. The system was grounded in the strong foundational assessments, policies and procedures of its members including supports and resources from institutions of higher education (IHEs) and workplace representatives. The Consortium expanded on these proven successes to create a high quality, balanced, multistate assessment system based on the Common Core State Standards (CCSS) in English language arts/literacy (ELA/literacy) and mathematics.

The Consortium's staff provide expert guidance and facilitate member driven decisions regarding the maintenance and enhancement of the system as required to fulfill its mission to improve teaching and learning. Smarter Balanced members retain flexibility regarding how to customize the system so that it may best be used as part of their approach to improving their local educational systems. The Consortium's assessment system strategically uses of a variety of item types including performance tasks to measure the full range of the CCSS. The Consortium also deploys essential resources that are embedded in the test to ensure accurate assessment of all students, including students with disabilities, English language learners, and low- and high-performing students. Smarter Balanced implemented a system that features

- assessment of CCSS using secure adaptive summative assessments that incorporate items that are deliberately designed to measure specific content. The assessments include a variety of item types including technology-enhanced items, items that require constructed response and performance tasks.
- interim assessments that incorporate items that are developed according to the same processes as the items incorporated in the summative assessment. The interim assessments are not secure and provide more flexible administration options to assist educators in determining what students know and can do in relation to the CCSS.
- a digital library that includes research-supported tools, processes, and practices developed by experts that support the formative process teachers can use to improve their professional practices.
- open sourced technology that members can use to delivery assessments and report results to educators.

- cross-member state communications to inform stakeholders about Smarter Balanced activities and to ensure a common focus on the goal of college- and career-readiness for all students.

The innovative and efficient use of technology serves as a central feature of this balanced assessment system. Some central notions concerning technology use are that

1. the Smarter Balanced system uses computer adaptive testing to increase the precision and efficiency of the summative tests,
2. the expanded use of technology enables the development of innovative and realistic item types that measure student achievement across a wide performance continuum providing opportunities for educator and administrator professional development and local capacity building, and
3. through the use of an interoperable electronic platform and leveraging of cross-member state resources, Smarter Balanced delivers assessments and produces standardized reports that are cost effective, timely, and useful for a range of audiences in tracking and analyzing student progress toward college- and career-readiness at the individual student, student subgroup, classroom, school, district, and state levels.

In summary, the Smarter Balanced learning and assessment system is grounded in a sound theory of action. This system promotes research-supported classroom practice and incorporates a balanced set of technology-enabled tools, innovative assessments, and classroom support materials intended to work coherently to facilitate teaching and learning.

Technical Report Approach

The intent of this report is to provide comprehensive and detailed evidence in support of the validity of Smarter Balanced assessment program. This report focuses on summative tests and will be supplemented with information about interim tests. Information about the overall system is provided to provide context. At the outset, it should be recognized that demonstration of validity is an ongoing process. Validity evidence provided here is from the first year of operational testing, referencing initial pilot and field test phases as needed.

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 the member, 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 necessary validity 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 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 that lead to valid inferences about student performance on the standards, both individually and aggregated,
- scores that reflect the full range of achievement standards,
- documentation that describes how the assessments consist of a coherent system across grades and subjects including studies establishing vertical scales, and
- identification 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 prepared for postsecondary success in college or a career through increased student learning and improved teaching. This approach suggests that enhanced learning will result from high-quality assessments that support ongoing improvements in instruction and learning. 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

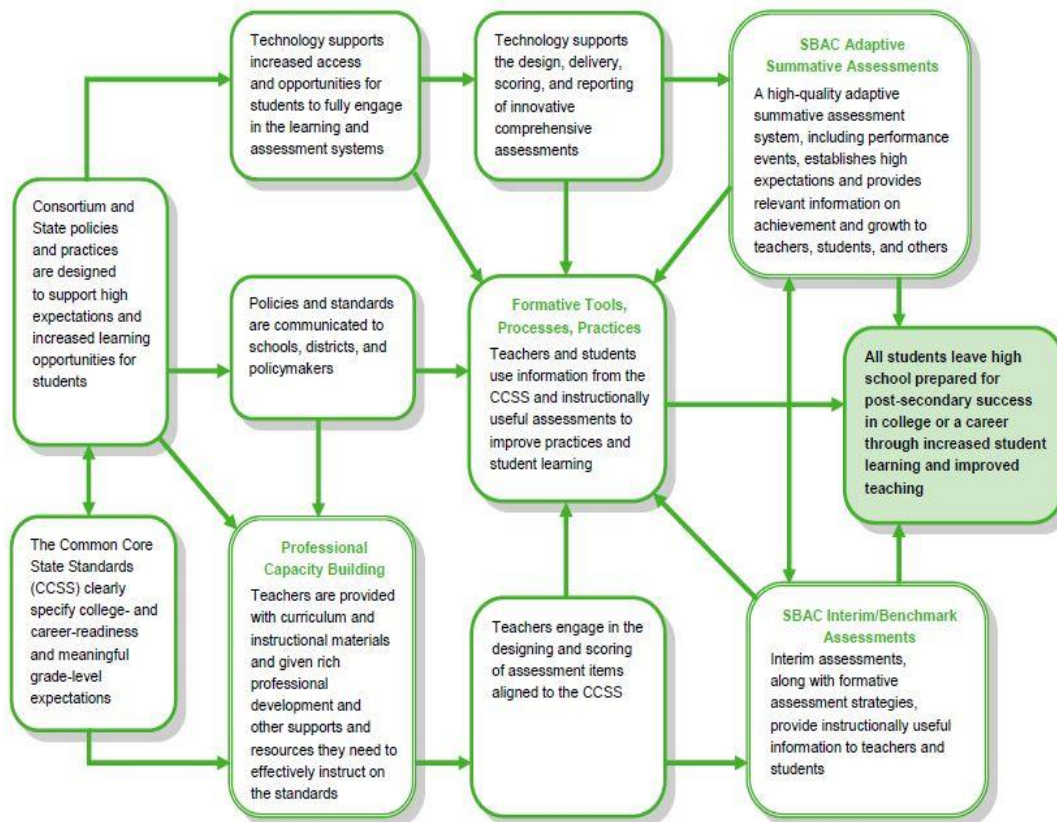
The Smarter Balanced assessment is guided by a set of six principles shared by systems in high-achieving nations and a number of high-achieving states in the U.S.

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 and interim/benchmark assessments and associated support tools 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 evaluate 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 US DOE Peer Review, relying heavily on the Standards as its core resource for quality design.

FIGURE 1. OVERVIEW OF SMARTER BALANCED THEORY OF ACTION



The primary rationale of the Smarter Balanced assessments is that these aspects can interact to improve the intended student outcomes (i.e., college- and career-readiness). Connection among these assessment components is presented in Figure 1.

Purposes for the Smarter Balanced Assessment System

The Smarter Balanced purpose statements refer to three categories: (a) summative assessments, (b) interim assessments, and (c) formative assessment resources.

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 those 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.

This report provides technical information about the summative assessments, but the purposes of interim assessments and formative resources are given here to provide context for summative assessments as a component of the assessment system.

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 measurement tools and resources to

- improve teaching and learning,
- provide resources to teachers to help them 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 the chapters contained 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.

Description: This chapter provides information in answer to the following questions. 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?

Content: Chapter 1 provides the statement of test purposes; valid score uses and interpretations; outline of validity evidence in the rest of the report.

Chapter 2: Reliability/Precision and Errors of Measurement

Information about simulated and operational performance of the test in delivering scores is provided in this chapter. 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.

Description: How do we know that scores are accurate? What kind of instructional decisions does precision support? Are scores accurate enough to evaluate change over time in growth models? How do we know they are stable and have the same meaning for all students?

Content: Chapter 2 presents simulated and operational conditional standard errors of measurement, marginal and fixed-form reliability, test information curve; observed decision consistency and accuracy; simulation results for bias and theta recovery.

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.

Description: How do we know that the test is fair to all students? How was fairness guaranteed in developing test questions and tasks? How is the test administered so that each student can demonstrate their skills?

Content: Chapter 3 presents the Smarter Balanced Conceptual Framework for Usability, Accessibility, and Accommodations, bias and sensitivity reviews conducted during item and task development and differential item functioning (DIF) analysis.

Chapter 4: Test Design

This chapter provides information pertaining to the content validity of the Smarter Balanced assessment system. It describes the how tasks and items are structured to achieve domain coverage. For Smarter Balanced tests, test design includes the relationship of claims and targets to the underlying CCSS and how adaptive and performance 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.

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

Content: Chapter 4 provides evidence that the CCSS address skills required for college and career readiness. 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.

Chapter: 5 Scores, Scales and Norms

This chapter describes the steps to adopting a psychometric model and how Smarter Balanced scales were constructed. It also covers the achievement level setting process and resulting cut scores. Normative information about the Consortium population is also included.

Description: What do the scores mean? How can we know that descriptions of achievement levels match criteria for those levels? Are criteria stable so teachers can work toward a fixed goal? How does a student's score compare to expected criteria; to those of his or her peers?

Content: Chapter 5 summarizes how scales were established in pilot and field test stages. It describes how cut scores were developed from foundational achievement level descriptors 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.

Chapter 6: Test Administration

Part of test validity rests on the assumption that assessments are given 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. This chapter includes analysis of field test items presented during test administration.

Description: 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?

Content: 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.

Chapter 7: Reporting and Interpretation

Examples of Smarter Balanced reports are shown here, along with an explanation of report elements. This chapter discusses intended uses of report information.

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

Content: Chapter 7 provides examples of the Smarter Balanced suite of reports and interpretive information.

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.

2014-15 Technical Advisory Committee.

- Jamal Abedi, Ph.D. *UC Davis/CRESST*
- Randy Bennett, Ph.D. *ETS*
- Derek C. Briggs, Ph.D. *University of Colorado*
- Gregory J. Cizek, Ph.D. *University of North Carolina*
- David T. Conley, Ph.D. *University of Oregon*
- Linda Darling-Hammond, Ph.D. *Stanford University*
- Brian Gong, Ph.D. *The Center for Assessment*
- Edward Haertel, Ph.D. *Stanford University*
- Joan Herman, Ph.D. *UCLA/CRESST*
- G. Gage Kingsbury, Ph.D. *Psychometric Consultant*
- James W. Pellegrino, Ph.D. *University of Illinois, Chicago*
- W. James Popham, Ph.D. *UCLA, Emeritus*
- Joseph Ryan, Ph.D. *Arizona State University*
- Martha Thurlow, Ph.D. *University of Minnesota/NCEO*

Contributors to the Accessibility Accommodations Framework.

In February 2012, the Smarter Balanced Assessment Consortium Accessibility and Accommodations Work Group began work on developing the Accessibility and Accommodations Framework. The primary goal of this effort was to develop uniform accessibility and accommodation policies and guidelines that will be adopted and used by all Smarter Balanced members. Recognizing the diversity in policies and practices that currently exist across member states, the legal issues that must be addressed by the policies, the mixed research findings regarding many accommodation practices, and the differences in opinion regarding accommodation policies, the work group undertook an iterative process designed to gather input from a large and diverse audience. This effort began by contracting with Measured Progress and its partners, who included:

- Members of the Measured Progress Innovation Lab who conducted work in accessibility in digital environments, developed the Accessible Test Design model, and were leaders in developing the Accessible Portable Item Protocol (APIP) Standard,
- Experts at Educational Testing Service who have conducted a variety of studies on test accommodations and accessibility for students with disabilities and for students who are English language learners, and who have developed industry-recognized guidelines for accessibility in the context of assessment,
- Experts at the George Washington University Center for Equity and Excellence in Education, who are nationally recognized experts in accessible assessment for students who are English language learners and who have worked with several states to develop policies on test accommodations for students who are English language learners, and
- Experts affiliated with the National Center on Educational Outcomes who have conducted extensive reviews of state test accommodation policies, worked with the Assessing Special Education Students (ASES) work group of the Council of Chief State School Officers (CCSSO) to develop test accommodation policies, and closely monitored research on test accommodations.

In addition to these partners, an expert panel was formed composed of the following members:

- Jamal Abedi assessment of English language learners, *UC Davis/CRESST*,
- Martha Thurlow assessment of students with disabilities, University of Minnesota/NCEO,
- Sheryl Lazarus test accommodations for students with disabilities, University of Minnesota/NCEO,
- Stephanie Cawthon accommodations for students who communicate in American Sign Language, University of Texas at Austin,
- Richard Jackson accommodations for students with visual impairments, Boston College,
- Rebecca Kopriva assessment of students who are English language learners, Wisconsin Center for Education Research, and
- Stephen Sireci validity of test accommodations, University of Massachusetts Amherst/CEA.

Other Acknowledgments.

This technical report leveraged the *Smarter Balanced Comprehensive Research Agenda* by Stephen G. Sireci (2012) as the primary validity framework and sources of evidence. Input was provided on critical aspects of the program and this report by the Smarter Balanced 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 <http://www.k12center.org/publications.html>.
- 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. This validation process does not rely on a single study or gathering 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 development, field-testing and analyses of items, test scaling and linking, scoring, and reporting. This chapter provides an evaluative framework for the validation of the Smarter Balanced summative assessment. It points the reader to supporting evidence in other parts of the technical report and other sources that seek to demonstrate that the Smarter Balanced Assessment System adheres to guidelines for fair and high quality assessment.

The validity argument begins with a statement of summative assessment intended purposes, followed by the evidentiary framework supporting the validity argument. 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*, and the *Smarter Balanced Assessment Consortium: Comprehensive Research Agenda* (Sireci, 2012).

The *Standards* are considered to be “the most authoritative statement of professional consensus regarding the development and evaluation of educational and psychological tests” (Linn, 2006, p. 27) currently available. The 2014 *Standards* differ from earlier versions in the emphasis given to the increased prominence of technology in testing, including computer adaptive testing (CAT). Dr. Sireci based his research agenda work on the *Standards* and his work in operational interpretation of validity argumentation (Sireci, 2013).

Purposes of the Smarter Balanced System for Summative Assessments

To derive the statements of purpose listed below, panels consisting of Smarter Balanced leadership, including the Executive Director, Smarter Balanced staff, Dr. Stephen Sireci and key personnel from Consortium states were convened.

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

7. Students' ELA/literacy and mathematics achievement with respect to those CCSS measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school.
8. Whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college readiness.

9. 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.
10. Students' annual progress toward college and career readiness in ELA/literacy and mathematics.
11. How instruction can be improved at the classroom, school, district, and state levels.
12. Students' ELA/literacy and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems.
13. Students' achievement in ELA/literacy and mathematics that is equitable for all students and subgroups of students.

Summary of Validity Argument

The crux of the argument presented here is that the technical quality of the summative assessments supports these purposes. The Common Core State Standards (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. Most of each content area test is delivered adaptively so that blueprint requirements are met, but scores are more accurate 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. The vertical scale was constructed to provide measurement across grades, facilitating estimates of progress toward career and 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 descriptors and continued through a rigorous process of setting achievement criteria. These processes involved participants from member higher education systems to ensure that readiness criteria represented skills needed for success in first year college courses.

This is a high-level view of the validity argument. A detailed description of evidence appears below.

Validity Framework for Summative Assessments

The *Standards* describe a process of validation that consists of developing a sufficiently convincing argument, based on empirical evidence, that the interpretations and actions based on test scores are sound. Kane (1992, 2006) characterized this process as a validity argument, which is consistent with the validation process described by the 2014 *Standards*.

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 validity framework corresponds to validity evidence sources described in the *Standards* (AERA et al. 2014, pp. 26-31). They are also the organizing principles for the Smarter Balanced research framework document (Sireci, 2012). These five sources of validity evidence consist of:

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 for Validity and Consequences of Testing.

This Technical Report provides part of the evidence for this framework. As many observers have noted, 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. It is comprised of members who deliver the test, score operational items and provide reports. Members use Smarter Balanced test scores in their own accountability models. In this report, guidelines for administration and use are documented. For complete validity evidence, member documentation on specific test administration procedures, reporting, and use should be consulted.

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.

In many cases, validity evidence will come from an outside auditor, such as the federal peer reviewer process or from an external study. Table 1.1 shows the components of validity covered in this report, other sources, and future studies.

TABLE 1.1 SOURCES OF VALIDITY ELEMENTS

Purpose	Source of Validity Evidence for Summative Assessments			
	Test Content	Internal Structure	Relations to Other Variables	Response Processes
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.	✓	✓		✓

Table 1.2 through Table 1.5 provide details on sources of validity evidence, including the location of the evidence in this report and from other sources.

TABLE 1.2 SOURCES OF VALIDITY EVIDENCE BASED ON TEST CONTENT

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
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.	<ul style="list-style-type: none"> 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. 	3, 4	<ul style="list-style-type: none"> <i>Evaluating the Content and Quality of Next Generation Assessments</i> (Doorey & Polikoff, 2016) <i>Evaluating the Content and Quality of Next Generation High School Assessments</i> (Schultz, Michaels, Dvorak, & Wiley, 2016)
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.	<ul style="list-style-type: none"> 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. 	4	<ul style="list-style-type: none"> <i>Development Process</i> (NGA Center & CCSSO, 2016) <i>Evaluating the Content and Quality of Next Generation Assessments</i> (Doorey & Polikoff, 2016)
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.	<ul style="list-style-type: none"> CCSS are based on CCR. Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. 	4	<ul style="list-style-type: none"> <i>Development Process</i> (NGA Center & CCSSO, 2016) <i>Evaluating the Content and Quality of Next Generation High School Assessments</i> (Schultz, Michaels, Dvorak, & Wiley, 2016)
4. Measure students' annual progress toward college and career readiness in ELA/literacy and mathematics.	<ul style="list-style-type: none"> 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. 	4	<ul style="list-style-type: none"> <i>Development Process</i> (NGA Center & CCSSO, 2016) <i>Evaluating the Content and Quality of Next Generation Assessments</i> (Doorey & Polikoff, 2016) <i>Evaluating the Content and Quality of Next Generation High School Assessments</i> (Schultz, Michaels, Dvorak, & Wiley, 2016)

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
5. Inform how instruction can be improved at the classroom, school, district, and state levels.	<ul style="list-style-type: none"> • 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. 	4, 7	<ul style="list-style-type: none"> • <i>Evaluating the Content and Quality of Next Generation Assessments</i> (Doorey & Polikoff, 2016) • <i>Evaluating the Content and Quality of Next Generation High School Assessments</i> (Schultz, Michaels, Dvorak, & Wiley, 2016) • <i>End of Grant Report</i> (Smarter Balanced, 2015, p. 28)
6. Report students' ELA/literacy and mathematics proficiency for federal accountability purposes and potentially for state and local accountability systems.	<ul style="list-style-type: none"> • 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. 	5, 6, 7	
7. Assess students' achievement in ELA/literacy and mathematics in a manner that is equitable for all students and subgroups of students.	<ul style="list-style-type: none"> • Bias is minimized through Universal Design and accessibility resources. • Assessments are administered in a standardized manner sufficient to yield data that supports valid inferences. 	3, 4, 6	

TABLE 1.3 SOURCES OF VALIDITY EVIDENCE BASED ON INTERNAL STRUCTURE

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
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.	<ul style="list-style-type: none"> The assessment supports precise measurement and consistent classification. Achievement levels were set consistent with best practice. 	2, 5	
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.	<ul style="list-style-type: none"> Scale is vertically articulated Achievement levels are vertically articulated 	5	2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 6, 9, 10)
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.	<ul style="list-style-type: none"> Scale is vertically articulated. Achievement levels are vertically articulated. 	5	2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 6, 9, 10)
4. Measure students' annual progress toward college and career readiness in ELA/literacy and mathematics.	<ul style="list-style-type: none"> 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. 	2, 5	2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 6, 9, 10)
5. Inform how instruction can be improved at the classroom, school, district, and state levels.	<ul style="list-style-type: none"> Threshold, Range and policy Achievement Level Descriptors 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. 	4, 5, 7	

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
6. Report students' ELA/literacy and mathematics proficiency for federal accountability purposes and potentially for state and local accountability systems.	<ul style="list-style-type: none"> 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. 	2, 5, 7	2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 10)
7. Assess students' achievement in ELA/literacy and mathematics in a manner that is equitable for all students and subgroups of students.	<ul style="list-style-type: none"> 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. 	2, 3, 6	2013-2015 Technical Manual (Smarter Balanced, 2016, Ch. 10)

TABLE 1.4 SOURCES OF VALIDITY EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
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.	<ul style="list-style-type: none"> Achievement levels are consistent with other measures 	5	<ul style="list-style-type: none"> <i>Study of the Relationship Between the Early Assessment Program and the Smarter balanced Field Tests</i> (ETS, 2015) <i>Linking Course Grades to Smarter Balanced Cut Scores</i> (OSPI, 2016)
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.	<ul style="list-style-type: none"> Achievement levels are consistent with other measures. 	5	
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.	<ul style="list-style-type: none"> Achievement levels are consistent with other measures. 		<ul style="list-style-type: none"> <i>Study of the Relationship Between the Early Assessment Program and the Smarter balanced Field Tests</i> (ETS, 2015)
4. Measure students' annual progress toward college and career readiness in ELA/literacy and mathematics.	Will be addressed in future studies of annual observed growth.		
5. Inform how instruction can be improved at the classroom, school, district, and state levels.	Will be addressed in future studies of instructional change in response to assessment results.		
6. Report students' ELA/literacy and mathematics proficiency for federal accountability purposes and potentially for state and local accountability systems.	N/A		
7. Assess students' achievement in ELA/literacy and mathematics in a manner that is equitable for all students and subgroups of students.	N/A		

TABLE 1.5 SOURCES OF VALIDITY EVIDENCE BASED ON RESPONSE PROCESSES

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
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.	<ul style="list-style-type: none"> 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 provides confirmation of content measurement. 	3, 4, 5	<i>Cognitive Laboratories Technical Report (AIR, 2013)</i>
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.	<ul style="list-style-type: none"> 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 provides confirmation of content measurement. 	4, 5	<ul style="list-style-type: none"> <i>Cognitive Laboratories Technical Report (AIR, 2013)</i>
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.	<ul style="list-style-type: none"> 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 provides confirmation of content measurement. 	4, 5	<ul style="list-style-type: none"> <i>Cognitive Laboratories Technical Report (AIR, 2013)</i>

Purpose	Summary of Evidence	Chapters	Other Sources Evidence
4. Measure students' annual progress toward college and career readiness in ELA/literacy and mathematics.	<ul style="list-style-type: none"> 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 provides confirmation of content measurement. 	4, 5	<ul style="list-style-type: none"> <i>Cognitive Laboratories Technical Report (AIR, 2013)</i>
5. Inform how instruction can be improved at the classroom, school, district, and state levels.	<ul style="list-style-type: none"> Test blueprint, content specifications, and item specifications are aligned to grade level content, process skills, and associated cognitive complexity. Threshold, Range and policy Achievement Level Descriptors were developed in consultation with educators, with the goal of providing information to educators. 		
6. Report students' ELA/literacy and mathematics proficiency for federal accountability purposes and potentially for state and local accountability systems.	<ul style="list-style-type: none"> 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 provides confirmation of content measurement. 	5, 7	<i>Cognitive Laboratories Technical Report (AIR, 2013)</i>
7. Assess students' achievement in ELA/literacy and mathematics in a manner that is equitable for all students and subgroups of students.	<ul style="list-style-type: none"> 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 provides confirmation of content measurement. 	3, 4, 6	<i>Cognitive Laboratories Technical Report (AIR, 2013)</i>

Essential Validity Evidence Derived from the *Standards*

The *Standards* (AERA et al. 2014, p.22) also present a set of essential validity elements consistent with evidence typically reported for large-scale educational assessment programs. The essential validity elements present a traditional synopsis of validity evidence, which form the basis for the evidence demonstrated for the Smarter Balanced initial operational administration.

The *Standards* describe these essential validity elements as

- A. evidence of careful test construction;
- B. adequate score reliability;
- C. appropriate test administration and scoring;
- D. accurate score scaling, equating, and standard setting; and
- E. attention to fairness, equitable participation and access.

Table 1.6 presents a brief description of the essential validity evidence. Many of these essential validity elements fall under the validity evidence based on test content (e.g., careful test construction) and internal structure (adequate score reliability, scaling, equating). The sources of evidence listed in Table 1.1 will reemerge when considering the five specific validity elements, which represent the full validity framework. This overlap underscores the fundamental nature of these elements for supporting the use of Smarter Balanced assessments for their intended purposes. Table 1.6 is followed by a brief description of the potential types of evidence associated with each essential element.

TABLE 1.6 SYNOPSIS OF ESSENTIAL VALIDITY EVIDENCE DERIVED FROM STANDARDS (AERA ET AL., 2014, P. 22)

Essential Element	Chapter	Type of Associated Validation Evidence
Careful Test Construction	4. Test Design	Description of test development steps, including construct definition (test specifications and blueprints), item writing and review, item data analysis, alignment studies
Adequate Measurement Precision (Reliability)	2. Reliability, Precision & Error	Analysis of test information, conditional standard errors of measurement, decision accuracy and consistency, and reliability estimates.
Appropriate Test Administration	6. Test Administration	Test administration procedures, including protocols for test irregularities; availability and assignment of test accommodations. Test, item and data security.
Appropriate Item Scoring	6. Test Administration	Scoring procedures, rater agreement analyses.
Accurate Scaling and Equating	5. Scales, Scores, and Norms	Documentation of test design, IRT model choice, scaling and equating procedures, IRT residuals, validating vertical scaling assumptions.

Essential Element	Chapter	Type of Associated Validation Evidence
Appropriate Standard Setting	5. Scales, Scores, and Norms	Comprehensive standard-setting documentation provided, including procedural, internal, and external validity evidence for all achievement-level standards.
Attention to Fairness, Equitable Participation and Access	3. Test Fairness	Accommodation policy guidelines, implementation of accommodations, sensitivity review, DIF analyses, analyses of accommodated tests; analysis of participation rates, availability of translations.

The 2014 Standards' Five Primary Sources of Validity Evidence

The five sources of validity evidence serve as organizing principles and represent a comprehensive framework for evaluating validity for Smarter Balanced. 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. These five sources of validity evidence consist of (1) test content, (2) response processes, (3) internal structure, (4) relations to other variables, and (5) consequences of testing. They are briefly described below:

6. 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). The degree to which (a) the Smarter Balanced test specifications captured the Common Core State Standards and (b) the items adequately represent the domains delineated in the test specifications, were demonstrated in the alignment studies. The major assumption here is that the knowledge, skills, and abilities measured by the Smarter Balanced assessments are consistent with the ones specified in the Common Core State Standards. Administration and scoring can be considered as aspects of content-based evidence. With computer adaptive testing, an extra dimension of test content is to ensure that the tests administered to students conform to the test blueprint.
7. Validity evidence based on *response processes* refers to “evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by examinees” (AERA et al., 1999 p. 12). This evidence might include documentation of such activities as
 - interviewing students concerning their responses to test items (i.e., speak alouds);
 - systematic observations of test response behavior;
 - evaluation of the criteria used by judges when scoring performance tasks, analysis of student item-response-time data, features scored by automated algorithms; and

- evaluation of the reasoning processes students employ when solving test items (Emberetson, 1983; Messick, 1989; Mislevy, 2009).

This type of evidence was used to confirm that the Smarter Balanced assessments are measuring the cognitive skills that are intended to be the objects of measurement and that students are using these targeted skills to respond to the items.

8. 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 analysis or multidimensional IRT scaling (both exploratory and confirmatory). 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. In addition, analysis of item functioning using Item Response Theory (IRT) and differential item functioning (DIF) fall under the internal structure category. For Smarter Balanced, a dimensionality study was conducted in the Pilot Test to determine the factor structure of the assessments and the types of scales developed as well as the associated IRT models used to calibrate them.
9. 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 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). This type of evidence is essential for supporting the validity of certain inferences based on scores from the Smarter Balanced assessments for certifying college and career readiness, which is one of the primary test purposes. A subset of students who took NAEP and PISA items also took Smarter Balanced items and performance tasks. A summary of the resulting item performance for NAEP, PISA, and all Smarter Balanced items was conducted, the results are discussed in Chapter 5. Usually, association with other assessment results requires a data set with a common set of either test items or examinees. The Consortium has not yet acquired a data set from its members so a study of this type is out of the scope of this manual. However, other organizations have studied the association between Smarter Balanced tests and other tests of similar constructs. These include linking Smarter Balanced to California’s current high school graduation tests (ETS, 2015) and linking course grades to Smarter Balanced cut scores in Washington (OSPI, 2016)
10. Finally, evidence based on *consequences of testing* refers to the evaluation of the intended and unintended consequences associated with a testing program. Examples of evidence based on testing consequences include investigations of adverse impact, evaluation of the effects of testing on instruction, and evaluation of the effects of testing on issues such as

high school dropout rates. With respect to educational tests, the *Standards* stress the importance of evaluating test consequences. For example, they state,

When educational testing programs are mandated . . . the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the use of the test, both intended and unintended, should also be examined by the test user (AERA et al., 2014, p. 145).

Investigations of testing consequences relevant to the Smarter Balanced goals include analyses of students' opportunity to learn with regard to the Common Core State Standards, and analyses of changes in textbooks and instructional approaches. Unintended consequences, such as changes in instruction, diminished morale among teachers and students, increased pressure on students leading to increased dropout rates, or the pursuit of college majors and careers that are less challenging, can be evaluated. These studies are beyond the scope of this report.

Conclusion for Summative Test Validity Results

Validation is an ongoing, essentially 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

- 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/>
- 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/>
- 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.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- 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.
- 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.
- 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/>
- Smarter Balanced Assessment Consortium. (2015). *End of grant report*. Los Angeles, CA: Author. Retrieved from <http://www.smarterbalanced.org/wp-content/uploads/2015/08/SB-EndOfGrantReport-F2.pdf>

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 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 concepts of reliability and precision are examined through analysis of measurement error in simulated and operational conditions. Reliability in physical instruments is checked by repeated measurement. For example, reliability of scales are verified by seeing that the scale always gives the same weight for the same object. For assessments, it isn't possible to give the same test more than once to the same individual without altering the object of measurement. Consequently, reliability is inferred from 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.

Simulations Studies for 2014-15 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 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 2014-15 tests using packaged pools with its own simulation engine. This serves 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, 2014b).

Results from CRESST's simulation are presented here. For each grade and content area, true ability (theta) values for 1,000 simulees were created using the population distribution of the sample used in standard setting.

TABLE 2.1 POPULATION PARAMETERS USED TO GENERATE ABILITY DISTRIBUTIONS FOR SIMULATED TEST ADMINISTRATIONS

Grade	ELA/Literacy		Mathematics	
	Mean	SD	Mean	SD
3	-1.240	1.06	-1.285	0.97
4	-0.748	1.11	-0.708	1.00
5	-0.310	1.10	-0.345	1.08
6	-0.055	1.11	-0.100	1.19
7	0.114	1.13	0.010	1.33
8	0.382	1.13	0.176	1.42
11	0.529	1.19	0.506	1.52

Using the adaptive algorithm (Cohen & Albright, 2014) 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 2014-15 administration, the test scoring algorithm resolved extreme scores by using the highest and lowest obtainable scale scores (HOSS and LOSS) derived during 2014 achievement level setting. Scores above HOSS or below LOSS are assigned HOSS and LOSS values. This provides a limit to the score range, which is desired in public reporting.

TABLE 2.2 HOSS/LOSS VALUES IN LOGIT UNITS AND PERCENTAGES OF AFFECTED SIMULATION RESULTS

Grade	Obtainable Score Range		Percentage of Affected Scores	
	LOSS	HOSS	LOSS	HOSS
English Language Arts/Literacy				
3	-4.59	1.34	0.7	1.1
4	-4.40	1.80	0.3	2.0
5	-3.58	2.25	1.0	2.1
6	-3.48	2.51	0.8	1.7
7	-2.91	2.75	1.4	1.6
8	-2.57	3.04	1.5	1.9
11	-2.44	3.34	1.5	1.4
Mathematics				
3	-4.11	1.33	0.5	0.9
4	-3.92	1.82	0.3	1.1
5	-3.73	2.33	1.0	1.6
6	-3.53	2.95	0.8	1.1
7	-3.34	3.32	2.2	1.2
8	-3.15	3.63	2.8	1.2
11	-2.96	4.38	3.3	1.2

Statistics for simulations computed 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.
- 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. It is the marginal reliability for the simulated population.
- Standard error 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.

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

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

where θ_i is the true score and $\hat{\theta}_i$ is the estimated (observed) score.

Variance of the bias:

$$\text{var}(bias) = \frac{1}{N(N-1)} \sum_{i=1}^N (\theta_i - \bar{\hat{\theta}})^2 \quad (3)$$

where, $\bar{\hat{\theta}}$ is an average of the estimated theta.

Significance of the bias is then tested as:

$$z = bias / \sqrt{\text{var}(bias)} \quad (4)$$

A p-value for the significance of the bias is reported from this z test.

The average standard error is computed as:

$$\text{mean}(se) = \sqrt{N^{-1} \sum_{i=1}^N se_i^2} \quad (5)$$

where $se(\hat{\theta}_i)^2$ is the standard error of the estimated θ for individual i .

To determine the number of students falling outside the 95% and 99% confidence interval coverage, a t-test is performed as follows:

$$t = \frac{\theta_i - \hat{\theta}_i}{se(\hat{\theta}_i)} \quad (6)$$

where $\hat{\theta}$ is the ability estimate for individual i , and θ is the true score for individual i . The percentage of students' estimated theta falling outside the coverage 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.

TABLE 2.3 BIAS OF THE ESTIMATED PROFICIENCIES: ENGLISH LANGUAGE ARTS/LITERACY

Grade	Mean Bias	SE of Mean Bias	p-value for the z-Test	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall						
3	0.00	0.03	0.89	0.10	4.6	1.3
4	0.01	0.04	0.81	0.11	6.2	1.2
5	-0.01	0.03	0.75	0.10	4.8	1.0
6	0.00	0.04	0.92	0.11	4.5	0.4
7	0.01	0.04	0.87	0.12	4.2	1.2
8	0.02	0.04	0.59	0.11	4.1	0.5
11	0.00	0.04	0.98	0.14	5.7	1.2
Claim 1: Reading						
3	0.09	0.03	0.01	0.35	6.8	2.8
4	0.06	0.04	0.07	0.38	5.4	1.9
5	0.04	0.04	0.20	0.32	5.7	1.8
6	0.07	0.04	0.04	0.43	4.5	1.5
7	0.06	0.04	0.12	0.42	5.6	1.2
8	0.08	0.04	0.03	0.39	5.5	2.0
11	0.04	0.04	0.34	0.43	5.7	2.0
Claim 2: Writing						
3	0.01	0.03	0.75	0.32	5.4	1.1
4	0.02	0.04	0.64	0.32	6.7	1.6
5	-0.02	0.03	0.62	0.32	6.2	1.3
6	0.02	0.04	0.55	0.34	4.7	1.3
7	0.05	0.04	0.17	0.38	6.2	2.1
8	0.02	0.04	0.58	0.30	3.2	1.2
11	0.04	0.04	0.32	0.49	6.1	1.5
Claim 3: Speaking/Listening						
3	0.11	0.03	0.00	0.87	9.3	5.7
4	0.10	0.04	0.01	0.87	8.2	5.0
5	0.09	0.04	0.01	0.83	8.5	5.1
6	0.10	0.04	0.00	0.87	7.9	3.8
7	0.02	0.04	0.56	0.74	5.9	2.8
8	0.05	0.04	0.13	0.83	7.6	4.0
11	0.00	0.04	0.98	0.93	7.2	3.8
Claim 4: Research						
3	0.17	0.03	0.00	0.87	12.9	7.9
4	0.19	0.04	0.00	0.99	10.9	6.8
5	0.05	0.04	0.13	0.59	8.8	5.1
6	0.19	0.04	0.00	1.01	12.9	8.2
7	0.20	0.04	0.00	0.98	14.6	8.8
8	0.11	0.04	0.00	0.79	11.0	6.8
11	0.18	0.04	0.00	0.88	12.2	7.4

TABLE 2.4 BIAS OF THE ESTIMATED PROFICIENCIES: MATHEMATICS

Grade	Mean Bias	SE of Mean Bias	p-value for the z-Test	MSE	95% CI Miss Rate	99% CI Miss Rate
Overall						
3	0.00	0.03	0.99	0.06	4.5	0.9
4	0.01	0.03	0.69	0.08	5.5	1.6
5	0.03	0.03	0.33	0.13	4.5	1.3
6	0.01	0.04	0.80	0.11	4.2	0.8
7	0.00	0.04	0.93	0.19	5.3	1.0
8	0.00	0.05	0.99	0.20	4.3	0.8
11	0.02	0.05	0.72	0.25	4.8	1.2
Claim 1: Concepts and Procedures						
3	-0.01	0.03	0.83	0.12	5.3	0.8
4	0.03	0.03	0.28	0.15	4.5	0.9
5	0.06	0.03	0.06	0.25	4.9	1.7
6	0.02	0.04	0.54	0.21	4.2	0.6
7	0.06	0.04	0.18	0.37	7.3	1.8
8	0.04	0.05	0.38	0.36	5.8	0.7
11	0.04	0.05	0.42	0.46	4.8	1.4
Claim 2/4: Problem Solving/Modeling and Data Analysis						
3	0.10	0.03	0.00	0.39	8.4	4.9
4	0.13	0.03	0.00	0.55	10.1	5.2
5	0.29	0.04	0.00	1.03	15.5	9.1
6	0.17	0.04	0.00	0.82	12.4	6.7
7	0.23	0.04	0.00	1.30	15.8	7.3
8	0.36	0.05	0.00	1.64	20.1	10.2
11	0.39	0.05	0.00	1.73	18.2	9.9
Claim 3: Communicating Reasoning						
3	0.17	0.03	0.00	0.62	12.2	8.2
4	0.15	0.03	0.00	0.55	8.7	5.4
5	0.20	0.03	0.00	0.76	11.1	6.1
6	0.22	0.04	0.00	0.89	11.2	6.0
7	0.29	0.04	0.00	1.29	12.9	6.9
8	0.13	0.05	0.01	0.96	9.5	3.9
11	0.20	0.05	0.00	1.20	9.1	3.6

Bias in overall scores is both small and insignificant. Claim scores do include some systematic bias. This is likely caused by application of HOSS and LOSS values. In the simulation, the HOSS was applied more often than the LOSS which caused some asymmetry in estimates.

Table 2.5 and Table 2.6 below show marginal reliability (mean ρ) and precision overall and by 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 error is higher. Claims with smaller numbers of items and fewer points from the adaptive section of the test exhibit the lowest reliability. (These are claims 2 and 4 in English Language Arts/literacy (ELA/literacy) and claims 2/4 and 3 in mathematics.) This shows the importance of incorporating error in claim level reports.

Table 2.7 shows average error by decile of the true thetas, which were generated based on the achievement-level-setting population. One of the advantages of adaptive tests is that error can be controlled for all levels of theta. Table 2.7 shows that error at the high end of the test is consistent with overall error and only slightly above error for the mean population. There is some increase in error at the first and second deciles, caused by the relative difficulty of the pools noted in Chapter 4.

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. Details of simulations for Braille and stacked Spanish tests can be found in the full simulation report (National Center for Research on Evaluation, Standards, & Student Testing [CRESST], 2015a). Details of simulations for the American Sign Language and translated glossary pools can be found in a separate report (CRESST, 2016).

TABLE 2.5 OVERALL SCORE AND CLAIM SCORE PRECISION/RELIABILITY: ENGLISH LANGUAGE ARTS/LITERACY

Grade	Overall ELA/L					Claim 1					Claim 2					Claim 3					Claim 4				
	ave # items	SD($\hat{\theta}$)	mean SEM	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
3	45.4	1.1	.31	.31	.92	16.0	1.3	.51	.59	.79	12.0	1.3	.54	.57	.80	9.0	1.5	.85	.93	.59	8.4	1.5	.71	.94	.60
4	45.5	1.2	.32	.33	.92	16.0	1.3	.58	.62	.78	12.0	1.3	.53	.56	.81	9.0	1.5	.85	.93	.59	8.5	1.5	.78	.99	.58
5	45.8	1.2	.31	.31	.93	16.0	1.3	.54	.57	.80	12.0	1.3	.53	.57	.81	9.0	1.4	.87	.91	.60	8.8	1.4	.67	.77	.70
6	43.3	1.2	.33	.33	.92	14.0	1.3	.66	.65	.75	12.0	1.3	.55	.58	.80	9.0	1.5	.88	.93	.60	8.3	1.5	.77	1.01	.58
7	43.1	1.2	.35	.35	.91	14.0	1.3	.65	.65	.75	12.0	1.3	.58	.62	.78	9.0	1.4	.87	.86	.63	8.2	1.5	.79	.99	.58
8	43.5	1.2	.34	.34	.92	14.0	1.3	.61	.62	.78	12.0	1.3	.56	.54	.82	9.0	1.5	.90	.91	.61	8.5	1.5	.78	.89	.62
11	45.4	1.2	.37	.37	.91	16.0	1.4	.63	.65	.77	12.0	1.4	.67	.70	.75	9.0	1.5	.95	.96	.58	8.4	1.5	.86	.94	.63

TABLE 2.6 OVERALL SCORE AND CLAIM SCORE PRECISION/RELIABILITY: MATHEMATICS

Grade	Overall Mathematics					Claim 1					Claim 2/4					Claim 3				
	ave # items	SD($\hat{\theta}$)	mean SEM	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$	ave # items	SD($\hat{\theta}$)	mean SE($\hat{\theta}$)	RMSE	$\bar{\rho}$
3	39.7	1.0	.25	.25	.94	20.0	1.1	.35	.35	.89	9.9	1.2	.52	.63	.74	9.8	1.3	.61	.79	.63
4	39.2	1.1	.28	.28	.93	20.0	1.1	.38	.39	.88	9.6	1.3	.57	.74	.69	9.6	1.3	.62	.74	.67
5	39.7	1.2	.35	.36	.91	20.0	1.3	.48	.50	.84	9.8	1.6	.64	1.01	.61	9.9	1.4	.65	.87	.63
6	38.8	1.3	.35	.34	.93	19.0	1.3	.47	.46	.88	9.8	1.6	.67	.91	.67	10.0	1.6	.76	.94	.64
7	39.4	1.4	.44	.44	.90	20.0	1.5	.58	.61	.83	10.0	1.8	.81	1.14	.60	9.4	1.7	.95	1.14	.57
8	38.8	1.5	.46	.45	.91	20.0	1.5	.60	.60	.85	9.1	2.0	.86	1.28	.58	9.7	1.7	.88	.98	.66
11	41.3	1.6	.52	.50	.90	22.0	1.6	.69	.68	.83	9.3	2.1	.95	1.31	.60	10.0	1.9	1.04	1.10	.66

TABLE 2.7 AVERAGE STANDARD ERRORS BY GRADE AND BY DECILES OF TRUE PROFICIENCY SCORES

Proficiency Score Distribution											
Grade	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Overall
English Language Arts/Literacy											
3	0.49	0.32	0.29	0.27	0.26	0.25	0.25	0.25	0.25	0.28	0.30
4	0.45	0.33	0.30	0.29	0.28	0.28	0.28	0.28	0.28	0.30	0.31
5	0.43	0.31	0.29	0.28	0.27	0.27	0.27	0.27	0.28	0.31	0.30
6	0.48	0.37	0.32	0.30	0.28	0.28	0.28	0.28	0.28	0.31	0.32
7	0.50	0.39	0.35	0.33	0.31	0.30	0.29	0.29	0.29	0.31	0.34
8	0.48	0.37	0.33	0.32	0.31	0.30	0.30	0.30	0.30	0.32	0.34
11	0.53	0.43	0.37	0.35	0.33	0.32	0.31	0.31	0.31	0.33	0.36
Mathematics											
3	0.38	0.28	0.25	0.23	0.22	0.22	0.21	0.21	0.21	0.23	0.25
4	0.45	0.32	0.27	0.25	0.23	0.22	0.21	0.21	0.21	0.23	0.26
5	0.63	0.42	0.34	0.30	0.27	0.25	0.23	0.21	0.21	0.22	0.31
6	0.57	0.41	0.36	0.32	0.31	0.28	0.26	0.25	0.24	0.25	0.33
7	0.72	0.58	0.48	0.42	0.37	0.32	0.29	0.26	0.23	0.23	0.40
8	0.73	0.57	0.50	0.45	0.40	0.37	0.34	0.30	0.27	0.26	0.43
11	0.85	0.67	0.57	0.52	0.45	0.39	0.34	0.31	0.27	0.26	0.47

Item exposure

Table 2.8 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, 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 and one item in grade 11 mathematics 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 2.8 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	591	1.35	97.29	1.35	0	0	0
4	567	0.35	97.00	2.65	0	0	0
5	546	5.86	91.58	2.20	0	0	0.37
6	548	4.56	91.42	3.65	0.37	0	0
7	508	5.71	90.16	3.94	0.20	0	0
8	499	1.00	94.79	4.21	0	0	0
11	1455	0.21	99.45	0.34	0	0	0
Mathematics							
3	829	0.48	99.16	0.36	0	0	0
4	818	0.12	99.14	0.73	0	0	0
5	807	0.12	99.38	0.50	0	0	0
6	739	0.14	99.05	0.81	0	0	0
7	670	0.15	98.66	1.19	0	0	0
8	612	0.00	98.04	1.80	0.16	0	0
11	1711	0.70	99.18	0.06	0	0	0.06

Blueprint fidelity

For target and Depth of Knowledge (DOK) constraints, the simulated ELA CAT test events met blueprint specifications with the exceptions noted in Table 2.9: nine Grade 4 tests did not have the minimum number of items at DOK level 2; one Grade 5 test did not have the minimum number of Claim 1, Target 9 items; one Grade 6 test did not have the minimum number of Claim 1, Target 2 items; and 44 Grade 6 tests did not have the minimum number of items with DOK level 2 or greater. In mathematics, all CAT portions met the blueprint requirements for targets and DOK.

TABLE 2.9 CAT TESTS WITH BLUEPRINT DEVIATIONS

Grade	Subject	Blueprint Specification	Blueprint Requirement			Percentage of Tests				
			Pg. #	Min	Max	Below Min. by 2 or more	Below Min. by 1	Above Max. by 1	Above Max. by 2 or more	Total
4	ELA	Claim 1, DOK=2	4	6	6	0	0.9	0	0	0.9
5	ELA	Claim 1 (Informational), Target 9: Central Ideas	4	1	2	0	0.1	0	0	0.1
6	ELA	Claim 1 (Literary), Target 2: Central Ideas	7	1	1	0	0.1	0	0	0.1
6	ELA	Claim 2, DOK≥2	7	5	--	0.1	4.2	0	0	4.3

As shown in Table 2.10, not all performance tasks conform to the blueprint specifications. In ELA/literacy, some have fewer research items; in mathematics some have fewer than six items. The remaining PT items still allow claim level reporting when combined with CAT items. There was a decision on the part of Smarter Balanced leadership to accept these performance tasks as operational. The Consortium will adjust the blueprint requirements to match operational PTs.

TABLE 2.10 PT TESTS WITH BLUEPRINT DEVIATIONS

Grade	Subject	Blueprint Specification	Blueprint Requirement			Percentage of Tests				
			Pg. #	Min	Max	Below Min. by 2 or more	Below Min. by 1	Above Max. by 1	Above Max. by 2 or more	Total
4	ELA	Claim 4	6	2	3	0.0	4.8	0.0	0.0	4.8
4	ELA	Claim 4, DOK≥3	6	2	3	0.0	4.8	0.0	0.0	4.8
3	Math	Claim 2	5	1	2	0.0	0.0	35.3	0.0	35.3
3	Math	Claim 4	5	1	3	0.0	11.6	4.8	0.0	16.4
3	Math	Claim 3	5	0	2	0.0	0.0	5.5	0.0	5.5
4	Math	Claim 2	7	1	2	0.0	0.0	4.8	0.0	4.8
4	Math	Claim 4	7	1	3	0.0	4.4	0.0	0.0	4.4
5	Math	Claim 2	9	1	2	0.0	0.0	13.1	0.0	13.1
6	Math	Claim 3	11	0	2	0.0	0.0	4.3	0.0	4.3
7	Math	Claim 2	13	1	2	0.0	0.0	6.1	0.0	6.1
8	Math	Claim 4	15	1	3	0.0	22.3	0.0	0.0	22.3
11	Math	Claim 2	17	1	2	0.0	0.0	5.9	0.0	5.9
11	Math	Claim 4	17	1	3	0.0	10.4	0.0	0.0	10.4
11	Math	Claim 3	17	0	2	0.0	0.0	0.0	5.9	5.9

Observed Reliability

Observed reliability is derived from standard errors of measurement computed from the test form each student took. The method of standard error calculation for both total and score reporting category scores, as described in Smarter Balanced Scoring Specifications for 2014-15 (AIR, 2014a), is displayed below:

The standard error (SE) for student i is:

$$SE(\theta_i) = \frac{1}{\sqrt{I(\theta_i)}}$$

where $I(\theta_i)$ is the test information for student i , calculated as:

$$I(\theta_i) = \sum_{j=1}^I D^2 a_j^2 \left(\frac{\sum_{l=1}^{m_j} l^2 \text{Exp}(\sum_{k=1}^l D a_j (\theta_i - b_{jk}))}{1 + \sum_{l=1}^{m_j} \text{Exp}(\sum_{k=1}^l D a_j (\theta_i - b_{jk}))} - \left(\frac{\sum_{l=1}^{m_j} l \text{Exp}(\sum_{k=1}^l D a_j (\theta_i - b_{jk}))}{1 + \sum_{l=1}^{m_j} \text{Exp}(\sum_{k=1}^l D a_j (\theta_i - b_{jk}))} \right)^2 \right)$$

where m_j is the maximum possible score point (starting from 0) for the j th item, D is the scale factor, 1.7. Values of a_j and b_{jk} are item parameters for item j and score level k .

SE is calculated based only on the answered items. The upper bound of SE is set to 2.5 on theta metric. Any value larger than 2.5 is truncated at 2.5 on theta metric.

Standard errors reported here have been transformed to the reporting scale metric. This transformation is:

$$SE_{vs} = a * SE_{\theta_i}$$

where SE_{θ} is the standard error of the ability estimate on the θ scale and a is the slope of the scaling constants that take θ to the reporting scale.

Because the set of items administered to each student in a Smarter Balanced adaptive test is virtually unique, standard error is estimated for each test event. Reliability for each total score and claim/reporting category score is derived from the SE using the following relationship:

$$SE(\theta_i) = s_x \sqrt{1 - r}$$

Where s_x is the standard deviation of the score and r is reliability for that particular test event.

$$r = 1 - \left(\frac{SE}{s_x} \right)^2$$

Reliability for a test instrument as a whole is called marginal reliability and is estimated as one minus the ratio of mean error variance to observed score variance. Marginal reliability for the data submitted by members¹ using the general test pools are given in Table 2.11 and Table 2.12. Because claim scores are based on fewer items, they have lower reliability than total scores.

TABLE 2.11 ELA SUMMATIVE SCALE MARGINAL RELIABILITY ESTIMATES

Grade	N	Total score	Claim 1	Claim 2	Claim 3	Claim 4
3	710,836	.920	.739	.768	.544	.608
4	674,936	.918	.737	.769	.559	.624
5	698,426	.922	.754	.790	.517	.640
6	689,274	.908	.690	.777	.477	.587
7	677,287	.915	.730	.778	.480	.607
8	679,863	.915	.739	.766	.508	.610
11	610,761	.918	.737	.770	.528	.637

TABLE 2.12 MATH SUMMATIVE SCALE SCORE MARGINAL RELIABILITY ESTIMATES

Grade	N	Total Score	Claim 1	Claim 2/4	Claim 3
3	717,519	.940	.884	.654	.601
4	702,093	.941	.889	.601	.698
5	699,713	.929	.875	.561	.596
6	689,045	.925	.867	.567	.616
7	681,387	.909	.848	.537	.486
8	681,197	.918	.850	.539	.662
11	557,386	.893	.817	.496	.541

Table 2.13 and Table 2.14 show that reliability varies by overall score levels. All students take the same number of items, but the information delivered by the items differs. The most information, and hence lower error and higher reliability, is found where the pool has the most items. Smarter Balanced pools are difficult relative to the population. Students with lower scores (deciles 1 and 2) have lower reliability than those with higher scores (deciles 8 and 9.)

Because of the differences by score level, demographic groups with lower average scores tend to have lower reliability than the population as a whole. Table 2.15 to Table 2.18 show marginal reliability by demographic group.

¹ Data for the marginal reliability analysis provided by the following Consortium members: Delaware, Hawaii, Idaho, Oregon, South Dakota, US Virgin Islands, Vermont, Washington, West Virginia, California, Montana, Nevada, and North Dakota.

TABLE 2.13 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	710,836	7604.7	0.92	0.86	0.91	0.92	0.93	0.93	0.93	0.93	0.93	0.93	0.92
4	674,936	8482.2	0.92	0.87	0.91	0.92	0.92	0.93	0.93	0.93	0.93	0.93	0.91
5	698,426	8676.2	0.92	0.88	0.92	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.91
6	689,274	8265.2	0.91	0.84	0.90	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.90
7	677,287	9136.0	0.92	0.86	0.91	0.92	0.92	0.93	0.93	0.93	0.93	0.93	0.91
8	679,863	8927.5	0.92	0.87	0.91	0.92	0.92	0.92	0.93	0.92	0.92	0.92	0.91
11	610,761	11648.4	0.92	0.87	0.91	0.92	0.93	0.93	0.93	0.93	0.93	0.93	0.91

TABLE 2.14 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	717,519	6344.6	0.94	0.87	0.93	0.94	0.95	0.95	0.95	0.96	0.96	0.96	0.94
4	702,093	6516.4	0.94	0.88	0.93	0.94	0.95	0.95	0.95	0.96	0.96	0.96	0.94
5	699,713	7931.6	0.93	0.83	0.90	0.92	0.93	0.94	0.95	0.95	0.96	0.96	0.95
6	689,045	10227.9	0.93	0.76	0.90	0.93	0.94	0.95	0.95	0.96	0.96	0.96	0.95
7	681,387	11601.8	0.91	0.67	0.87	0.91	0.92	0.94	0.95	0.96	0.96	0.97	0.96
8	681,197	13377.6	0.92	0.77	0.89	0.90	0.92	0.93	0.94	0.95	0.96	0.96	0.96
11	557,386	15263.9	0.89	0.65	0.83	0.87	0.90	0.92	0.93	0.95	0.96	0.96	0.96

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

TABLE 2.15 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY ETHNIC GROUP – ELA/LITERACY

Grade	Group	N	Var	MSE	Marginal Reliability
3	All	710,836	7604.7	608.8	0.92
	American Indian/Alaska Native	14,921	5989.6	640.2	0.89
	Asian	55,234	7454.8	575.5	0.92
	Black/African American	40,326	6698.8	653.2	0.90
	Hispanic/Latino Ethnicity	298,651	6191.9	640.6	0.90
	White	287,185	7063.8	578.7	0.92
4	All	674,936	8482.2	694.7	0.92
	American Indian/Alaska Native	15,171	6793.9	723.4	0.89
	Asian	54,808	8208.7	676.6	0.92
	Black/African American	38,088	7425.6	734.3	0.90
	Hispanic/Latino Ethnicity	275,647	6956.0	716.5	0.90
	White	280,544	7711.2	671.6	0.91
5	All	698,426	8676.2	673.5	0.92
	American Indian/Alaska Native	15,639	7153.0	693.8	0.90
	Asian	58,312	8492.5	682.3	0.92
	Black/African American	39,826	7835.9	701.9	0.91
	Hispanic/Latino Ethnicity	282,000	7239.3	679.2	0.91
	White	290,464	7813.3	660.9	0.92
6	All	689,274	8265.2	762.2	0.91
	American Indian/Alaska Native	15,029	7280.1	806.9	0.89
	Asian	57,697	7975.0	743.5	0.91
	Black/African American	39,385	7673.3	819.7	0.89
	Hispanic/Latino Ethnicity	275,926	6989.1	787.2	0.89
	White	288,296	7510.4	735.2	0.90
7	All	677,287	9136.0	774.2	0.92
	American Indian/Alaska Native	14,642	7967.0	811.1	0.90
	Asian	55,716	8660.1	763.4	0.91
	Black/African American	39,389	8178.6	825.3	0.90
	Hispanic/Latino Ethnicity	271,234	7650.7	797.0	0.90
	White	284,851	8267.9	747.5	0.91
8	All	679,863	8927.5	758.1	0.92
	American Indian/Alaska Native	14,763	8125.6	783.1	0.90
	Asian	56,159	8506.1	751.7	0.91
	Black/African American	40,677	8260.7	801.4	0.90
	Hispanic/Latino Ethnicity	270,892	7382.9	769.4	0.90
	White	286,694	8338.0	742.3	0.91
11	All	610,761	11648.4	955.1	0.92
	American Indian/Alaska Native	12,247	10778.5	984.2	0.91
	Asian	52,572	11248.4	946.6	0.92
	Black/African American	35,219	11289.4	1023.1	0.91
	Hispanic/Latino Ethnicity	241,033	10209.5	971.1	0.90
	White	259,062	11125.0	935.0	0.92

TABLE 2.16 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY ETHNIC GROUP – MATHEMATICS

Grade	Group	N	Var	MSE	Marginal Reliability
3	All	717,519	6344.6	381.2	0.94
	American Indian/Alaska Native	15,225	5314.7	413.1	0.92
	Asian	56,156	6233.9	351.5	0.94
	Black/African American	40,576	5754.2	446.8	0.92
	Hispanic/Latino Ethnicity	300,768	5142.9	411.6	0.92
	White	290,552	5757.4	348.6	0.94
4	All	702,093	6516.4	386.9	0.94
	American Indian/Alaska Native	15,362	5404.0	423.6	0.92
	Asian	57,666	6597.1	365.9	0.94
	Black/African American	39,623	5564.6	445.1	0.92
	Hispanic/Latino Ethnicity	290,156	5047.1	418.5	0.92
	White	286,800	5859.4	353.9	0.94
5	All	699,713	7931.6	563.3	0.93
	American Indian/Alaska Native	15,679	6484.3	652.8	0.90
	Asian	58,936	7911.6	449.5	0.94
	Black/African American	39,859	6782.1	713.9	0.89
	Hispanic/Latino Ethnicity	282,803	6135.0	653.1	0.89
	White	290,143	7025.9	484.0	0.93
6	All	689,045	10227.9	767.0	0.93
	American Indian/Alaska Native	15,007	8885.5	908.4	0.90
	Asian	58,250	9682.1	594.3	0.94
	Black/African American	39,489	9295.4	1005.0	0.89
	Hispanic/Latino Ethnicity	276,665	8462.5	896.1	0.89
	White	286,870	9056.9	653.4	0.93
7	All	681,387	11601.8	1052.0	0.91
	American Indian/Alaska Native	14,742	9556.9	1201.5	0.87
	Asian	56,448	11164.6	701.0	0.94
	Black/African American	39,618	10137.3	1485.0	0.85
	Hispanic/Latino Ethnicity	273,640	9442.6	1304.0	0.86
	White	285,296	10117.0	832.7	0.92
8	All	681,197	13377.6	1103.1	0.92
	American Indian/Alaska Native	14,800	10961.5	1255.0	0.89
	Asian	56,575	13344.8	784.9	0.94
	Black/African American	40,781	10913.0	1412.6	0.87
	Hispanic/Latino Ethnicity	272,763	10489.2	1309.7	0.88
	White	285,717	12231.1	941.3	0.92
11	All	557,386	15263.9	1637.9	0.89
	American Indian/Alaska Native	10,261	11823.0	1894.8	0.84
	Asian	49,068	16186.1	1022.5	0.94
	Black/African American	33,071	11970.3	2200.7	0.82
	Hispanic/Latino Ethnicity	231,810	11726.4	1929.2	0.84
	White	216,482	14880.3	1423.0	0.90

TABLE 2.17 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY GROUP – ELA/LITERACY

Grade	Group	N	Var	MSE	Marginal Reliability
3	All	710836	7,605	608.8	0.92
	LEP	177,467	5333.8	670.6	0.87
	IDEA Indicator	74,867	7090.8	745.0	0.89
	Section 504	5,412	7348.1	597.9	0.92
	Economically Disadvantaged	415,784	6279.1	636.8	0.90
4	All	674936	8,482	694.7	0.92
	LEP	135,787	5144.7	763.6	0.85
	IDEA Indicator	76,171	7548.2	826.5	0.89
	Section 504	6,631	7696.2	681.9	0.91
	Economically Disadvantaged	389,689	7064.1	714.7	0.90
5	All	698426	8,676	673.5	0.92
	LEP	118,989	4876.8	726.1	0.85
	IDEA Indicator	80,326	7222.4	802.2	0.89
	Section 504	8,278	7769.3	658.8	0.92
	Economically Disadvantaged	397,152	7317.3	680.0	0.91
6	All	689274	8,265	762.2	0.91
	LEP	89,606	4935.7	915.0	0.81
	IDEA Indicator	75,816	6614.0	982.9	0.85
	Section 504	8,917	7105.8	741.8	0.90
	Economically Disadvantaged	385,768	7105.2	789.8	0.89
7	All	677287	9,136	774.2	0.92
	LEP	77,054	4818.7	941.9	0.80
	IDEA Indicator	71,493	6559.0	977.6	0.85
	Section 504	9,753	8077.3	755.1	0.91
	Economically Disadvantaged	375,293	7833.1	799.4	0.90
8	All	679863	8,928	758.1	0.92
	LEP	68,766	4650.4	883.9	0.81
	IDEA Indicator	69,971	6319.8	911.4	0.86
	Section 504	10,787	8172.6	753.1	0.91
	Economically Disadvantaged	370,756	7687.4	773.5	0.90
11	All	610761	11,648	955.1	0.92
	LEP	44,608	6096.9	1203.2	0.80
	IDEA Indicator	51,151	8729.6	1185.2	0.86
	Section 504	10,574	10573.9	935.4	0.91
	Economically Disadvantaged	304,694	10660.5	980.4	0.91

TABLE 2.18 MARGINAL RELIABILITY OF TOTAL SUMMATIVE SCORES BY GROUP – MATHEMATICS

Grade	Group	N	Var	MSE	Marginal Reliability
3	All	717519	6,345	381.2	0.94
	LEP	180,367	4995.0	434.9	0.91
	IDEA Indicator	75,305	7706.7	569.7	0.93
	Section 504	5,481	6578.8	371.7	0.94
	Economically Disadvantaged	419,249	5353.1	408.7	0.92
4	All	702093	6,516	386.9	0.94
	LEP	144,463	4417.3	471.4	0.89
	IDEA Indicator	79,083	6917.9	556.8	0.92
	Section 504	6,806	6179.7	373.3	0.94
	Economically Disadvantaged	406,485	5302.2	414.7	0.92
5	All	699713	7,932	563.3	0.93
	LEP	120,938	4920.0	797.8	0.84
	IDEA Indicator	80,149	7281.0	902.7	0.88
	Section 504	8,253	7320.6	520.0	0.93
	Economically Disadvantaged	397,967	6427.2	642.5	0.90
6	All	689045	10,228	767.0	0.93
	LEP	91,336	7299.7	1278.8	0.82
	IDEA Indicator	75,647	9673.6	1549.5	0.84
	Section 504	8,880	8891.9	691.8	0.92
	Economically Disadvantaged	386,612	8744.7	891.2	0.90
7	All	681387	11,602	1052.0	0.91
	LEP	79,345	8095.7	2032.8	0.75
	IDEA Indicator	71,853	9760.5	2271.6	0.77
	Section 504	9,768	10301.3	941.3	0.91
	Economically Disadvantaged	378,213	9756.5	1277.5	0.87
8	All	681197	13,378	1103.1	0.92
	LEP	70,952	8762.6	1930.5	0.78
	IDEA Indicator	70,183	9466.8	1963.0	0.79
	Section 504	10,749	12371.6	1034.6	0.92
	Economically Disadvantaged	372,791	11057.8	1288.8	0.88
11	All	557386	15,264	1637.9	0.89
	LEP	43,248	9595.6	3137.9	0.67
	IDEA Indicator	47,480	9193.7	3063.7	0.67
	Section 504	8,347	14429.0	1572.4	0.89
	Economically Disadvantaged	284,662	12484.5	1926.9	0.85

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. The paper/pencil tests are on the Smarter Balanced scale with parameters estimated using a set of anchors from online administrations (CRESST, 2015b)². The number of paper pencil tests administered is presented in the Chapter 5 Addendum.

TABLE 2.19 RELIABILITY OF PAPER PENCIL TESTS, FORM 1 ENGLISH LANGUAGE ARTS/LITERACY

Gr	Full test					Claim 1		Claim 2		Claim 3		Claim 4	
	N items	Reliability	SEM	Avg. <i>b</i>	Avg. <i>a</i>	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 2.20 RELIABILITY OF PAPER PENCIL TEST, FORM 1 MATHEMATICS

Gr	Full test					Claim 1		Claim 2/4		Claim 3	
	N items	Reliability	SEM	Avg. <i>b</i>	Avg. <i>a</i>	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

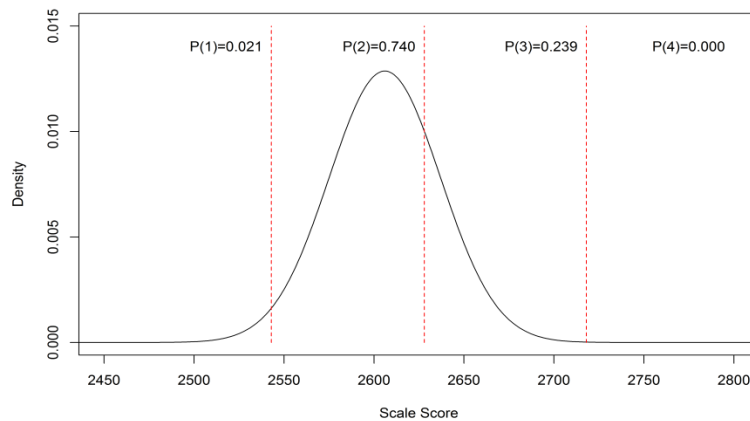
Classification accuracy is a measure of how accurately test scores or sub-scores place students into reporting category levels. The likelihood of inaccurate placement depends on the amount of error associated with scores, especially those nearest cut points. For this report, classification accuracy was calculated in the following manner³. For each examinee, analysts constructed a normal distribution with means equal to the scale score estimate and standard deviation equal to the standard error of measurement as a plausible true score distribution. For each student, the proportion of that normal distribution that fell within each level was calculated.

² Data for the paper/pencil parameter estimation provided by Michigan.

³ Data for the analysis of classification accuracy provided by the following Consortium members: Delaware, Hawaii, Idaho, Oregon, South Dakota, US Virgin Islands, Vermont, Washington, West Virginia, California, Montana, Nevada, and North Dakota.

The figure below illustrates the approach for one examinee in Grade 11 mathematics. In this example, the examinee's overall scale score is 2606, with a standard error of measurement of 31 points. Accordingly, a normal distribution with mean of 2606 and standard deviation of 31 is used to approximate a plausible distribution for this examinee's true score, based on the observed test performance. From this distribution, we obtain the probability that the true score falls within each level.

FIGURE 2.1 ILLUSTRATIVE EXAMPLE OF A NORMAL DISTRIBUTION USED TO CALCULATE CLASSIFICATION ACCURACY



The table below shows the results for 10 examinees from Grade 11 Mathematics (the examinee illustrated above is Student #6).

TABLE 2.21 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 assigned performance/achievement level. These sums can then be expressed as a proportion of the total sample.

TABLE 2.22 EXAMPLE ESTIMATED DISTRIBUTION OF TRUE ACHIEVEMENT LEVELS FOR EACH ASSIGNED ACHIEVEMENT LEVEL

Assigned Level	N	P	Expected Frequency				Expected Proportion			
			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 assigned level (excluding the “At/Near Standard” classification for claims). This is the proportion of students whose true classification matches the assigned level, among the subset of students with that assigned level. In the table below, the calculations are illustrated.

TABLE 2.23 ILLUSTRATIVE EXAMPLE OF CORRECT CLASSIFICATION RATES

Assigned Level	P	Expected Proportion				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 classification rate is the sum of the proportions of students whose true score level matches the assigned level, divided by the total proportion of students assigned to a level. This denominator is 1 for the overall score (i.e., all students are assigned to a level). For the claim scores, the denominator is one minus the proportion whose level is deemed “At/Near Standard.” Note that for the claim scores, the “Below” 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

TABLE 2.24 GRADE 3 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	230,414	.324	.289	.035	.000	.000	.893	.800
Level 2	186,940	.263	.037	.190	.036	.000	.722	
Level 3	151,684	.213	.000	.035	.148	.030	.694	
Level 4	141,798	.199	.000	.000	.026	.173	.867	
Claim 1								
Below	266,716	.375	.300	.069	.006	.001	.983	.983
At/Near	297,493	.419	.051	.168	.148	.053		
Above	146,627	.206	.000	.004	.036	.167	.982	
Claim 2								
Below	244,147	.343	.286	.052	.005	.000	.985	.984
At/Near	327,043	.460	.063	.181	.159	.057		
Above	139,646	.196	.000	.003	.033	.161	.984	
Claim 3								
Below	154,375	.217	.184	.025	.006	.002	.964	.966
At/Near	441,122	.621	.137	.180	.158	.146		
Above	115,339	.162	.001	.004	.017	.141	.969	
Claim 4								
Below	207,099	.291	.249	.034	.007	.002	.969	.975
At/Near	365,352	.514	.124	.157	.139	.093		
Above	138,385	.195	.000	.003	.024	.167	.983	
All Students	710,836	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.25 GRADE 4 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	238,881	.354	.319	.034	.000	.000	.902	.792
Level 2	145,023	.215	.037	.141	.037	.000	.655	
Level 3	148,495	.220	.000	.038	.149	.033	.678	
Level 4	142,537	.211	.000	.000	.028	.183	.865	
Claim 1								
Below	237,575	.352	.306	.040	.006	.001	.983	.983
At/Near	296,804	.440	.085	.144	.145	.066		
Above	140,557	.208	.000	.003	.031	.174	.984	
Claim 2								
Below	216,274	.320	.277	.039	.004	.000	.986	.984
At/Near	323,360	.479	.079	.165	.166	.069		
Above	135,302	.200	.000	.004	.031	.165	.980	
Claim 3								
Below	140,813	.209	.186	.017	.005	.001	.969	.968
At/Near	421,188	.624	.176	.150	.155	.144		
Above	112,935	.167	.001	.004	.018	.144	.967	
Claim 4								
Below	189,995	.282	.251	.023	.006	.002	.973	.976
At/Near	354,113	.525	.146	.138	.144	.097		
Above	130,828	.194	.000	.003	.025	.165	.981	
All Students	674,936	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.26 GRADE 5 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	213,337	.305	.274	.031	.000	.000	.898	.802
Level 2	149,681	.214	.034	.145	.035	.000	.678	
Level 3	205,097	.294	.000	.039	.223	.032	.758	
Level 4	130,311	.187	.000	.000	.027	.159	.854	
Claim 1								
Below	244,005	.349	.297	.047	.005	.000	.984	.984
At/Near	299,292	.429	.064	.151	.178	.035		
Above	155,129	.222	.000	.004	.057	.162	.983	
Claim 2								
Below	213,827	.306	.260	.042	.004	.000	.987	.986
At/Near	314,885	.451	.063	.158	.190	.040		
Above	169,714	.243	.000	.004	.059	.180	.983	
Claim 3								
Below	151,681	.217	.193	.016	.006	.002	.963	.967
At/Near	435,139	.623	.192	.137	.171	.123		
Above	111,606	.160	.001	.004	.024	.132	.973	
Claim 4								
Below	132,548	.190	.161	.022	.005	.001	.965	.975
At/Near	369,356	.529	.109	.153	.201	.066		
Above	196,522	.281	.000	.005	.060	.216	.981	
All Students	698,426	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.27 GRADE 6 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	178,252	.259	.226	.032	.000	.000	.876	.791
Level 2	198,124	.287	.038	.208	.041	.000	.724	
Level 3	214,529	.311	.000	.042	.238	.031	.765	
Level 4	98,369	.143	.000	.000	.024	.119	.831	
Claim 1								
Below	243,616	.353	.294	.052	.007	.000	.979	.980
At/Near	326,392	.474	.080	.175	.178	.040		
Above	119,266	.173	.000	.003	.047	.123	.982	
Claim 2								
Below	210,339	.305	.229	.072	.004	.000	.985	.984
At/Near	330,967	.480	.034	.200	.216	.030		
Above	147,968	.215	.000	.004	.064	.147	.981	
Claim 3								
Below	121,172	.176	.153	.018	.004	.001	.970	.961
At/Near	467,046	.678	.152	.172	.200	.153		
Above	101,056	.147	.002	.006	.022	.118	.950	
Claim 4								
Below	118,079	.171	.143	.022	.005	.001	.963	.974
At/Near	406,351	.590	.125	.185	.219	.060		
Above	164,844	.239	.000	.004	.066	.169	.982	
All Students	689,274	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.28 GRADE 7 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	189,691	.280	.248	.032	.000	.000	.887	.804
Level 2	171,352	.253	.035	.181	.037	.000	.715	
Level 3	226,914	.335	.000	.040	.266	.029	.795	
Level 4	89,330	.132	.000	.000	.023	.109	.826	
Claim 1								
Below	229,903	.339	.279	.054	.006	.000	.983	.982
At/Near	311,688	.460	.064	.173	.196	.027		
Above	135,696	.200	.000	.004	.066	.130	.981	
Claim 2								
Below	191,658	.283	.226	.053	.004	.000	.985	.983
At/Near	314,759	.465	.045	.182	.214	.024		
Above	170,870	.252	.000	.005	.089	.159	.982	
Claim 3								
Below	146,319	.216	.184	.025	.006	.001	.968	.963
At/Near	437,756	.646	.151	.170	.208	.118		
Above	93,212	.138	.001	.005	.025	.106	.954	
Claim 4								
Below	154,095	.228	.192	.027	.007	.001	.965	.973
At/Near	363,839	.537	.119	.165	.210	.043		
Above	159,353	.235	.000	.004	.077	.154	.981	
All Students	677,287	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.29 GRADE 8 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	160,753	.236	.206	.030	.000	.000	.872	.804
Level 2	193,649	.285	.036	.211	.038	.000	.741	
Level 3	236,489	.348	.000	.039	.278	.030	.800	
Level 4	88,972	.131	.000	.000	.023	.108	.826	
Claim 1								
Below	203,254	.299	.231	.062	.006	.000	.980	.982
At/Near	315,715	.464	.047	.187	.210	.020		
Above	160,894	.237	.000	.004	.090	.143	.983	
Claim 2								
Below	183,827	.270	.206	.060	.004	.000	.985	.984
At/Near	338,760	.498	.044	.200	.225	.030		
Above	157,276	.231	.000	.004	.074	.153	.982	
Claim 3								
Below	142,443	.210	.177	.027	.005	.001	.972	.967
At/Near	446,244	.656	.133	.191	.223	.110		
Above	91,176	.134	.001	.004	.026	.103	.959	
Claim 4								
Below	145,401	.214	.178	.028	.007	.001	.964	.973
At/Near	376,328	.554	.113	.182	.217	.042		
Above	158,134	.233	.000	.004	.075	.153	.981	
All Students	679,863	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.30 GRADE 11 ELA/LITERACY CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	114,056	.187	.162	.024	.000	.000	.869	.803
Level 2	145,745	.239	.029	.175	.034	.000	.735	
Level 3	207,092	.339	.000	.038	.264	.037	.779	
Level 4	143,868	.236	.000	.000	.034	.201	.855	
Claim 1								
Below	121,518	.199	.151	.044	.004	.000	.977	.982
At/Near	295,699	.484	.052	.190	.209	.033		
Above	193,544	.317	.000	.005	.090	.222	.984	
Claim 2								
Below	132,061	.216	.171	.041	.004	.000	.982	.984
At/Near	282,023	.462	.043	.177	.206	.036		
Above	196,677	.322	.000	.005	.085	.233	.985	
Claim 3								
Below	123,272	.202	.171	.026	.004	.001	.975	.970
At/Near	377,876	.619	.128	.175	.192	.123		
Above	109,613	.179	.001	.005	.031	.142	.965	
Claim 4								
Below	95,874	.157	.129	.021	.006	.001	.957	.974
At/Near	312,382	.511	.093	.164	.198	.057		
Above	202,505	.332	.000	.006	.082	.243	.982	
All Students	610,761	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

Mathematics Classification Accuracy

TABLE 2.31 GRADE 3 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	214,514	.299	.268	.031	.000	.000	.896	.825
Level 2	188,588	.263	.034	.195	.033	.000	.742	
Level 3	200,599	.280	.000	.034	.223	.023	.797	
Level 4	113,818	.159	.000	.000	.019	.139	.878	
Claim 1								
Below	264,624	.369	.279	.086	.004	.000	.990	.989
At/Near	257,074	.358	.019	.168	.165	.007		
Above	195,821	.273	.000	.003	.095	.175	.989	
Claim 2/4								
Below	244,464	.341	.274	.057	.007	.002	.972	.977
At/Near	309,502	.431	.054	.173	.181	.023		
Above	163,553	.228	.000	.003	.069	.155	.984	
Claim 3								
Below	180,380	.251	.209	.032	.008	.002	.958	.971
At/Near	373,327	.520	.118	.178	.187	.037		
Above	163,812	.228	.000	.003	.059	.166	.986	
All Students	717,519	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.32 GRADE 4 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	190,576	.271	.241	.031	.000	.000	.887	.837
Level 2	240,465	.342	.034	.277	.032	.000	.810	
Level 3	168,423	.240	.000	.030	.190	.020	.791	
Level 4	102,629	.146	.000	.000	.017	.129	.885	
Claim 1								
Below	301,407	.429	.269	.157	.004	.000	.992	.991
At/Near	233,114	.332	.005	.171	.149	.008		
Above	167,572	.239	.000	.002	.077	.159	.989	
Claim 2/4								
Below	245,780	.350	.257	.082	.009	.002	.968	.973
At/Near	325,653	.464	.044	.211	.176	.033		
Above	130,660	.186	.000	.003	.048	.135	.983	
Claim 3								
Below	248,107	.353	.256	.088	.008	.001	.974	.978
At/Near	309,439	.441	.039	.209	.169	.023		
Above	144,547	.206	.000	.003	.059	.144	.985	
All Students	702,093	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.33 GRADE 5 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	256,607	.367	.328	.039	.000	.000	.895	.829
Level 2	204,737	.293	.035	.228	.030	.000	.778	
Level 3	119,809	.171	.000	.027	.122	.022	.714	
Level 4	118,560	.169	.000	.000	.019	.151	.889	
Claim 1								
Below	331,381	.474	.343	.127	.004	.000	.991	.990
At/Near	217,897	.311	.010	.159	.122	.020		
Above	150,435	.215	.000	.002	.044	.168	.989	
Claim 2/4								
Below	293,649	.420	.322	.083	.010	.005	.965	.971
At/Near	279,683	.400	.043	.182	.131	.044		
Above	126,381	.181	.000	.002	.031	.147	.986	
Claim 3								
Below	261,967	.374	.301	.062	.009	.003	.969	.973
At/Near	326,357	.466	.084	.190	.129	.064		
Above	111,389	.159	.000	.003	.022	.134	.983	
All Students	699,713	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.34 GRADE 6 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	232,706	.338	.304	.034	.000	.000	.900	.822
Level 2	215,999	.313	.037	.242	.035	.000	.772	
Level 3	130,921	.190	.000	.031	.136	.023	.715	
Level 4	109,419	.159	.000	.000	.018	.140	.884	
Claim 1								
Below	312,526	.454	.330	.120	.004	.000	.991	.990
At/Near	233,889	.339	.010	.169	.136	.024		
Above	142,630	.207	.000	.003	.042	.162	.987	
Claim 2/4								
Below	249,557	.362	.289	.060	.010	.004	.963	.969
At/Near	325,500	.472	.071	.206	.148	.047		
Above	113,988	.165	.000	.003	.031	.132	.984	
Claim 3								
Below	223,541	.324	.265	.049	.008	.002	.969	.974
At/Near	347,880	.505	.100	.200	.144	.061		
Above	117,624	.171	.000	.003	.027	.141	.984	
All Students	689,045	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.35 GRADE 7 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	232,356	.341	.302	.039	.000	.000	.885	.819
Level 2	201,779	.296	.038	.222	.035	.000	.751	
Level 3	139,382	.205	.000	.029	.154	.021	.751	
Level 4	107,870	.158	.000	.000	.017	.142	.894	
Claim 1								
Below	293,866	.431	.322	.104	.004	.000	.989	.989
At/Near	233,086	.342	.016	.172	.139	.015		
Above	154,435	.227	.000	.002	.058	.166	.990	
Claim 2/4								
Below	213,689	.314	.258	.041	.009	.005	.956	.968
At/Near	336,069	.493	.109	.195	.151	.038		
Above	131,629	.193	.000	.002	.042	.149	.988	
Claim 3								
Below	121,879	.179	.153	.017	.006	.004	.948	.966
At/Near	436,269	.640	.195	.221	.171	.054		
Above	123,239	.181	.000	.003	.036	.142	.985	
All Students	681,387	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.36 GRADE 8 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	259,877	.382	.340	.041	.000	.000	.892	.817
Level 2	180,638	.265	.039	.190	.035	.000	.718	
Level 3	120,541	.177	.000	.028	.127	.022	.719	
Level 4	120,141	.176	.000	.000	.018	.159	.900	
Claim 1								
Below	302,823	.445	.356	.084	.005	.000	.989	.989
At/Near	225,109	.330	.025	.160	.126	.019		
Above	153,265	.225	.000	.002	.046	.177	.991	
Claim 2/4								
Below	197,283	.290	.246	.032	.008	.004	.959	.971
At/Near	349,939	.514	.142	.179	.143	.051		
Above	133,975	.197	.000	.002	.035	.159	.987	
Claim 3								
Below	218,426	.321	.274	.040	.006	.001	.976	.980
At/Near	340,693	.500	.115	.183	.140	.062		
Above	122,078	.179	.000	.002	.027	.150	.987	
All Students	681,197	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

TABLE 2.37 GRADE 11 MATHEMATICS CLASSIFICATION ACCURACY

Assigned Level	N	P	Expected Proportion				Correct Classification	
			1	2	3	4	by level ^(a)	overall ^(b)
Overall								
Level 1	251,419	.451	.404	.047	.000	.000	.895	.827
Level 2	140,955	.253	.039	.180	.034	.000	.711	
Level 3	103,889	.186	.000	.025	.145	.015	.779	
Level 4	61,123	.110	.000	.000	.012	.098	.894	
Claim 1								
Below	274,399	.492	.401	.085	.006	.000	.988	.989
At/Near	181,865	.326	.031	.162	.127	.006		
Above	101,122	.181	.000	.002	.058	.121	.990	
Claim 2/4								
Below	188,839	.339	.288	.036	.011	.004	.956	.965
At/Near	283,689	.509	.158	.174	.149	.028		
Above	84,858	.152	.000	.002	.044	.106	.985	
Claim 3								
Below	167,608	.301	.257	.033	.009	.002	.965	.971
At/Near	308,821	.554	.167	.183	.161	.044		
Above	80,957	.145	.000	.002	.034	.109	.984	
All Students	557,386	1.000						

Notes: (a) Correct classification rate by level is the expected proportion of students among those assigned to a particular level who are correctly assigned; (b) overall correct classification rate is the expected proportion of students among those assigned to any level (excluding the "At/Near" classification) who are correctly assigned.

Standard Errors of Measurement

The reliability of reported test scores can be characterized by the standard errors associated with the students' test scores. The standard errors of measurement (SEM), the inverse of the square root of information, are related to reliability in that they represent the standard deviation of repeated test scores.

The conditional standard errors of measurement (CSEM) express the degree of measurement error in scale score units and are conditioned on the ability of the student. When using a computer-adaptive assessment, the CSEM will vary for the same scale score; therefore, it is necessary to report averages.

Table 2.38 presents the overall SEM and the CSEM by scale score decile for ELA/literacy and mathematics. This table shows that the CSEM is relatively similar between deciles 2 and 10. The CSEM tends to be higher at the first decile.

TABLE 2.38 OVERALL SEM AND CONDITIONAL STANDARD ERRORS OF MEASUREMENT (CSEM) BY DECILE, ELA/LITERACY AND MATHEMATICS 2014-2015 ADMINISTRATION

Content Area	Grade	Overall SEM	1	2	3	4	5	6	7	8	9	10
ELA/Literacy	3	24.7	32.9	26.2	24.4	23.6	23.0	22.6	22.4	22.3	22.7	24.7
	4	26.4	33.0	27.4	26.1	25.4	24.9	24.6	24.5	24.5	24.7	27.3
	5	26.0	32.0	25.7	24.7	24.3	24.4	24.6	24.6	24.7	25.2	28.0
	6	27.6	36.0	28.6	26.7	26.0	25.7	25.6	25.7	25.8	26.1	28.1
	7	27.8	35.7	29.0	27.2	26.5	26.1	25.7	25.5	25.5	26.1	29.3
	8	27.5	33.7	28.3	27.0	26.4	25.9	25.7	25.9	26.2	26.6	28.6
	11	30.9	39.3	33.2	30.8	29.3	28.4	28.0	28.0	28.3	29.3	32.4
Mathematics	3	19.5	28.9	20.9	19.4	18.4	17.7	17.2	16.8	16.6	16.8	19.1
	4	19.7	28.4	21.6	19.8	18.7	17.9	17.3	16.9	16.8	16.8	19.4
	5	23.7	37.2	28.5	25.4	23.1	21.4	20.1	19.0	18.3	17.8	19.2
	6	27.7	49.4	31.5	27.3	24.9	23.5	22.3	21.4	20.7	20.3	22.7
	7	32.4	61.9	38.1	32.9	29.7	27.2	24.9	22.8	21.2	19.9	22.3
	8	33.2	55.6	39.1	35.7	33.4	31.0	28.4	25.7	23.4	21.8	23.7
	11	40.5	73.2	51.1	44.0	39.2	35.2	31.9	28.9	26.2	23.7	23.8

Table 2.39 and Table 2.40 show the average CSEM near the achievement level cut scores.

TABLE 2.39 CONDITIONAL STANDARD ERRORS OF MEASUREMENT NEAR (± 10 POINTS) OF ACHIEVEMENT LEVEL CUT SCORES, GRADES 3-8 & 11 ELA/LITERACY 2014-2015 ADMINISTRATION

Grade	cut 1/2			cut 2/3			cut 3/4		
	N	M	SD	N	M	SD	N	M	SD
3	56747	23.69	1.5	59873	22.48	1.2	47513	22.41	1.1
4	51264	25.38	1.4	54759	24.59	1.1	45104	24.43	1.1
5	49211	24.38	1.1	55718	24.58	1.0	44430	24.87	1.0
6	47367	26.6	1.5	60458	25.59	1.3	38822	26.07	1.4
7	44471	26.87	1.4	54000	25.74	1.3	35845	26.21	1.3
8	44346	27.06	1.3	52998	25.72	1.2	35833	26.68	1.2
11	26770	32.12	1.6	40795	28.46	1.4	40651	28.39	1.2

TABLE 2.40 CONDITIONAL STANDARD ERRORS OF MEASUREMENT NEAR (± 10 POINTS) OF ACHIEVEMENT LEVEL CUT SCORES, GRADES 3-8 & 11 MATHEMATICS 2014-2015 ADMINISTRATION

Grade	cut 1/2			cut 2/3			cut 3/4		
	N	M	SD	N	M	SD	N	M	SD
3	66917	18.8	1.1	74678	17.1	0.9	47464	16.7	0.9
4	61304	19.5	1.0	67388	17	0.9	39812	16.8	0.9
5	59128	22.8	1.1	54641	18.9	1.0	42302	17.9	1.0
6	51372	25	1.2	55523	21.4	1.1	37205	20.3	1.1
7	46992	29.9	1.6	49491	23	1.4	33623	19.9	1.3
8	45905	32.6	1.9	43162	25.8	1.6	31595	22.1	1.3
11	35540	35.1	2.1	31453	27.4	2.3	16923	22.6	1.8

Figure 2.2 to Figure 2.15 below plot the CSEM for the overall and claim scale scores for Grades 3 through 8 and 11 for ELA/literacy and mathematics. Together with Table 2.39 and Table 2.40, the figures show that the CSEM tends to minimize around cut scores 2/3 and 3/4.

FIGURE 2.2 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 3 ELA/LITERACY 2014-2015 ADMINISTRATION

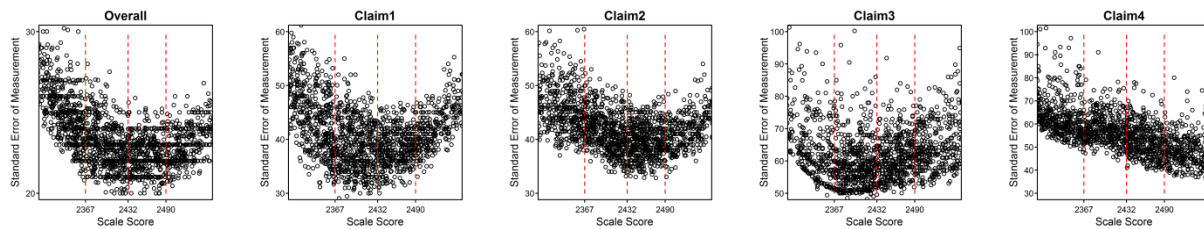


FIGURE 2.3 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 4 ELA/LITERACY 2014-2015 ADMINISTRATION

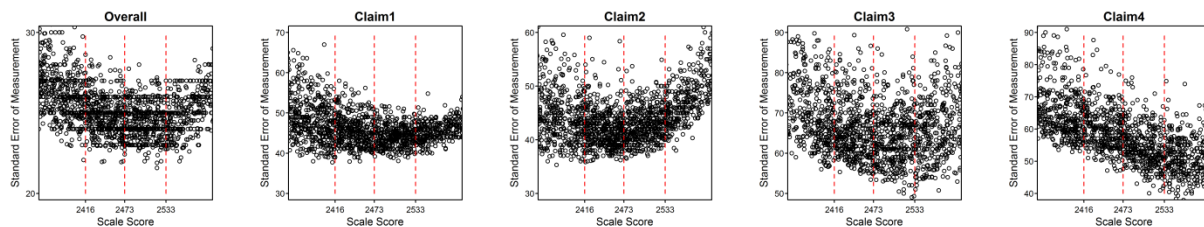


FIGURE 2.4 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 5 ELA/LITERACY 2014-2015 ADMINISTRATION

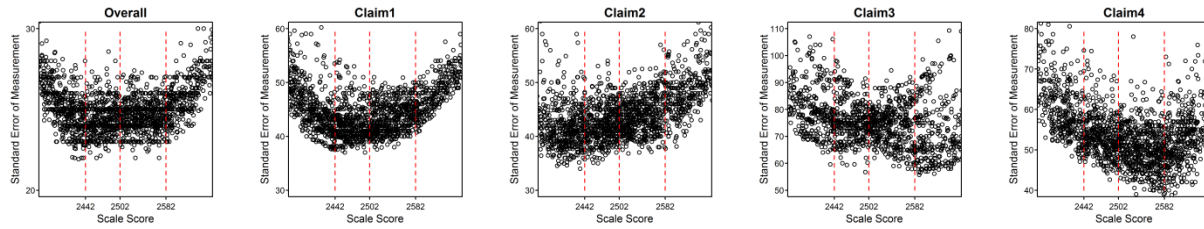


FIGURE 2.5 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 6 ELA/LITERACY 2014-2015 ADMINISTRATION

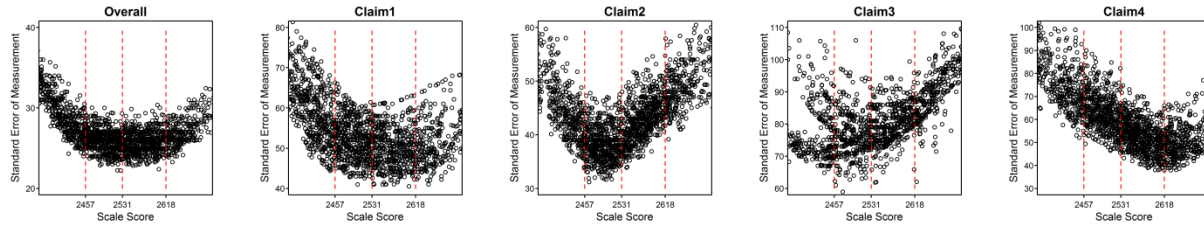


FIGURE 2.6 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 7 ELA/LITERACY 2014-2015 ADMINISTRATION

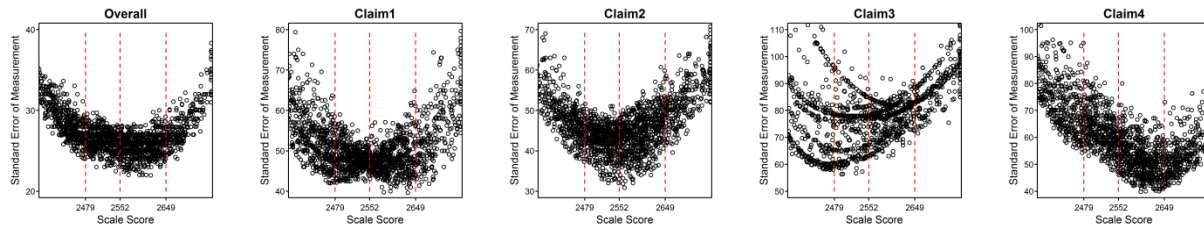


FIGURE 2.7 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 8 ELA/LITERACY 2014-2015 ADMINISTRATION

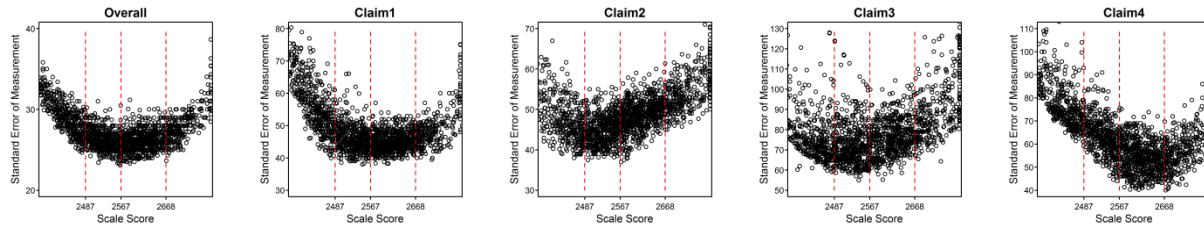


FIGURE 2.8 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 11 ELA/LITERACY 2014-2015 ADMINISTRATION

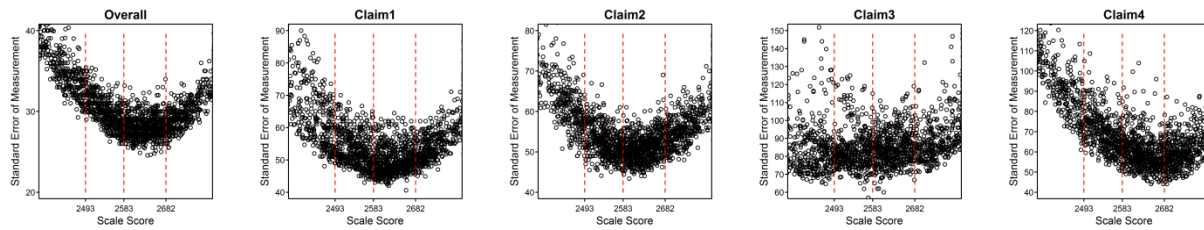


FIGURE 2.9 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 3 MATHEMATICS 2014-2015 ADMINISTRATION

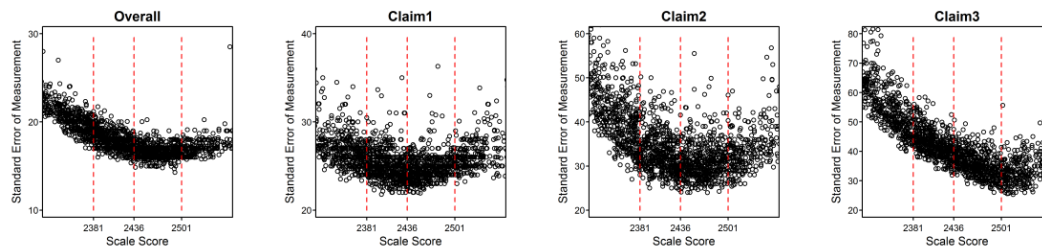


FIGURE 2.10 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 4 MATHEMATICS 2014-2015 ADMINISTRATION

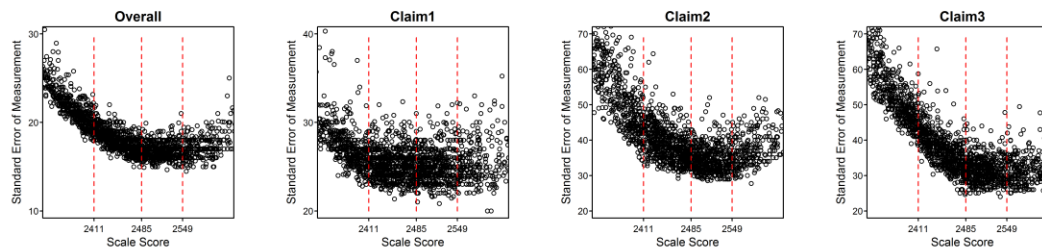


FIGURE 2.11 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 5 MATHEMATICS 2014-2015 ADMINISTRATION

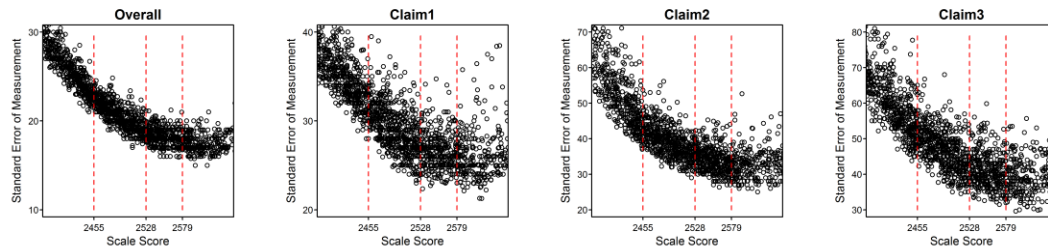


FIGURE 2.12 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 6 MATHEMATICS 2014-2015 ADMINISTRATION

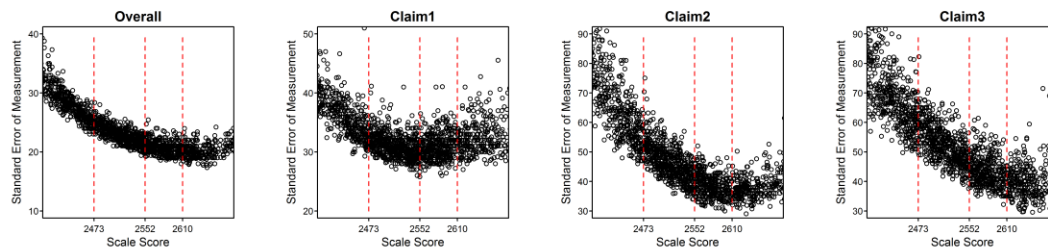


FIGURE 2.13 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 7 MATHEMATICS 2014-2015 ADMINISTRATION

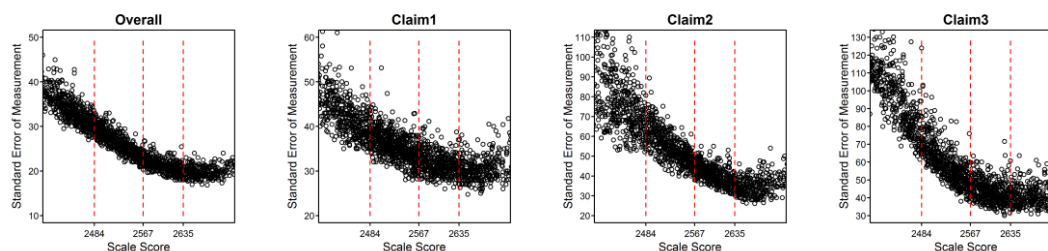


FIGURE 2.14 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 8 MATHEMATICS 2014-2015 ADMINISTRATION

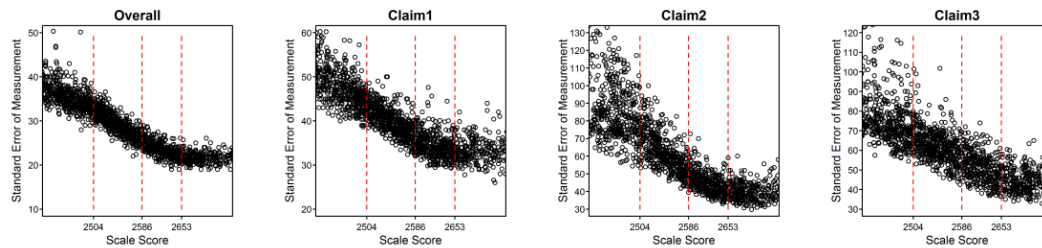
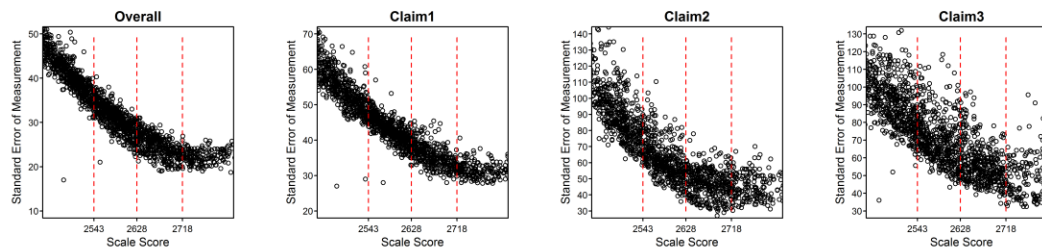


FIGURE 2.15 CONDITIONAL STANDARD ERRORS OF MEASUREMENT FOR OVERALL AND CLAIM SCALE SCORES, GRADE 11 MATHEMATICS 2014-2015 ADMINISTRATION



Online Platform Effects

Test are delivered through a variety of configurations based on equipment available to students. Smarter Balanced makes every effort to ensure comparability in assessment results regardless of the device available to students. For online assessments, Smarter Balanced is currently able to detect whether students used a device with a mouse (desktop/laptop) or used a tablet. The analysis of the 2014-15 operational data for these different modes is described below.

CRESST analyzed operational assessment data⁴ to evaluate platform effects for students who used either a desktop/laptop or tablet. For this analysis, calibrations of the items within each online platform (desktop/laptop or tablet) were compared against previous estimates (i.e., the item parameter values used in operational scoring). These calibrations were performed separately by platform.

As an initial step, CRESST screened all multiple choice items in the pool (for a given grade level and subject) as potential anchors, as it was expected that these items would be the least impacted by platform. Items were tested for invariance across the platforms so long as there were at least 500 observed scores for the particular platform. If the number of observed item scores exceeded 10,000, we randomly sampled 10,000 cases.

⁴ Data for platform effect analysis provided by California.

Likelihood ratio tests provided a formal evaluation of the null hypothesis that platform-specific item parameters were exactly equal to the previously specified values. However, this test tends to be quite sensitive (rejecting the null hypothesis for the majority of items). Accordingly, the weighted absolute area between expected score curves (wABC; see, e.g., Stucky, Edelen, & Chandra, 2015) was used as our primary criterion for judging the severity of differential item functioning across platforms.

Multiple choice items with $wABC > 0.10$ were rejected as anchor items, while those with $wABC \leq 0.10$ were retained. In all calibrations, the latent variable mean and variance were freely estimated (because the group of individuals administered a particular item cannot be assumed to be representative of the population, due to adaptive item selection). Results of this screening of candidate anchors are summarized in Tables 3 and 4. The vast majority of multiple choice items tested were retained as anchors.

After screening the multiple choice items to obtain a final set of anchors, CRESST proceeded with the calibration of all non-anchor items (i.e., the multiple choice items rejected as anchors, as well as items of any other type) for which a minimum of 500 scores were available. Due to differences in platform use, more items could be tested (calibrated) for desktop/laptop than for tablet. Results of the analyses are summarized in Tables 3 (ELA/literacy) and 4 (Mathematics) below. Not all items in the pools could be evaluated. However, among those that were tested, the vast majority of items had very small wABC values, suggesting minimal differences in item functioning across the platforms. Specifically, across grades, subjects, and platforms, less than 1% of items showed $wABC > 0.20$, except for grade 11 ELA/literacy tablets (for which 3% of the items had $wABC > 0.20$). These results suggest that items functioned similarly across platforms.

TABLE 2.41 SCREENING OF CANDIDATE ANCHOR ITEMS FOR EXAMINATION OF PLATFORM EFFECT IN ELA/LITERACY, BY GRADE AND PLATFORM

Platform	# MC items	# items tested	# anchors retained	LRT $p < .01$		weighted absolute area between expected score curves (wABC)											
						Mean	SD	(.00, .05)		(.05, .10)		(.10, .15)		(.15, .20)		(.20,1.00)	
				N	P			N	P	N	P	N	P	N	P	N	P
Grade 3 ELA/Literacy																	
Desktop/Laptop	283	275	256	243	.884	.033	.027	230	.846	32	.118	10	.037	0	.000	0	.000
Tablet		191	177	139	.728	.038	.031	147	.778	30	.159	12	.063	0	.000	0	.000
Grade 4 ELA/Literacy																	
Desktop/Laptop	258	258	231	210	.814	.034	.032	199	.790	35	.139	17	.067	1	.004	0	.000
Tablet		198	179	142	.717	.039	.033	139	.724	40	.208	12	.063	1	.005	0	.000
Grade 5 ELA/Literacy																	
Desktop/Laptop	224	220	204	188	.855	.033	.029	180	.837	25	.116	8	.037	2	.009	0	.000
Tablet		152	137	118	.776	.039	.033	120	.805	17	.114	9	.060	3	.020	0	.000
Grade 6 ELA/Literacy																	
Desktop/Laptop	224	213	203	169	.793	.031	.026	181	.858	22	.104	7	.033	1	.005	0	.000
Tablet		137	131	110	.803	.035	.028	108	.794	23	.169	5	.037	0	.000	0	.000
Grade 7 ELA/Literacy																	
Desktop/Laptop	191	183	180	138	.754	.025	.018	169	.929	11	.060	2	.011	0	.000	0	.000
Tablet		107	105	85	.794	.025	.020	94	.887	11	.104	1	.009	0	.000	0	.000
Grade 8 ELA/Literacy																	
Desktop/Laptop	215	214	206	158	.738	.025	.028	188	.883	19	.089	3	.014	3	.014	0	.000
Tablet		138	130	110	.797	.025	.032	116	.847	14	.102	5	.036	2	.015	0	.000
Grade 11 ELA/Literacy																	
Desktop/Laptop	567	377	326	326	.865	.025	.047	259	.712	71	.195	14	.038	14	.038	6	.016
Tablet		162	125	107	.660	.025	.047	85	.578	40	.272	12	.082	8	.054	2	.014

Notes: "LRT $p < .01$ " indicates the number (N) and proportion (P) of items with p-values $< .01$. The likelihood ratio tests and expected score curves (the basis of the weighted area between the curves) were based on comparisons of models in which an item's parameters are freely estimated and a second, nested model in which the item's parameters were fixed to their prior estimates (the scoring parameters). Items with $wABC > 0.10$ or for which calibrations did not converge were rejected as anchors for analysis of platform differences.

TABLE 2.42 SCREENING OF CANDIDATE ANCHOR ITEMS FOR EXAMINATION OF PLATFORM EFFECT IN MATHEMATICS, BY GRADE AND PLATFORM

Platform	# MC items	# items tested	# anchors retained	LRT p < .01		weighted absolute area between expected score curves (wABC)											
						Mean	SD	(.00, .05)		(.05, .10)		(.10, .15)		(.15, .20)		(.20,1.00)	
				N	P			N	P	N	P	N	P	N	P	N	P
Grade 3 ELA/Literacy																	
Desktop/Laptop	140	122	122	102	.836	.036	.019	94	.770	28	.230	0	.000	0	.000	0	.000
Tablet		59	58	49	.831	.040	.019	40	.690	18	.310	0	.000	0	.000	0	.000
Grade 4 ELA/Literacy																	
Desktop/Laptop	154	152	152	119	.783	.027	.018	137	.901	15	.099	0	.000	0	.000	0	.000
Tablet		99	99	65	.657	.036	.022	75	.758	24	.242	0	.000	0	.000	0	.000
Grade 5 ELA/Literacy																	
Desktop/Laptop	203	205	203	179	.873	.025	.015	186	.916	17	.084	0	.000	0	.000	0	.000
Tablet		168	164	117	.696	.030	.024	144	.873	20	.121	0	.000	0	.000	1	.006
Grade 6 ELA/Literacy																	
Desktop/Laptop	88	92	91	83	.902	.027	.013	87	.956	4	.044	0	.000	0	.000	0	.000
Tablet		63	60	41	.651	.029	.017	53	.883	7	.117	0	.000	0	.000	0	.000
Grade 7 ELA/Literacy																	
Desktop/Laptop	79	83	83	61	.735	.023	.012	82	.988	1	.012	0	.000	0	.000	0	.000
Tablet		42	42	28	.667	.030	.017	39	.929	3	.071	0	.000	0	.000	0	.000
Grade 8 ELA/Literacy																	
Desktop/Laptop	123	129	129	101	.783	.023	.013	124	.961	5	.039	0	.000	0	.000	0	.000
Tablet		89	89	63	.708	.027	.015	81	.910	8	.090	0	.000	0	.000	0	.000
Grade 11 ELA/Literacy																	
Desktop/Laptop	486	353	314	264	.748	.034	.034	287	.854	27	.080	16	.048	5	.015	1	.003
Tablet		35	33	14	.400	.033	.022	26	.788	7	.212	0	.000	0	.000	0	.000

Notes: "LRT $p < .01$ " indicates the number (N) and proportion (P) of items with p-values $< .01$. The likelihood ratio tests and expected score curves (the basis of the weighted area between the curves) were based on comparisons of models in which an item's parameters are freely estimated and a second, nested model in which the item's parameters were fixed to their prior estimates (the scoring parameters). Items with $wABC > 0.10$ or for which calibrations did not converge were rejected as anchors for analysis of platform differences.

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/>
- 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. (2016). *Simulation-based evaluation of the 2014-2015 smarter balanced summative assessments: accommodated item pools*. Los Angeles, CA: Smarter Balanced Assessment Consortium. Retrieved from <http://www.smarterbalanced.org/assessments/development/additional-technical-documentation/>

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 Standards⁵ (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 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

⁵ <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 and sensitivity 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 *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.

The Smarter Balanced Accessibility and Accommodations Framework

Smarter Balanced has built a framework of accessibility for all students, including English Language Learners (ELLs), students with disabilities, and ELLs with disabilities, but not limited to those groups. Three resources—the Smarter Balanced *Item Specifications* (2015c), the Smarter Balanced *Usability, Accessibility, and Accommodations Guidelines* (2014b), 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 develop an *Accessibility and Accommodations Framework* (2014a) that guided the Consortium as it worked to reach agreement on the specific universal tools, designated supports, and accommodations available for the assessments. This work also incorporated research and practical lessons learned through Universal Design, accessibility tools, and accommodations (Thompson, Johnstone, & Thurlow, 2002).

In the process of developing its next-generation assessments to measure students' knowledge and skills as they progress toward college and career readiness, Smarter Balanced recognized that the validity of assessment results depends on each student having appropriate universal tools, designated supports, and/or accommodations when needed, based on the constructs being measured by the assessment. The Smarter Balanced Assessment System uses technology intended to deliver assessments that meet the needs of individual students. Online/electronic delivery of the assessments helps ensure that students are administered a test individualized to meet their needs while still measuring the same construct. During the administration of tests, items and tasks are delivered using a variety of accessibility resources and accommodations that can be administered to students automatically based on their individual profiles. Accessibility resources include but are not limited to foreground and background color flexibility, tactile presentation of content (e.g., braille), and translated presentation of assessment content in signed form and selected spoken languages.

One of Smarter Balanced's main goals was to adopt a common set of accessibility resources and accommodations. As a starting point, Smarter Balanced surveyed all members to determine their

past practices. From these data, Smarter Balanced used a deliberative analysis strategy as described in *Accommodations for English Language Learners and Students with Disabilities: A Research-Based Decision Algorithm* (Abedi & Ewers, 2013) to determine which accessibility resources should be made available during the assessment and whether access to these resources should be moderated by an adult. As a result, some accessibility resources that states traditionally had identified as accommodations, were instead embedded in the test or otherwise incorporated into the Smarter Balanced assessments as universal tools. Other resources were not incorporated into the assessment because access to these resources were not grounded in research. The final list of accessibility resources and the recommended use of the resources can be found in the *Usability Accessibility and Accommodations Guidelines* (2014b, pp. 6-20).

A fundamental goal was to design an assessment that is accessible for all students, regardless of English language proficiency, disability, or other individual circumstances. The three components of the *Accessibility and Accommodations Framework* are designed to meet that need. The intent was to ensure that the following steps were achieved for Smarter Balanced.

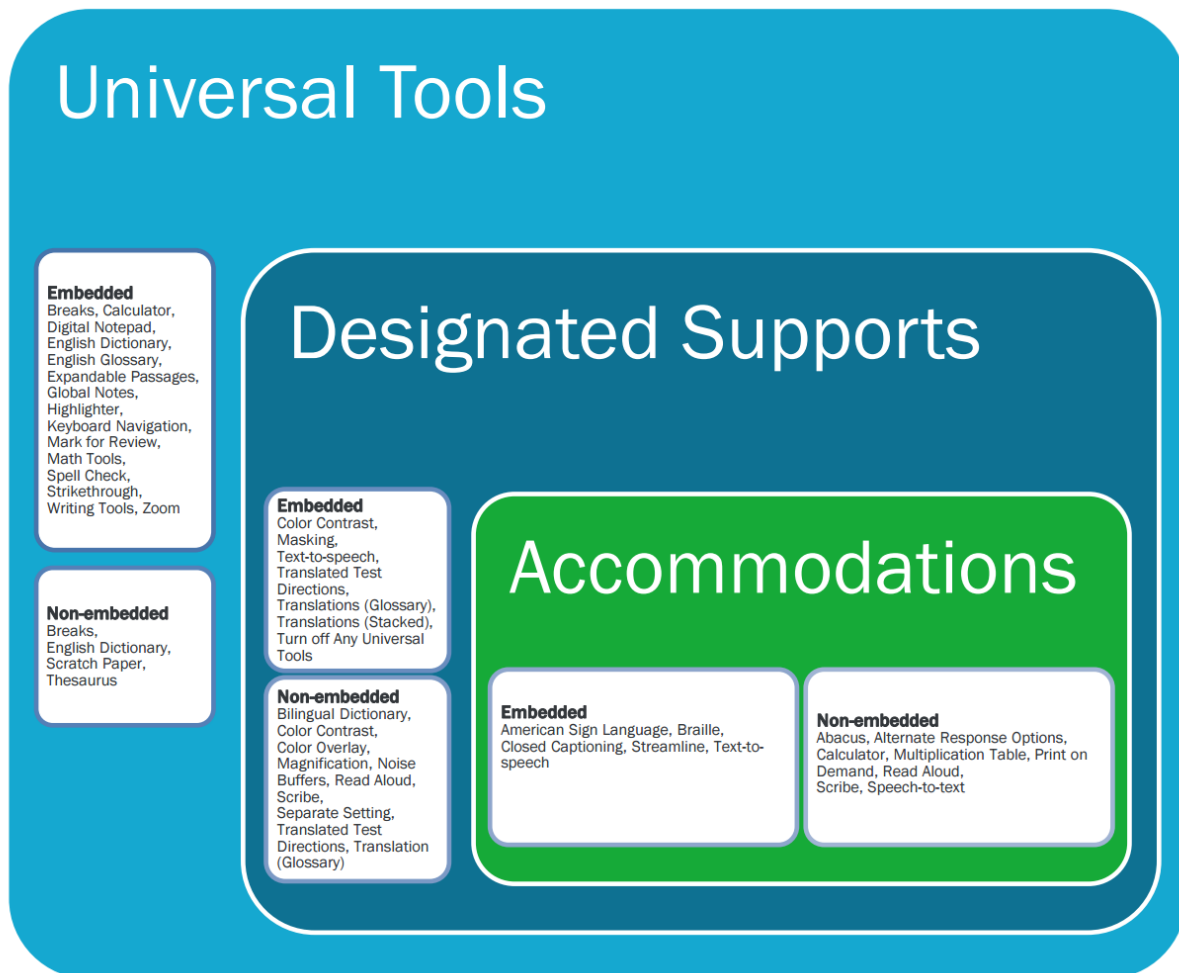
- Design and develop items and tasks to ensure that all students have access to the items and tasks designed to measure the targeted constructs. In addition, deliver items, tasks, and the collection of student responses in a way that maximizes validity for each student.
- Adopt the conceptual model embodied 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 from those available to all students ranging 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.

The conceptual framework that serves as the basis underlying the usability, accessibility, and accommodations is shown in 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 target and 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 3.1 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 separate from test delivery.

FIGURE 3.1 CONCEPTUAL MODEL UNDERLYING THE SMARTER BALANCED USABILITY, ACCESSIBILITY, AND ACCOMMODATIONS GUIDELINES. FROM *USABILITY ACCESSIBILITY AND ACCOMMODATIONS GUIDELINES* (P. 4), 2014.



Meeting the Needs of Traditionally Underrepresented Populations.

The policy decision was 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 ELLs, students with disabilities, and ELLs with disabilities is presented here.

How the Framework Meets Needs of Students Who Are ELLs.

Students who are ELLs have needs that are unique from those students with disabilities, including language-related disabilities. The needs of ELLs 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 ELLs 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 and an audio presentation of test navigation features, are available to all students, including those at various stages of English language development. Other tools, such as the audio presentation of items and 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 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 Tool⁶ 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

⁶ <http://52.11.155.96/static/isaap/index.html>

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 3.1 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 avoidance of 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 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: Intended Audience and Recommended Applications.

Smarter Balanced 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 (2014b, p. 2), all eligible students (including students with disabilities, ELLs, and ELLs 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 (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.

Guidelines for Accessibility for English Language Learners.

In addition to the use of Universal Design features, Smarter Balanced has built a framework of accessibility for all students, including English Language Learners (ELLs) that were established in the *Smarter Balanced Guidelines for Accessibility for English Language Learners* (Young, Pitoniak, King, & Ayad, 2012). ELLs have not yet acquired complete proficiency in English. For ELLs, the most significant accessibility issue concerns the nature of the language used in the assessments. 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 ELLs, 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 ELL 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 ELLs measure the intended targets. The ELL *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. ELLs may have had cultural and social experiences that differ from those of other students. Caution should be exercised in assuming that ELLs 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 ELLs. 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 ELLs 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 ELLs 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 ELLs 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, ELLs 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 ELLs 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.

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 in tests with a minimum of barriers. These accessibility resources are considered embedded accommodations or embedded designated supports. The specialized pools that were available in 2014-15 are shown in Table 3.1.

TABLE 3.1 SPECIALIZED TESTS AVAILABLE TO QUALIFYING STUDENTS IN 2014-15

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

Fairness as a Lack of Measurement Bias: Differential Item Functioning Analyses

As part of the validity evidence from internal structure, differential item functioning (DIF) analyses were conducted on items using data from the 2014 field test. 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. Fairness requires assuring that differences in response patterns to be valid. 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.

Differential Item Functioning (DIF) Analyses for the Calibration Item Pool

Differential item functioning (DIF) analyses were performed on the items during field testing. DIF analyses are used to identify those items that identify groups of students (e.g., males versus females) with the same underlying level of ability that 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 ranked relative to their total test score (conditioning on ability). Item performance from the focal group to be examined (e.g., females) is compared conditionally based on ability with the reference group (e.g., males). The definitions for the focal and references groups used are given in Table 3.2. 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?" DIF in this context is viewed as a potential source of invalidity.

DIF statistics are used to identify items that are functioning differentially. Subsequent reviews by content experts and bias/sensitivity committees are required to determine the source and meaning of performance differences. If the item is differentially more difficult for an identifiable subgroup when conditioned on ability, it 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 or statistical Type I error. Final decisions about the resolution of item DIF are made by a 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)	Absolute value of the MH D-DIF is not significantly different from zero, or is less than one.
B (slight to moderate)	Absolute value of the MH D-DIF is significantly different from zero but not from one, and is at least one; or Absolute value of the MH D-DIF is significantly different from one, but less than 1.5. Positive values are classified as “B+” and negative values as “B-”
C (moderate to large)	Absolute value of the MH D-DIF is significantly different from 1, and is at least 1.5; and Absolute value of the MH D-DIF is larger than 1.96 times the standard error of MH D-DIF. Positive values 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 \leq 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 slight or moderate DIF, and Category C items have moderate to large 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 were not conducted if the sample size for either the reference group or the focal group was less than 100 or if the sample size for the two combined groups was less than 400. In subsequent tables, 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 consisting of the Mantel-Haenszel procedure (Mantel & Haenszel, 1959) and the standardized mean difference (SMD) procedure (Dorans & Kulick, 1983, 1986) are used for dichotomous and polytomous items

respectively. Selected-response 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

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

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):

$$MH\ D-DIF = -2.35 \ln[\alpha_{MH}].$$

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

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

where $E(R_{rm}) = N_{rm} R_{Nm} / N_m$, $Var(R_{rm}) = \frac{N_{rm} N_{fm} R_{Nm} W_{Nm}}{N_m^2 (N_m - 1)}$, 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. The classification logic for selected-response items is based on a combination of absolute differences and significance testing, is shown in Figure 3.1.

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:

$$SMD = \frac{\sum p_{Fk} m_{Fk} - \sum p_{Rk} m_{Rk}}{\sqrt{\sum p_{Fk} m_{Fk}^2 - (\sum p_{Fk} m_{Fk})^2 + \sum p_{Rk} m_{Rk}^2 - (\sum p_{Rk} m_{Rk})^2}}$$

where p_{Fk} is the proportion of the focal group members who are at the k^{th} level of the matching variable, m_{Fk} is the mean score for the focal group at the k^{th} level, and m_{Rk} is the mean item score for the reference group at the k^{th} level. 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 3.4. 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.

A relatively small number of items showed some performance differences between student groups as indicated by C-DIF flagging criteria. Table 3.5 and Table 3.6 show the number of items flagged for all categories of DIF for ELA/literacy and mathematics in grades 3 – 8 and 11. A relatively small percentage of items with moderate or significant levels of DIF (B or C DIF) were included in summative pools. All items had previously undergone bias reviews. 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 3.5 NUMBER OF DIF ITEMS IN SUMMATIVE POOLS FLAGGED BY CATEGORY (ELA, 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	1	1	183	77	3	502	83	38
3	A	618	604	435	543	630	122	544	599
3	B-	4	15	9	8	3	6	5	0
3	B+	10	12	10	8	1	7	5	0
3	C-	0	1	0	1	0	0	0	0
3	C+	4	4	0	0	0	0	0	0
4	N/A	2	2	165	97	8	492	71	31
4	A	600	592	452	523	623	125	549	600
4	B-	10	14	10	13	3	11	11	4
4	B+	16	20	7	2	2	8	4	1
4	C-	1	2	1	1	0	0	1	0
4	C+	7	6	1	0	0	0	0	0
5	N/A	0	0	0	0	0	0	0	0
5	A	573	590	606	605	624	610	608	623
5	B-	16	19	11	13	2	11	8	3
5	B+	21	14	6	6	1	3	7	1
5	C-	2	0	1	2	0	3	2	0
5	C+	15	4	3	1	0	0	2	0
6	N/A	0	0	0	0	0	0	0	0
6	A	565	559	583	574	591	586	584	591
6	B-	10	14	7	10	2	5	7	3
6	B+	16	15	6	9	4	5	4	3
6	C-	2	5	1	4	0	1	2	0
6	C+	4	4	0	0	0	0	0	0
7	N/A	0	0	0	0	0	0	0	0
7	A	525	535	556	561	566	569	564	574
7	B-	22	16	9	11	4	2	7	0
7	B+	21	15	7	3	5	4	4	1
7	C-	2	1	2	0	0	0	0	0
7	C+	5	8	1	0	0	0	0	0
8	N/A	0	0	0	0	0	0	0	0
8	A	523	540	558	556	579	568	560	576
8	B-	17	18	12	15	1	10	12	5
8	B+	25	12	8	5	1	2	9	0
8	C-	4	3	2	2	0	1	0	0
8	C+	12	8	1	3	0	0	0	0
11	N/A	0	0	0	0	0	0	0	0
11	A	1499	1468	1519	1511	1541	1523	1520	1539
11	B-	33	34	11	22	2	8	14	5
11	B+	11	38	14	13	4	14	11	4
11	C-	5	3	2	1	1	1	2	0
11	C+	0	5	2	1	0	2	1	0

TABLE 3.6 NUMBER OF DIF ITEMS IN SUMMATIVE POOLS FLAGGED BY CATEGORY (MATHEMATICS, GRADES 3-8 AND 11)

Grade	DIF Category	Focal group/Referent Group							
		Female/Male	Asian/White	Black/White	Hispanic/White	NativeAmerican/White	IEP/Non-IEP	LEP/Non-LEP	Title1/Non-Title1
3	N/A	0	178	18	0	894	2	0	0
3	A	936	692	856	895	55	928	912	949
3	B-	3	19	22	16	0	9	10	0
3	B+	10	38	49	37	1	8	23	1
3	C-	0	4	0	0	0	1	1	0
3	C+	1	19	5	2	0	2	4	0
4	N/A	0	144	102	0	737	16	1	0
4	A	902	704	783	886	178	897	881	920
4	B-	8	17	9	10	1	9	10	0
4	B+	13	42	23	24	7	1	26	4
4	C-	0	3	2	1	0	1	4	0
4	C+	1	14	5	3	1	0	2	0
5	N/A	0	157	76	0	642	1	22	0
5	A	879	674	783	888	240	875	851	895
5	B-	6	18	12	5	5	5	8	1
5	B+	13	32	27	4	7	14	12	2
5	C-	0	3	0	1	1	2	5	0
5	C+	0	14	0	0	3	1	0	0
6	N/A	0	100	173	0	816	52	37	0
6	A	802	658	636	808	14	773	774	823
6	B-	3	10	4	2	0	1	3	1
6	B+	21	30	11	15	0	4	16	6
6	C-	2	4	2	1	0	0	0	0
6	C+	2	28	4	4	0	0	0	0
7	N/A	0	118	67	0	733	48	58	0
7	A	734	552	665	734	16	682	676	745
7	B-	8	10	1	3	0	5	1	2
7	B+	7	43	13	12	0	13	13	2
7	C-	0	4	1	0	0	0	0	0
7	C+	0	22	2	0	0	1	1	0
8	N/A	0	219	122	0	712	64	184	0
8	A	722	462	585	714	16	643	528	728
8	B-	3	13	8	4	0	4	7	0
8	B+	3	22	10	9	0	12	7	0
8	C-	0	3	1	1	0	1	0	0
8	C+	0	9	2	0	0	4	2	0
11	N/A	0	1073	382	0	1678	1186	1372	0
11	A	1630	556	1261	1636	14	485	302	1671
11	B-	14	7	17	13	0	3	2	2
11	B+	37	29	26	38	0	17	13	16
11	C-	5	0	3	0	0	0	1	0
11	C+	6	27	3	5	0	1	2	3

Note: In February, 2016, the Consortium's Technical Advisory Committee suggested that these criteria may be too conservative. Based on this advice, the Consortium may loosen its thresholds for determining DIF. Any such change will be reported in subsequent technical manuals.

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 ELLs. 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 <http://www.smarterbalanced.org/wp-content/uploads/2015/08/Accommodations-for-under-represented-students.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-web.ad.education.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
- 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 http://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA_Content_Specs.pdf
- Smarter Balanced Assessment Consortium. (2015b). *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. (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.

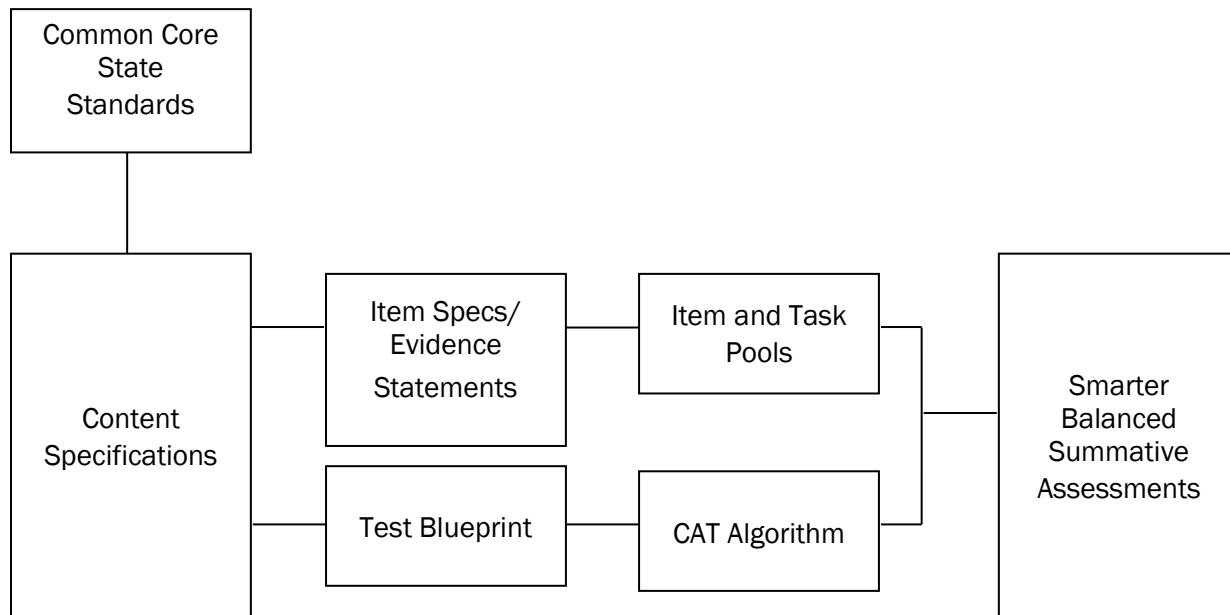
Chapter 4: Test Design



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 a system of assessments, the goal of Smarter Balanced was to ensure that its measurement properties reflected the expectations of content, rigor, and performance that comprise the Common Core State Standards (CCSS). The primary mechanism for this was to ensure the alignment of the Smarter Balanced assessments with the 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) 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, distilling 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 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 Claim 3 is assessed. In mathematics, Claims 2 and 4 are reported together, 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 statement 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, 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.

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 3–8 and high school—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 and content constraints embedded in 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 and the summative test blueprints reflect refinements and revisions after the analysis of the Pilot and Field Tests.

Synopsis of Assessment System Components

The summative assessment for each content area consists of two parts: a CAT and a 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 inferences 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 conception 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. The layers focus in turn on the identification of the substantive domain to be assessed; the assessment argument; the structure of assessment elements such as tasks, rubrics, and psychometric models; the implementation of these elements; and the way they function in an operational assessment, as described below. For Smarter Balanced, a subset of the general ECD elements was used.

- **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 or weighting of different content strata (e.g., claims) if there is any weighting beyond the proportions of items and points;
- 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.

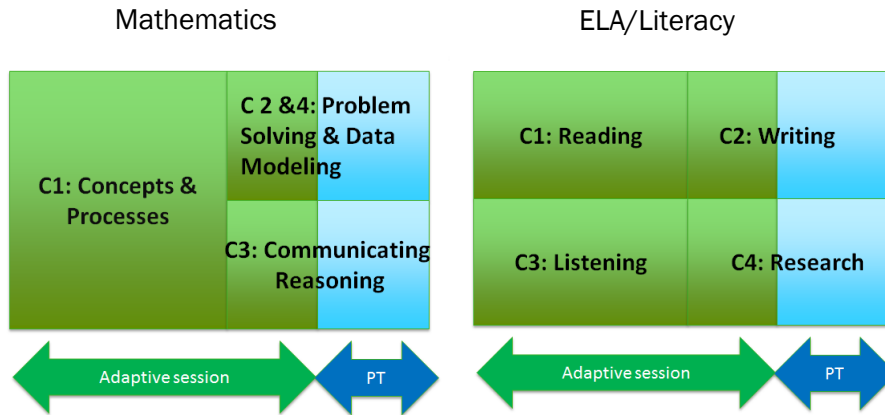
The test blueprint is an essential guide for both assessment developers and for curriculum and instruction. For assessment developers, the blueprint and related test-specification documents define how the test will ensure coverage of the full breadth and depth of content and how it will maintain fidelity 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 for test developers and signals to the broader education community both the full complexity of the CCSS and how performance on these standards are 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. These were fully implemented in the 2014-15 school year. The complete blueprints details for each grade and content area (Smarter Balanced, 2015a; Smarter Balanced, 2015b).

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 blueprint overviews summarize the claim score\reporting category, content category, stimuli used, items by CAT or performance tasks, and total number of items by claim. 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). 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 the IRT target information function. In establishing target requirements for the CAT, 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 targets in a domain area is large, blueprints allow choice within target clusters rather than limiting the number of targets. Since all targets are represented in the pool, any student could 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. Figure 4.4 and

Figure 4.5 present blueprint requirements for grade six mathematics, by claim and assessment target. It displays the number of items overall by claim and shows the contribution of the CAT and performance task portions to the overall design. 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 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. Here, the DOK requirements are applied at the target level. Performance tasks are delivered as a fixed set of items within a theme common to a class or school.

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

¹ 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/smarter-balanced-assessments/>.

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.

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.	1, 2, 3	1		
		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).				
	Modeling and Data Analysis (drawn across content domains)	A. Apply mathematics to solve problems arising in everyday life, society, and the workplace.	2, 3	1	1–3	
		D. Interpret results in the context of a situation.	2, 3, 4	1		
		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.	1, 2, 3	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).				
G. Identify, analyze, and synthesize relevant external resources to pose or solve problems.	3, 4	0				
3. Communicating Reasoning	Communicating Reasoning (drawn across content domains)	A. Test propositions or conjectures with specific examples.	2, 3	3	0-2	8-10
		D. Use the technique of breaking an argument into cases.				
		B. Construct, autonomously, chains of reasoning that will justify or refute propositions or conjectures.	2, 3, 4	3		
		E. Distinguish correct logic or reasoning from that which is flawed, and—if there is a flaw in the argument—explain what it is.				
		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.)				

- 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 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. Several performance tasks are associated with a common theme. A theme is assigned to school for each grade and the performance tasks within the theme are randomly distributed within the grade.

Adaptive Test Design and Algorithm Overview

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 (Luecht, 1998).

For the operational test, an item-level, fully adaptive test component was administered in ELA/literacy and mathematics. The adaptive part delivers blueprints in a manner that efficiently minimizes measurement error and maximizes information. Smarter Balanced provides a specific CAT delivery engine, but states may choose to use other engines as long as they can deliver a conforming test blueprint with a minimum degree of error, avoid item over- or under-exposure, and provide the design features specified by Smarter Balanced. This section outlines some of the design features for the operational adaptive test component.

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. 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.

This section describes the adaptive algorithm design for the Smarter Balanced Test Delivery System. To assert the administration of comparable tests, 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, provided that in concert, they can deliver the published blueprint as expected.

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.

In general, an adaptive algorithm is the 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. All items in the bank must have complete information about these elements available to algorithm software. 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 used in the selection process that affects test delivery. By allowing for the specification of weights, the general algorithm can be customized for varying conditions of population and pool distribution. This function can help assure that a test best matches the purpose for which it is designed. For example, weights can be shifted to emphasize measurement precision or content coverage, depending on policy priorities. Final weights were established during the last stages of test design when all item parameters were known and simulation results were 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 functional 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.

Test Operation Walkthrough

Preparation: The 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 contain only items written for the intended grade. Under certain circumstances (described below) the filter is dropped and the adjacent grade items are added. 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.

Item selection. The initialization and selection processes control underuse and overuse of items, also known as exposure control. Exposure control enhances item security, discouraging 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 initial item response has been given, 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, 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, the student ability estimate is updated. The selection procedure is then repeated until the blueprint has been satisfied. 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.

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.

The algorithm proceeds in this manner until a percentage of the test (coverage in mathematics, 61%; ELA, 62%) has been administered, sampling items from all claim areas. At this point the distance of the estimated score from the college content readiness cut score is evaluated. This is Level 3 as defined in the Smarter Balanced *Achievement Level Setting Final Report* (2015f). If there is a determination that the student is in either Level 1 or Level 4 as defined by the Achievement Level Setting Report, the item pool is expanded to include items from no more than two adjacent grades in either direction. In grade 3, the expansion includes items from adjacent upper grades only; in grade 11 only adjacent lower grades are included. Items from adjacent grades have been screened for appropriateness by content experts to assure that they are instructionally and developmentally appropriate for 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. When a determination of being in Level 1 or level 4 cannot be made, the test continues with on-grade items. The algorithm delivers the remainder of the blueprint until termination.

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 level 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 content proportions 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 span from the content specifications and Achievement Level Descriptors (ALDs) to the assessment itself. While the content specifications established the Consortium's claims and the types of evidence or targets, that would need to be collected in order to support these claims, more specificity was needed in order 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 review 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, they 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 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 included on test questions in 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.

This is important because it underscores 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 (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 necessitates 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, 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 took 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 used an iterative process for creating and revising each item as well as the collection of items. 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 played 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 worked with educators throughout the test development cycle to develop items. 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. All K-12 participants:

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

Qualifications for Higher Education Faculty included:

- 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;
- Completing training and achieving qualifications through the certification process.

The selected educators were trained on the Consortium's content specifications, the item and task specifications, and ELA/literacy stimulus specifications, as well as the item authoring system in which the items were developed. In addition, professional item writers and the Consortium held regular meetings to provide direction and feedback to the educators. Educators, state partners, and assessment vendors developed the items in the Consortium's item pool.

Training

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.

The item writers were 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 helped ensure that item writers were trained to write items that allowed the widest possible range of students to demonstrate their knowledge, skills, and cognitive processes with regard to the content. This meant that item writers needed 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 were also trained to be aware of issues that might unintentionally bias an item for or against a particular group.

Educator Participation

Consistent with the Consortium process, educators were the primary developers of items. The active involvement of educators was 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 helped ensure that the assessment system is accurate and efficient, and provides valid evidence of student learning.

State-Managed Item Development

The Consortium invited member states to participate in a separate effort to write items. This voluntary effort, known as State-Managed Item Development, was 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. These items were not operational in the Smarter Balanced 2015 summative assessments but were included as embedded field test items.

Item Reviews

Once items were written, groups of educators reviewed items prior to their pilot test administration in spring 2013 and their field test administration in spring 2014. Items that met technical quality criteria from the pilot test were again reviewed prior to their use in the spring 2014 field test.

Accessibility, Bias/Sensitivity, and Content Reviews

Panels of educators reviewed 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 the figures and graphics used on the mathematics assessments. Prior to the spring 2013 pilot test, 122 ELA/L educators and 106 mathematics educators reviewed items and performance tasks for accessibility, bias/sensitivity, or content, and 60 educators reviewed the ELA/L stimuli. Prior to the spring 2014 field test, 107 ELA/L educators and 157 mathematics educators from 14 states reviewed items and performance, and 95 educators from 13 states reviewed the ELA/L stimuli.

The educator qualifications for the accessibility, bias/sensitivity, and content reviews were the same as the educator qualifications for item writing except that participants were not required to submit a statement of interest. In addition, it was preferred (but not required) that educators have previous experience reviewing items, tasks, and/or stimuli.

During the accessibility reviews, panelists identified issues that may negatively affect a student's ability to access stimuli, items, or performance tasks, or to elicit valid evidence about an assessment target. During the bias and sensitivity review, panelists identified content in stimuli, items, or performance tasks that may negatively affect a student's ability to produce a correct response because of their background. The content review focused 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 checked the accuracy of the content, answer keys, and scoring materials. Items flagged for accessibility, bias/sensitivity, and/or content concerns were either revised to address the issues identified by the panelists 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

When all of the developmental reviews had been conducted, items that passed data review became part of the 2014 Field Test. Details of the 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.

Item Scoring

For those items that could not be machine scored, the Consortium engaged 102 participants from 20 states in range finding activities for those items requiring human scoring following the spring 2013 pilot. After the spring 2014 field test, 104 educators participated in range finding. Range finding improves the consistency and validity of scoring for the assessment. During range finding, educators focused on the performance tasks for mathematics and ELA/literacy. The participants reviewed student responses against item rubrics, validated the rubrics' accuracy, and selected the anchor papers that would be used by scorers during operational scoring of test items. In mathematics, educators also reviewed constructed response items for grades 7, 8, and high school.

The educator qualifications for range finding were the same as the educator qualifications for item writing, except that participants were not required to submit a statement of interest. In addition, it was preferred (but not required) that educators had previous range finding experience.

To verify correct scoring for machine-scored items, a rubric validation activity was 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.

Review of Item Data from Field Testing

The items developed for the 2015 operational item pool were administered during the spring 2014 field test, and data were analyzed to examine the statistical quality of the items. The Consortium established statistical criteria to flag items for possible defects in quality related to content, bias, or accessibility. Criteria that triggered item review are in Table 4.1. In addition, items with C-level of differential item functioning (DIF) were flagged for further review. (Details of DIF criteria can be found in Chapter 3 of this report.)

Following the spring 2014 field test, 57 ELA/literacy educators from 16 states and 30 mathematics educators from 12 states reviewed items with statistical flags, looking for possible content or bias issues. At least two educators reviewed each item. These educators were trained via webinars on the flagging criteria and on how to evaluate flagged items. Educators made recommendations on whether to accept the item with no change, revise and re-field test the item, or reject the item from the pool. McGraw-Hill CTB content experts reviewed all items where the reviewers' recommendations disagreed. In addition, McGraw-Hill CTB content experts and psychometricians also reviewed and provided recommendations for all items where both reviewers recommended accepting the item. In each situation, the content expert provided the Consortium with a final recommendation for the item.

The educator qualifications for the item data reviews were 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 CLASSICAL STATISTICS AND JUDGEMENTAL REVIEW

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 (p -value less than 0.30)
V	Item more difficult at the higher-grade level for vertical linking items
Z	Item needs content review (judgmental decision)

Items with no statistical flags were eligible for use in the operational pools. Flagged items moved into operational pools if they were not rejected or revised in data review. Not all operational items were used in summative test pools. Some items were used in achievement levels setting. Those and additional items were used for the interim assessments. Table 4.3 shows how summative pools were derived from the original field test pool.

TABLE 4.3 DISPENSATION OF FIELD TEST ITEMS

Content Area	Grade	Initial Field Test Pool	Reason for Pool Removal				Summative Pool
			Content Issues	Small Sample Size	Poor Item Statistics	Interim/ Other uses	
English Language Arts	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
Mathematics	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

Composition of Summative Item Pools

The numbers of items in each summative content area, grade and claim are shown in Table 4.4.

TABLE 4.4 COMPOSITION OF SUMMATIVE ITEM POOLS

	CLAIMS				
GRADE	1	2	3	4	Total
ELA/Literacy					
3	217	194	118	125	654
4	177	205	127	144	653
5	194	201	108	142	645
6	175	199	116	123	613
7	183	194	117	95	589
8	161	190	131	112	594
11	499	429	334	300	1562
Total	1606	1612	1051	1041	5310
Mathematics					
3	551	130	160	116	957
4	525	131	149	127	932
5	462	117	182	152	913
6	519	107	137	91	854
7	452	100	125	90	767
8	425	81	142	87	735
11	1022	196	460	192	1870
Total	3956	862	1355	855	7028

The Consortium developed many different types of items beyond the traditional multiple-choice item. This was 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 4.5. Distribution of item types is shown in Table 4.6 and Table 4.7. 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 4.5 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 4.6 DISTRIBUTION OF ELA/LITERACY ITEM TYPES BY GRADE AND CLAIM

ELA/Literacy									
Gr	Claim	Item Type							Total
		EBSR	HTQ	MC	MI	MS	SA	WER	
3	1	33	35	106		29	14		217
3	2	0	34	82		38	12	28	194
3	3	32		47	12	27			118
3	4	0	10	48	8	34	25		125
3	Total	65	79	283	20	128	51	28	654
4	1	31	26	67		32	21		177
4	2	0	37	85		30	15	38	205
4	3	30		57	12	28			127
4	4	0	11	49	9	39	36		144
4	Total	61	74	258	21	129	72	38	653
5	1	36	30	69		34	25		194
5	2	0	35	67		42	17	40	201
5	3	29		43	13	23			108
5	4	0	13	45	11	35	38		142
5	Total	65	78	224	24	134	80	40	645
6	1	28	37	59		33	18		175
6	2	0	39	74		42	16	28	199
6	3	32		51	13	20			116
6	4	0	8	40	6	42	27		123
6	Total	60	84	224	19	137	61	28	613
7	1	30	42	59		37	15		183
7	2	0	39	63		40	14	38	194
7	3	33		51	9	24			117
7	4	0	19	18	4	17	37		95
7	Total	63	100	191	13	118	66	38	589
8	1	30	34	47		29	21		161
8	2	0	34	56		40	18	42	190
8	3	17		90	3	21			131
8	4	0	19	22	8	22	41		112
8	Total	47	87	215	11	112	80	42	594
11	1	92	131	114		113	49		499
11	2	0	94	141		119	27	48	429
11	3	59		191	12	72			334
11	4	0	54	121	14	63	48		300
11	Total	151	279	567	26	367	124	48	1562
All Grades	Total	512	781	1962	134	1125	534	262	5310

TABLE 4.7 DISTRIBUTION OF MATHEMATICS ITEM TYPES BY GRADE AND CLAIM

Mathematics									
Gr	Claim	Item Type							Total
		EQ	GI	MC	MI	MS	SA	TI	
3	1	356	43	70	49	1		32	551
3	2	69	33	9	7	4	5	3	130
3	3	5	55	37	16	19	28		160
3	4	34	19	24	5	10	11	13	116
3	Total	464	150	140	77	34	44	48	957
4	1	278	60	64	115	0		8	525
4	2	75	15	29	6	3	1	2	131
4	3	14	59	20	7	20	28	1	149
4	4	30	20	41	5	8	16	7	127
4	Total	397	154	154	133	31	45	18	932
5	1	252	34	134	42	0			462
5	2	80	17	8	2	5		5	117
5	3	19	68	41	14	12	27	1	182
5	4	61	34	20	4	3	15	15	152
5	Total	412	153	203	62	20	42	21	913
6	1	268	54	50	60	85		2	519
6	2	76	14	2	2	6	2	5	107
6	3	12	38	30	16	8	33		137
6	4	45	9	6	2	4	13	12	91
6	Total	401	115	88	80	103	48	19	854
7	1	250	18	49	46	89			452
7	2	72	7	5	6	8		2	100
7	3	16	40	17	8	11	33		125
7	4	50	24	8	2	3	1	2	90
7	Total	388	89	79	62	111	34	4	767
8	1	200	53	97	26	49			425
8	2	44	16	4	5	1		11	81
8	3	16	47	12	16	17	34		142
8	4	35	22	10	5	5	7	3	87
8	Total	295	138	123	52	72	41	14	735
11	1	275	194	283	178	89		3	1022
11	2	89	46	25	13	13	1	9	196
11	3	42	153	134	55	32	44		460
11	4	83	29	45	15	8	7	5	192
11	Total	489	422	487	261	142	52	17	1870
All Grades	Total	2846	1221	1274	727	513	306	141	7028

Each grade's item pool for the Consortium's test was large enough to support the summative blueprint. Unlike a traditional paper-and-pencil test where all students take the same items, students taking the Consortium's CAT take items and tasks targeted to their ability level. This means that the Consortium needed to develop a very large number of items in order to meet the needs of the student population.

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.

TABLE 4.8 NUMBER OF PERFORMANCE TASKS BY GRADE

Grade	ELA	Math
3	14	18
4	19	19
5	20	15
6	14	18
7	19	16
8	21	18
11	24	17

The distribution of item parameters by grade and claim are shown below. 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 4.9 ITEM DIFFICULTY (B-PARAMETER) AND DISCRIMINATION (A-PARAMETER), ELA/LITERACY

ELA/LITERACY						
Grade	Claim	# of items	b-parameter			a-parameter
			Mean	Min	Max	Mean
3	1	217	-0.444	-2.725	2.743	0.72
	2	194	-0.671	-2.896	2.810	0.68
	3	118	-0.011	-2.283	3.815	0.56
	4	125	-0.071	-2.027	3.032	0.69
	Total	654	-0.362	-2.896	3.815	0.67
4	1	177	0.298	-2.101	3.133	0.64
	2	205	-0.224	-3.252	2.935	0.60
	3	127	0.136	-2.822	4.254	0.55
	4	144	0.591	-1.761	3.727	0.58
	Total	653	0.167	-3.252	4.254	0.60
5	1	194	0.669	-1.604	4.806	0.65
	2	201	0.208	-2.535	4.954	0.64
	3	108	0.676	-2.401	3.481	0.52
	4	142	0.741	-1.494	3.832	0.64
	Total	645	0.542	-2.535	4.954	0.62
6	1	175	1.053	-1.203	4.779	0.59
	2	199	0.818	-2.719	4.607	0.57
	3	116	1.026	-1.447	4.921	0.50
	4	123	1.198	-0.929	3.609	0.61
	Total	613	1.001	-2.719	4.921	0.57
7	1	183	1.165	-1.877	3.914	0.58
	2	194	0.885	-1.979	5.124	0.61
	3	117	0.869	-1.706	4.775	0.49
	4	95	1.793	-0.449	5.525	0.60
	Total	589	1.115	-1.979	5.525	0.57
8	1	161	1.490	-1.170	5.572	0.59
	2	190	1.019	-3.013	4.558	0.58
	3	131	0.974	-1.535	4.266	0.47
	4	112	1.868	-0.669	5.188	0.59
	Total	594	1.297	-3.013	5.572	0.56
11	1	499	1.843	-1.340	5.567	0.57
	2	429	1.612	-1.880	5.929	0.47
	3	334	1.304	-1.247	5.618	0.45
	4	300	2.024	-0.270	5.124	0.51
	Total	1562	1.699	-1.880	5.929	0.50

TABLE 4.10 ITEM DIFFICULTY (B-PARAMETER) AND DISCRIMINATION (A-PARAMETER), MATHEMATICS

MATHEMATICS						
Grade	Claim	# of items	b-parameter			a-parameter
			Mean	Min	Max	Mean
3	1	551	-1.146	-3.381	2.402	0.85
	2	130	-0.449	-2.537	1.967	0.99
	3	160	-0.200	-2.324	3.464	0.79
	4	116	-0.194	-2.677	1.818	0.84
	Total	957	-0.778	-3.381	3.464	0.86
4	1	525	-0.275	-3.260	4.113	0.83
	2	131	-0.042	-1.897	2.574	0.89
	3	149	0.380	-1.950	3.157	0.79
	4	127	0.354	-1.320	2.219	0.70
	Total	932	-0.052	-3.260	4.113	0.82
5	1	462	0.369	-2.526	3.606	0.77
	2	117	0.928	-1.147	3.409	0.93
	3	182	0.996	-1.219	5.278	0.71
	4	152	1.259	-0.991	4.452	0.74
	Total	913	0.714	-2.526	5.278	0.77
6	1	519	0.854	-3.934	4.347	0.68
	2	107	1.020	-2.978	5.099	0.82
	3	137	1.458	-1.727	4.709	0.63
	4	91	1.383	-0.410	3.894	0.78
	Total	854	1.028	-3.934	5.099	0.70
7	1	452	1.664	-1.792	5.643	0.71
	2	100	1.591	-1.085	5.071	0.85
	3	125	2.108	-1.345	6.174	0.67
	4	90	1.996	-0.924	4.373	0.81
	Total	767	1.766	-1.792	6.174	0.73
8	1	425	2.084	-1.542	6.321	0.62
	2	81	2.551	0.046	5.751	0.79
	3	142	2.597	-0.878	6.698	0.57
	4	87	2.229	-0.656	5.354	0.69
	Total	735	2.252	-1.542	6.698	0.64
11	1	1022	2.466	-3.364	7.297	0.55
	2	196	2.921	-1.101	6.680	0.62
	3	460	2.927	-1.793	7.194	0.47
	4	192	3.187	-0.069	6.379	0.54
	Total	1870	2.701	-3.364	7.297	0.54

Although there is a wide distribution of item difficulty, pools tend to be difficult in relation to the population and to proficiency cut scores (the cut between levels 2 and 3). The charts below show mean item difficulty, proficiency cut scores and mean student scores (all in theta units).

FIGURE 4.6 COMPARISON OF ITEM DIFFICULTY, MEAN, STUDENT SCORES, CUT SCORES FOR ELA/LITERACY

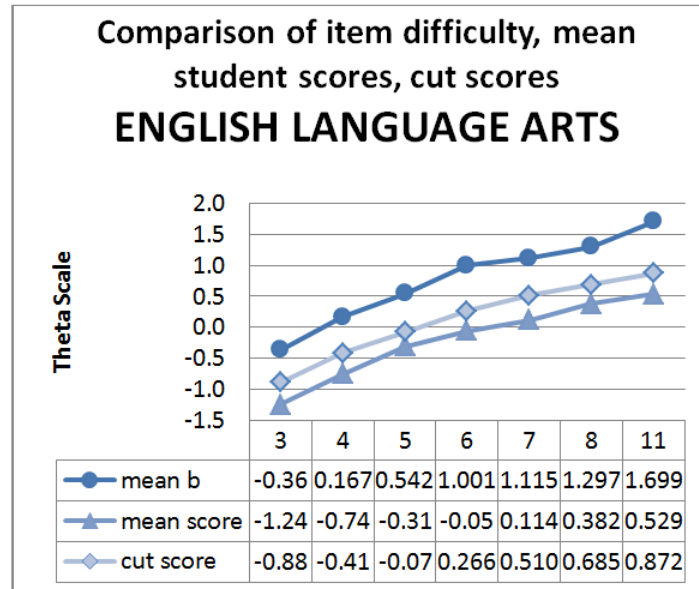
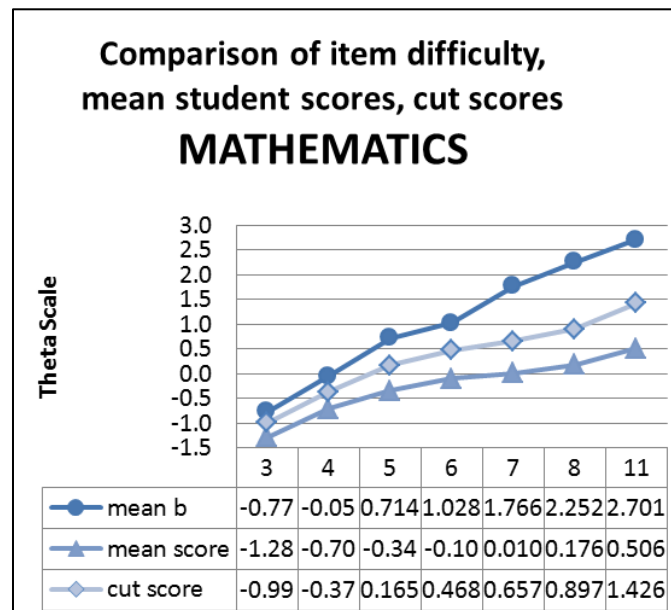


FIGURE 4.7 COMPARISON OF ITEM DIFFICULTY, MEAN, STUDENT SCORES, CUT SCORES FOR MATHEMATICS



Content Alignment

In developing a system of assessments, Smarter Balanced is committed to ensuring that its measurement reflects the expectations of content, rigor, and performance that correspond to the CCSS. To that end, Smarter Balanced designed item specifications to demonstrate alignment through methodologies that reflect ECD theory. According to Webb (2002), “Alignment of expectations for student learning and assessments for measuring students’ attainment of these expectations is an essential attribute for an effective standards-based education system.” DeMauro (2004) states, “Alignment activities . . . should be the guiding principle of test design, and item alignment studies should be sources of validity documentation, as should any studies of test content.” Test content alignment is at the core of content validity and consequential validity (Martone & Sireci, 2009). Content alignment addresses the appropriateness of inferences drawn from test results concerning “how well all policy elements [e.g., expectations and assessments] guide instruction and, ultimately, impact student learning” (Webb, 1997). Since Consortium states have adopted the CCSS in ELA/literacy and mathematics, it was imperative that Smarter Balanced conduct the appropriate alignment studies. Accordingly, the Consortium contracted with the Human Resources Research Organization to conduct an alignment study (HumRRO, 2014).

Webb (1997) identified several categories of criteria for judging content alignment. The Smarter Balanced alignment study describes how well the Smarter Balanced tests address expectations embodied in the Smarter Balanced content specifications and the CCSS. Test content alignment is at the core of content validity and consequential validity (Martone and Sireci, 2009). Because of the high stakes associated with statewide testing and the need to communicate learning goals during the NCLB era, attention was directed at test alignment in addition to individual item alignment. The emphasis on test content in alignment and validity studies is understandable. After all, a test is a small sampling of items from a much larger universe of possible items/tasks representing a very broad domain. For inferences from test results to be justifiable, that sample of items has to be an adequate representation of the broad domain, providing strong evidence to support claims based on the test results.

Assessment is always constrained to some extent by time and resources. Items and tasks that require extensive time (performance tasks and text responses), items that require expensive scoring, and items that require a lot of computer bandwidth (videos, animations) must be limited and chosen carefully. Smarter Balanced content experts carefully scrutinized each blueprint to assure optimal content coverage and prudent use of time and resources. In general, the Smarter Balanced blueprints represent content sampling proportions that reflect intended emphasis in instruction and mastery at each grade level. Specifications for numbers of items by claim, assessment target, depth-of-knowledge, and item type demonstrate the desired proportions within test delivery constraints. The blueprints were subject to state approval through a formal vote.

The alignment study conducted for the Consortium (HumRRO, 2014) discusses alignment among elements of content standards, content specifications, item specifications, and blueprints. The study itself is extensive, but its overall finding is that the Smarter Balanced summative tests and supporting item pools exceed levels of DOK representation recommended by Webb. The analysis is done with test blueprint, item and test specifications and item pools.

Fordham/HumRRO study

Extensive alignment studies were conducted in grades 5 and 8 by the Thomas B. Fordham Institute (Doorey & Polikoff, 2016) and in high school by the Human Resources Research Organization (HumRRO) (Schultz, Michaels, Dvorak, & Wiley 2016). Their evaluation was based on elements of the *Criteria for Procuring and Evaluating High Quality Assessments* developed by the Council of Chief State School Officers (CCSSO) and released in 2014. The study involved looking at items from simulated test events to evaluate operation of the test as experienced by students. Reviewers felt that the items were adequately aligned in almost all aspects, with many excellent ratings. Reviewers commented on the quality of the tasks with regard to the types of skills required for college and career readiness. Results of the study are shown below.

ELA/LITERACY	Ratings:	3-8	HS
I. CONTENT: Assesses the content most needed for College and Career Readiness		Excellent	Excellent
B.3 Reading1: Tests require students to read closely and use specific evidence from texts to obtain and defend correct responses.		Excellent	Excellent
B.5 Writing1: Tasks require students to engage in close reading and analysis of texts. Across each grade band, tests include a balance of expository, persuasive/argument, and narrative writing.		Excellent	Excellent
B.6 Vocabulary and language skills: Tests place sufficient emphasis on academic vocabulary and language conventions as used in real-world activities.		Good	Excellent
B.7 Research and inquiry: Assessments require students to demonstrate the ability to find, process, synthesize, and organize information from multiple sources.		Excellent	Excellent
B.8 Speaking and listening: Over time, and as assessment advances allow, the assessments measure speaking and listening communication skills.		Limited/ Uneven	Good
II. DEPTH: Assesses depth that reflects the demands of College and Career Readiness		Good	Excellent
B.1 Text quality and types1: Tests include an aligned balance of high-quality literary and informational texts.		Excellent	Excellent
B.2 Complexity of texts1, 3: Test passages are at appropriate levels of text complexity, increasing through the grades, and multiple forms of authentic, high-quality texts are used.		Good	Good
B.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.		Good	Excellent
B.9 High-quality items and variety of item types: Items are of high technical and editorial quality and test forms include at least two item types with at least one that requires students to generate a response.		Good	Excellent

In ELA/literacy, reviewers gave a limited/uneven match rating for B.8 because Smarter Balanced tests assess only listening and not speaking at this time.

MATHEMATICS	3-8	HS
I. CONTENT: Assesses the content most needed for College and Career Readiness	Good	Excellent
C.1 Focus1: Tests focus strongly on the content most needed in each grade or course for success in later mathematics (i.e., major work).	Good	Excellent
C.2 Concepts, procedures, and applications: Assessments place balanced emphasis on the measurement of conceptual understanding, fluency and procedural skill, and the application of mathematics.	N/A	Good
II. DEPTH: Assesses the depth that reflects the demands of College and Career Readiness	Good	Excellent
C.3 Connecting practice to content1, 5: Test questions meaningfully connect mathematical practices and processes with mathematical content.	Excellent	Excellent
C.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.	Good	Excellent
C.5 High-quality items and variety of item types: Items are of high technical and editorial quality and test forms include at least two item types, at least one that requires students to generate a response.	Limited/ Uneven	Good

Math C.5 received a limited/uneven match rating in grades 3-8 because reviewers encountered a few items with minor editorial issues. This has prompted an editorial review of the existing items.

Pool analysis and adequacy: Background and Recommendations

The quality of a CAT is highly dependent on the quality of the item pool. Quality is primarily related to how well the content constraints and statistical criteria can be met. The content specifications are defined as a combination of item attributes that tests delivered to students should have. There are typically constraints on item content such that they must conform to coverage of a test blueprint. If there are many content constraints and a limited pool, then it will be difficult to meet the CAT specifications. For a given content target, if the available difficulty/item information targeted at a given level ability is not available, then estimation error cannot be reduced efficiently. A third dimension is that there is usually some need to monitor the exposure of items such that the “best” items are not administered at high rates relative to other ones. Therefore, the quality of the item pools is critical to achieving the benefits that accrue for the CAT over fixed test forms. Quantification of pool adequacy prior to simulation could be accomplished either through the Reckase (2003) “bin” method or the van der Linden (2005) “shadow test” method. Both involve an inventory of items by required blueprint elements and information ranges. Smarter Balanced used the Reckase “bin” method to evaluate the pool and provide information for new item development. In general, the proportions of items in the pool were written to reflect test blueprints. Although item developers

strove to develop items covering the range of examinee achievement levels, the item pool is relatively difficult as compared to the performance that students displayed on the tests.

FIGURE 4.8 GAP ANALYSIS OF 2014-15 SUMMATIVE POOLS FOR ELA/LITERACY

Grade Level	Score Reporting Category	Claim	Targets	% of Test Blueprint*	# of 2014-15 ELA Operational Items	Difficulty				
						1	2	3	4	5
3	1	1	1 thru 14	0.40	217	5	30	38	51	93
	2	2	1,3,6,8,9	0.25	166	14	43	26	19	64
	3	3	4	0.20	118	2	12	17	16	71
	4	4	2,3,4	0.15	91	0	11	19	15	46
4	1	1	1 thru 14	0.40	177	5	25	26	29	92
	2	2	1,3,6,8,9	0.25	167	19	31	28	34	55
	3	3	4	0.20	127	10	18	21	21	57
	4	4	2,3,4	0.15	97	2	8	19	12	56
5	1	1	1 thru 14	0.40	194	10	28	24	32	100
	2	2	1,3,6,8,9	0.25	161	17	31	19	31	63
	3	3	4	0.20	108	3	11	14	26	54
	4	4	2,3,4	0.15	86	4	16	10	18	38
6	1	1	1 thru 14	0.44	175	3	29	19	24	100
	2	2	1,3,6,8,9	0.23	171	15	21	22	22	91
	3	3	4	0.19	116	8	15	15	22	56
	4	4	2,3,4	0.14	90	0	5	19	19	47
7	1	1	1 thru 14	0.44	183	10	20	26	28	99
	2	2	1,3,6,8,9	0.23	155	15	11	26	27	76
	3	3	4	0.19	117	8	15	25	20	49
	4	4	2,3,4	0.14	54	0	5	4	10	35
8	1	1	1 thru 14	0.44	161	6	23	17	22	93
	2	2	1,3,6,8,9	0.23	148	14	28	13	22	71
	3	3	4	0.19	131	12	25	24	19	51
	4	4	2,3,4	0.14	60	1	6	4	9	40
11	1	1	1 thru 14	0.43	499	12	45	60	80	302
	2	2	1,3,6,8,9	0.23	381	26	40	48	64	203
	3	3	4	0.20	334	27	53	62	57	135
	4	4	2,3,4	0.14	243	0	16	31	47	149

*percentage of blueprint needs

FIGURE 4.9 GAP ANALYSIS OF 2014-15 OPERATIONAL SUMMATIVE POOLS FOR MATHEMATICS

Grade Level	Score Reporting Category	Claim	Targets	% of Test Blueprint*	# of 2014-15 Math Operational Items	Difficulty				
						1	2	3	4	5
3	1	1	A thru K	0.59	547	88	141	95	100	138
	2 & 4	2	A thru D	0.09	76	3	3	8	18	44
	3	3	A thru F	0.24	123	1	6	11	26	79
	2 & 4	4	A thru F	0.09	83	2	8	4	14	55
4	1	1	A thru L	0.59	516	61	56	88	146	166
	2 & 4	2	A thru D	0.09	91	1	14	9	13	54
	3	3	A thru F	0.24	116	6	5	15	21	69
	2 & 4	4	A thru F	0.09	95	3	7	10	20	55
5	1	1	A thru K	0.59	459	12	52	74	148	173
	2 & 4	2	A thru D	0.09	81	0	1	9	15	56
	3	3	A thru F	0.24	146	0	8	17	39	82
	2 & 4	4	A thru F	0.09	121	0	2	7	13	99
6	1	1	A thru J	0.59	510	32	43	63	116	256
	2 & 4	2	A thru D	0.09	71	4	2	6	6	53
	3	3	A thru G	0.24	99	1	1	5	22	70
	2 & 4	4	A thru F	0.09	59	0	1	2	10	46
7	1	1	A thru I	0.59	452	9	11	32	76	324
	2 & 4	2	A thru D	0.09	67	0	2	3	8	54
	3	3	A thru G	0.24	97	1	1	6	12	77
	2 & 4	4	A thru F	0.09	54	0	0	1	8	45
8	1	1	A thru J	0.59	405	5	31	23	42	304
	2 & 4	2	A thru D	0.09	43	0	0	1	4	38
	3	3	A thru G	0.24	108	0	4	3	7	94
	2 & 4	4	A thru F	0.09	56	0	2	3	9	42
11	1	1	A thru P	0.59	979	38	41	71	133	696
	2 & 4	2	A thru D	0.09	159	2	1	4	25	127
	3	3	A thru G	0.24	408	5	7	28	49	319
	2 & 4	4	A thru F	0.09	167	0	2	5	19	141

*percentage of blueprint needs

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>
- Council of Chief State School Officers (2014). *Criteria for procuring and evaluating high quality assessments*.
- Dana, T. M., & Tippins, D. J. (1993). Considering alternative assessment for middle level learners. *Middle School Journal*, 25, 3-5.
- DeMauro, G. E. (2004). *Test alignment considerations for the meaning of testing*. Paper presented at the CCSSO Annual Conference on Large Scale Assessment, Boston, MA.
- Doorey, N., & Polikoff, M. (2016). *Evaluating the content and quality of next generation assessments*. Washington DC: Thomas B. Fordham Institute.
- 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.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79, 1-76.
- 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.
- Reckase, M. D. (2003). *Item pool design for computerized adaptive tests*. Paper presented at the National Council on Measurement in Education, Chicago, IL.
- 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.
- 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*. Alexandria, VA: HumRRO.
- 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.

- 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 http://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA_Blueprint.pdf
- Smarter Balanced Assessment Consortium. (2015b). *Mathematics summative assessment blueprint*. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/Mathematics_Blueprint.pdf
- 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 Assessment Consortium. (2016). 2013-2014 Technical Report. Retrieved from http://www.smarterbalanced.org/wp-content/uploads/2015/08/2013-14_Technical_Report.pdf
- van der Linden, W. J. (2005). *Linear models for optimal test design*. New York: Springer.
- Webb, N. L. (1997a, April). Criteria for alignment of expectations and assessments in mathematics and science education. Research Monograph No. 6. Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (March 28, 2002) *Depth-of-Knowledge Levels for Four Content Areas*, unpublished paper.
- 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 level descriptors help 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 2015 operational administration is shared at the end of 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. Table 5.1 summarizes the analyses completed during the pilot and field test phases that are related to calibration and scaling.

TABLE 5.1. ANALYSES COMPLETED DURING PILOT AND FIELD TEST TO SELECT CALIBRATION AND SCALING MODEL ⁷

Phase	Analysis	Summary
Pilot	Dimensionality	Multidimensional IRT was used as a factor analytic approach to examine the dimensional structure of the assessments. This 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, the ability estimates were compared across the three models. Based on the results of the studies and the considerations outline above, Smarter Balanced chose to employ the 2PL/GPC model.

⁷ 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 and the plots described previously. 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 linking tests horizontally within a grade level and content area, and a second linking tests vertically to adjacent grade levels within the content area. Following scaling of the vertical scaling sample, the remainder of the item pool was calibrated. The horizontal scale was created using a hybrid approach using both common items and randomly equivalent groups (implemented using LOFT administration). Items were calibrated using the IRT program PARSCALE. Consistent with the CCSS articulation of content through the grade levels, Smarter Balanced constructed a vertical scale to illustrate the amount of change that occurs in student learning from one grade level to the next. To create the scale, Smarter Balanced first evaluated the horizontal scaling of items in the targeted item pool. Once horizontal scaling was completed, vertical scaling was completed using common items. 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. After completion of the vertical scale for the targeted item pool, the remaining items were linked onto the scale for each content area 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

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

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

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

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 items and MLE estimation of 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 placing consecutive grades onto the vertical scale. In the second horizontal calibration step, the remaining, and much larger, item pool (containing non-common items, each administered only to one grade) was scaled using the items from the first phase as linking/common items. For detailed description of the methods used in vertical scaling, see Chapter 9 of the *2013-2014 Technical Report* (Smarter Balanced, 2016).

Transforming the Theta Metric to the Scale Score

The results from the calibration are in a theta metric. These results are transformed onto a four-digit scale that is more meaningful for stakeholders. The equation for this transformation is:

$$\text{Scale score} = (\text{logit} * \text{slope}) + \text{intercept}$$

Table 5.2 shows the slope and intercept for ELA/literacy and mathematics.

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 result in theta estimates for students with perfect or zero scores. Scale scores can be established for these extreme values following a non-maximum likelihood but logical procedure. These minimum and maximum values 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 $CSEM(HOSS) < 10 * \text{Minimum}(CSEMs \text{ for all scale scores})$, where CSEM is the conditional standard error of measurement. The LOSS should be high enough so that $CSEM(LOSS) < 15 * \text{Minimum}(CSEMs \text{ for all scale scores})$.
3. For multiple test levels placed on the same vertical scale, the HOSS and LOSS values should increase and transition smoothly over levels.

Table 5.3 provides the Smarter Balanced LOSS and HOSS values. 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	CSEM	HOSS	CSEM
ELA/literacy				
3	2114	2613	2623	2538
4	2131	2611	2663	2554
5	2201	2597	2701	2548
6	2210	2603	2724	2543
7	2258	2580	2745	2545
8	2288	2601	2769	2538
HS	2299	2594	2795	2549
Mathematics				
3	2189	2594	2621	2550
4	2204	2574	2659	2552
5	2219	2628	2700	2542
6	2235	2684	2748	2567
7	2250	2710	2778	2559
8	2265	2677	2802	2569
HS	2280	2675	2862	2566

Achievement-Level Setting

The Consortium used a multi-step process, called achievement level setting also known as standard setting, 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, 2015a).

Pre-Step: Development of the Achievement Level Descriptors

Smarter Balanced developed an interconnected system of achievement level descriptors (ALDs) for English language arts/literacy (ELA/literacy) and mathematics (CTB/McGraw Hill, 2014). These ALDs were developed through a series of workshops and review cycles that allowed participation from a variety of Smarter Balanced stakeholders. Key to the achievement level setting process, Smarter Balanced established Threshold ALDs that aligned to the Smarter Balanced content specifications and the Common Core State Standards that represent the knowledge and skills of a student at the borderline of a given achievement level. These Threshold ALDs provided panelists with a detailed description of the expectations of the students just entering each achievement level. The Threshold ALDs are important, because they ultimately guided the work of the achievement level setting process.

Step 1: Distributed Standard Setting (Online Panel)

Smarter Balanced selected the bookmark standard setting procedure (Lewis, Mitzel, Mercado, & Schultz, 2012) to set achievement levels because it is appropriate for assessments with a mixture of item types. The Smarter Balanced assessments are calibrated and scaled using item-response theory models, and the bookmark procedure utilizes those same 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), and its usefulness has been well established through adoption of cut scores produced by bookmark-based standard-setting activities.

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 score 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, a nationally-recognized expert in achievement level setting) 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

Results for the 2014-2015 Assessments

Results presented below are data aggregated across the Smarter Balanced members that submitted de-identified student results data for the 2014-2015 assessment⁸. The results in the tables in this chapter presented as evidence of reliability and validity of the scores from the Smarter Balanced assessments and should not be used for accountability purposes.

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 the assessment claims. The scale score quantifies student achievement, and the achievement-levels plainly define the meaning of the scores to stakeholders. Together, scale scores and achievement levels provide a comprehensive set of tools to assess student achievement by content and grade level. Table 5.5 provides the claims associated with the overall level for each content area.

⁸ Data for aggregated results provided by the following Consortium members: Delaware, Hawaii, Idaho, Maine, Oregon, South Dakota, US Virgin Islands, Vermont, Washington, West Virginia, California, Montana, Nevada, and North Dakota.

TABLE 5.5. 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 5.6 through Table 5.19 presents aggregate student results for the average overall scale score and for the percentage of students in each achievement level. These results are presented at the aggregate level (all submitted Consortium data) and disaggregated by gender, by race/ethnicity, and by various status flags: limited English proficiency, IDEA indicator, Section 504, and economically disadvantaged.

TABLE 5.6. GRADE 3 ELA/LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	776842	2412.9	87.1	31.5%	26.2%	21.7%	20.6%	42.3%
Female	379753	2422.9	86.2	27.2%	26.2%	23.1%	23.6%	46.7%
Male	397058	2403.4	86.8	35.7%	26.3%	20.3%	17.7%	38.0%
American Indian or Alaska Native	9536	2375.2	76.7	47.8%	29.0%	15.6%	7.7%	23.3%
Asian	55908	2458.7	86.4	15.8%	20.4%	24.8%	39.0%	63.8%
Black/African American	44474	2375.6	80.2	48.4%	26.5%	16.1%	9.0%	25.1%
Native Hawaiian or Pacific Islander	7516	2390.7	80.0	39.3%	29.9%	19.4%	11.4%	30.8%
Hispanic/Latino Ethnicity	275218	2380.1	78.5	45.0%	29.0%	16.9%	9.1%	26.0%
White	302323	2440.2	82.8	19.7%	24.4%	26.3%	29.6%	55.9%
Two or More Races	65339	2412.4	85.8	31.4%	27.2%	21.5%	19.9%	41.4%
Unidentified Race	16528	2441.5	80.6	18.3%	25.1%	27.2%	29.4%	56.6%
LEP Status	183892	2363.2	73.1	54.0%	28.3%	12.6%	5.1%	17.7%
IDEA Indicator	75412	2350.8	84.2	62.1%	20.1%	10.5%	7.3%	17.8%
Section 504 Status	5679	2415.7	85.8	29.7%	26.9%	22.6%	20.7%	43.3%
Economic Disadvantage Status	418082	2383.2	79.3	43.5%	28.9%	17.6%	10.0%	27.6%

TABLE 5.7. GRADE 4 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	738046	2454.9	92.0	34.8%	21.4%	22.3%	21.5%	43.8%
Female	361377	2466.0	90.7	30.2%	21.5%	23.4%	24.9%	48.3%
Male	376637	2444.2	92.0	39.2%	21.4%	21.2%	18.2%	39.4%
American Indian or Alaska Native	9311	2412.1	82.7	53.9%	22.0%	15.8%	8.2%	24.0%
Asian	55038	2506.8	90.8	17.0%	16.1%	24.2%	42.7%	66.9%
Black/African American	42197	2412.0	86.2	53.5%	21.2%	16.2%	9.2%	25.4%
Native Hawaiian or Pacific Islander	8334	2431.6	85.8	43.6%	23.6%	20.0%	12.8%	32.8%
Hispanic/Latino Ethnicity	252264	2419.8	83.3	49.4%	23.3%	17.6%	9.6%	27.2%
White	296265	2481.9	86.8	22.7%	20.7%	26.8%	29.9%	56.7%
Two or More Races	59189	2453.1	89.9	35.2%	22.4%	22.3%	20.1%	42.4%
Unidentified Race	15448	2486.3	84.5	21.0%	19.9%	27.9%	31.2%	59.1%
LEP Status	140655	2387.2	72.0	66.6%	21.0%	9.5%	2.8%	12.3%
IDEA Indicator	76464	2379.6	86.9	69.7%	14.6%	9.5%	6.1%	15.6%
Section 504 Status	6964	2455.8	87.6	34.4%	21.8%	23.4%	20.5%	43.9%
Economic Disadvantage Status	390792	2423.5	84.1	47.6%	23.4%	18.4%	10.5%	28.9%

TABLE 5.8. GRADE 5 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	761900	2495.3	93.0	30.0%	21.4%	29.5%	19.1%	48.6%
Female	372892	2508.9	90.9	24.7%	21.0%	31.4%	22.8%	54.2%
Male	388990	2482.3	93.2	35.0%	21.8%	27.7%	15.5%	43.2%
American Indian or Alaska Native	9809	2447.5	84.5	49.9%	23.6%	19.9%	6.6%	26.5%
Asian	58577	2547.4	92.3	14.3%	14.8%	31.4%	39.6%	71.0%
Black/African American	43949	2451.1	87.1	48.2%	22.7%	21.6%	7.5%	29.1%
Native Hawaiian or Pacific Islander	8589	2472.0	86.3	37.6%	24.2%	27.4%	10.8%	38.2%
Hispanic/Latino Ethnicity	259323	2461.2	85.1	43.0%	24.3%	24.2%	8.5%	32.7%
White	307315	2521.4	87.6	19.2%	19.7%	34.9%	26.2%	61.1%
Two or More Races	57938	2492.9	90.1	30.3%	22.7%	29.8%	17.2%	47.0%
Unidentified Race	16400	2527.3	86.2	16.8%	19.5%	35.1%	28.6%	63.7%
LEP Status	122514	2416.8	69.8	65.8%	22.4%	10.5%	1.4%	11.9%
IDEA Indicator	80620	2409.3	85.0	69.4%	15.7%	10.9%	4.0%	14.9%
Section 504 Status	8635	2497.4	87.9	28.2%	22.8%	31.4%	17.6%	49.0%
Economic Disadvantage Status	398182	2463.6	85.6	41.7%	24.2%	25.1%	8.9%	34.0%

TABLE 5.9. GRADE 6 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	751374	2517.6	91.3	25.9%	28.5%	31.1%	14.5%	45.6%
Female	367791	2531.6	88.7	20.4%	27.9%	34.1%	17.6%	51.7%
Male	383560	2504.1	91.7	31.2%	29.1%	28.2%	11.5%	39.7%
American Indian or Alaska Native	9148	2469.6	85.3	45.3%	30.4%	19.8%	4.5%	24.3%
Asian	57880	2571.7	90.1	11.6%	18.2%	36.3%	33.9%	70.2%
Black/African American	43562	2474.0	87.6	43.6%	29.5%	21.4%	5.4%	26.8%
Native Hawaiian or Pacific Islander	8161	2491.5	86.1	34.2%	31.9%	26.8%	7.0%	33.8%
Hispanic/Latino Ethnicity	254212	2486.6	83.6	36.2%	33.3%	24.5%	5.9%	30.4%
White	306052	2540.2	86.5	17.3%	26.3%	36.9%	19.5%	56.4%
Two or More Races	55601	2513.7	89.8	26.9%	29.4%	30.9%	12.8%	43.7%
Unidentified Race	16758	2551.4	83.4	13.3%	24.8%	39.4%	22.6%	62.0%
LEP Status	92325	2431.7	70.2	64.7%	27.4%	7.3%	0.7%	8.0%
IDEA Indicator	75997	2426.5	81.3	67.7%	21.5%	8.7%	2.1%	10.8%
Section 504 Status	9371	2517.7	84.7	23.9%	32.0%	31.4%	12.7%	44.1%
Economic Disadvantage Status	386416	2487.9	84.3	35.9%	32.8%	25.0%	6.3%	31.3%

TABLE 5.10. GRADE 7 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	739048	2540.1	95.8	27.6%	25.2%	33.6%	13.6%	47.2%
Female	362161	2555.2	92.7	21.9%	24.7%	36.9%	16.6%	53.5%
Male	376837	2525.6	96.6	33.0%	25.8%	30.5%	10.7%	41.2%
American Indian or Alaska Native	8966	2493.8	89.1	45.4%	27.6%	22.2%	4.7%	26.9%
Asian	56003	2596.8	93.5	12.2%	16.4%	38.7%	32.7%	71.4%
Black/African American	43651	2494.2	90.7	45.1%	27.1%	23.1%	4.7%	27.8%
Native Hawaiian or Pacific Islander	7779	2505.7	90.4	39.3%	28.3%	26.9%	5.4%	32.3%
Hispanic/Latino Ethnicity	249983	2505.1	87.3	39.6%	29.2%	26.2%	5.0%	31.2%
White	303446	2565.9	90.6	17.8%	23.2%	40.3%	18.7%	59.0%
Two or More Races	52657	2538.0	93.7	27.5%	26.4%	33.7%	12.4%	46.1%
Unidentified Race	16563	2573.1	86.0	15.0%	21.8%	43.8%	19.4%	63.2%
LEP Status	79810	2442.2	69.9	71.9%	21.4%	6.2%	0.5%	6.7%
IDEA Indicator	71646	2442.9	81.0	71.0%	18.6%	8.7%	1.7%	10.4%
Section 504 Status	10247	2541.0	90.2	25.7%	28.7%	33.1%	12.5%	45.6%
Economic Disadvantage Status	375934	2507.2	88.5	38.9%	28.8%	26.8%	5.5%	32.3%

TABLE 5.11. GRADE 8 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	742001	2559.6	94.7	23.4%	28.2%	35.1%	13.3%	48.4%
Female	362892	2576.2	90.7	17.4%	27.3%	38.9%	16.4%	55.3%
Male	379084	2543.8	95.7	29.1%	29.1%	31.5%	10.3%	41.8%
American Indian or Alaska Native	8813	2513.3	89.4	40.3%	31.4%	23.7%	4.6%	28.3%
Asian	56431	2615.6	92.3	9.9%	17.8%	40.6%	31.7%	72.3%
Black/African American	44921	2514.6	90.7	39.4%	31.4%	24.4%	4.8%	29.2%
Native Hawaiian or Pacific Islander	7407	2528.4	88.7	33.4%	32.3%	28.4%	5.9%	34.3%
Hispanic/Latino Ethnicity	249580	2527.5	85.7	32.7%	33.9%	28.2%	5.1%	33.3%
White	305957	2582.7	91.1	15.7%	24.9%	41.2%	18.1%	59.3%
Two or More Races	52036	2557.9	93.1	23.3%	29.3%	34.9%	12.5%	47.4%
Unidentified Race	16856	2592.1	83.6	11.5%	23.6%	46.4%	18.4%	64.8%
LEP Status	71747	2462.4	68.9	65.6%	27.3%	6.6%	0.5%	7.1%
IDEA Indicator	70143	2461.6	79.5	66.6%	23.0%	9.0%	1.4%	10.4%
Section 504 Status	11276	2559.8	90.6	22.0%	30.6%	35.0%	12.5%	47.5%
Economic Disadvantage Status	371394	2528.8	87.7	32.7%	33.0%	28.5%	5.7%	34.2%

TABLE 5.12. GRADE 11 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	630625	2598.0	108.2	18.3%	23.5%	33.9%	24.3%	58.2%
Female	309799	2613.8	102.0	13.4%	22.5%	36.3%	27.8%	64.1%
Male	320803	2582.8	111.8	23.0%	24.5%	31.5%	21.0%	52.5%
American Indian or Alaska Native	6970	2558.6	104.6	28.0%	29.2%	29.7%	13.2%	42.9%
Asian	53184	2651.6	105.7	9.3%	14.4%	31.1%	45.2%	76.3%
Black/African American	35149	2550.6	106.5	31.2%	28.2%	28.8%	11.7%	40.5%
Native Hawaiian or Pacific Islander	6360	2556.3	102.7	28.0%	29.7%	30.8%	11.4%	42.2%
Hispanic/Latino Ethnicity	214609	2567.0	101.2	24.4%	29.3%	33.0%	13.4%	46.4%
White	247703	2620.7	105.3	13.2%	19.6%	35.5%	31.7%	67.2%
Two or More Races	49600	2593.9	105.3	18.2%	25.0%	35.1%	21.7%	56.8%
Unidentified Race	17050	2634.2	92.4	8.2%	18.0%	40.2%	33.5%	73.7%
LEP Status	45401	2472.7	78.7	61.6%	29.7%	8.0%	0.8%	8.8%
IDEA Indicator	52742	2487.4	94.1	56.2%	27.5%	12.9%	3.4%	16.3%
Section 504 Status	11387	2607.9	103.1	14.4%	23.6%	36.1%	25.9%	62.0%
Economic Disadvantage Status	311642	2567.2	103.5	25.0%	28.5%	32.4%	14.2%	46.6%

TABLE 5.13. GRADE 3 MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	781627	2422.8	79.3	29.2%	26.3%	28.3%	16.2%	44.5%
Female	382059	2422.3	76.4	29.1%	27.2%	28.6%	15.1%	43.7%
Male	399539	2423.3	82.0	29.4%	25.4%	28.1%	17.1%	45.2%
American Indian or Alaska Native	9651	2388.1	73.1	45.1%	28.9%	20.0%	5.9%	25.9%
Asian	56753	2475.7	79.0	11.9%	17.5%	31.3%	39.3%	70.6%
Black/African American	44612	2382.5	74.7	48.2%	27.6%	19.0%	5.2%	24.2%
Native Hawaiian or Pacific Islander	7564	2404.0	75.3	36.2%	29.9%	24.1%	9.8%	33.9%
Hispanic/Latino Ethnicity	276667	2394.0	71.2	41.9%	30.0%	21.9%	6.2%	28.1%
White	303904	2445.6	74.3	18.0%	24.2%	35.0%	22.7%	57.7%
Two or More Races	65864	2421.4	78.9	29.9%	26.7%	27.8%	15.7%	43.5%
Unidentified Race	16612	2447.9	71.2	16.6%	24.1%	36.8%	22.5%	59.3%
LEP Status	186551	2384.3	70.6	48.3%	29.2%	17.5%	5.0%	22.5%
IDEA Indicator	75660	2362.6	87.7	59.3%	20.1%	14.3%	6.3%	20.6%
Section 504 Status	5722	2424.2	80.8	29.8%	24.6%	28.5%	17.0%	45.5%
Economic Disadvantage Status	420509	2396.9	73.2	40.4%	29.5%	22.8%	7.3%	30.1%

TABLE 5.14. GRADE 4 ELA LITERACY AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	765272	2462.5	80.8	26.5%	34.1%	24.5%	14.9%	39.4%
Female	374732	2461.6	77.2	26.2%	35.5%	24.7%	13.6%	38.3%
Male	390509	2463.3	84.0	26.9%	32.7%	24.3%	16.1%	40.4%
American Indian or Alaska Native	9442	2426.5	73.9	42.0%	36.8%	15.9%	5.4%	21.3%
Asian	57904	2519.4	81.4	10.0%	22.5%	28.9%	38.7%	67.6%
Black/African American	43694	2419.1	75.7	45.6%	35.2%	14.8%	4.5%	19.3%
Native Hawaiian or Pacific Islander	8466	2446.3	73.7	31.5%	37.8%	22.6%	8.1%	30.7%
Hispanic/Latino Ethnicity	266598	2430.3	70.6	39.7%	38.3%	17.0%	5.0%	22.0%
White	302312	2486.4	75.5	15.5%	32.1%	31.7%	20.8%	52.5%
Two or More Races	60413	2462.5	79.3	25.9%	35.2%	24.5%	14.3%	38.8%
Unidentified Race	16443	2486.8	72.9	14.5%	33.2%	31.9%	20.4%	52.3%
LEP Status	149456	2409.8	66.7	52.3%	35.2%	9.9%	2.7%	12.6%
IDEA Indicator	79400	2397.8	83.2	59.8%	25.1%	10.2%	4.9%	15.1%
Section 504 Status	7138	2467.2	78.3	23.6%	35.5%	25.6%	15.3%	40.9%
Economic Disadvantage Status	407588	2435.1	72.9	37.4%	38.0%	18.4%	6.3%	24.7%

TABLE 5.15. GRADE 5 MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	763194	2490.0	89.2	36.0%	29.4%	17.3%	17.3%	34.6%
Female	373516	2490.2	85.4	35.5%	30.9%	17.3%	16.3%	33.6%
Male	389661	2489.8	92.7	36.5%	27.9%	17.3%	18.2%	35.5%
American Indian or Alaska Native	9821	2447.1	80.4	55.5%	28.4%	10.3%	5.7%	16.0%
Asian	59223	2552.5	89.1	14.9%	21.7%	21.2%	42.1%	63.3%
Black/African American	43939	2440.1	81.9	58.8%	26.5%	9.4%	5.3%	14.7%
Native Hawaiian or Pacific Islander	8604	2471.7	82.0	43.0%	31.0%	16.1%	10.0%	26.1%
Hispanic/Latino Ethnicity	260106	2453.5	77.7	52.5%	29.8%	11.6%	6.1%	17.7%
White	306903	2516.5	83.3	22.8%	30.7%	22.4%	24.1%	46.5%
Two or More Races	58155	2490.0	87.1	35.5%	30.3%	17.5%	16.6%	34.1%
Unidentified Race	16443	2517.4	81.5	22.7%	30.3%	22.9%	24.1%	47.0%
LEP Status	124573	2422.6	70.2	70.6%	22.1%	5.1%	2.2%	7.3%
IDEA Indicator	80450	2414.0	85.3	72.2%	17.1%	6.1%	4.5%	10.6%
Section 504 Status	8609	2497.0	85.5	32.1%	31.6%	18.2%	18.0%	36.2%
Economic Disadvantage Status	398993	2459.0	80.2	49.5%	30.3%	12.7%	7.5%	20.2%

TABLE 5.16. GRADE 6 MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	751146	2510.9	101.0	33.3%	31.3%	19.2%	16.2%	35.4%
Female	367617	2515.0	96.3	31.3%	32.5%	20.1%	16.1%	36.2%
Male	383506	2506.9	105.1	35.2%	30.1%	18.3%	16.4%	34.7%
American Indian or Alaska Native	9081	2460.2	94.3	53.2%	30.6%	11.4%	4.9%	16.3%
Asian	58440	2584.2	98.5	13.0%	21.6%	22.8%	42.6%	65.4%
Black/African American	43599	2456.3	95.4	55.3%	29.0%	10.9%	4.9%	15.8%
Native Hawaiian or Pacific Islander	8152	2485.9	94.5	41.2%	34.1%	16.3%	8.4%	24.7%
Hispanic/Latino Ethnicity	254805	2474.4	91.6	46.7%	33.3%	14.0%	6.1%	20.1%
White	304541	2536.8	93.7	22.3%	31.6%	24.3%	21.8%	46.1%
Two or More Races	55678	2503.5	99.8	35.9%	32.0%	17.9%	14.2%	32.1%
Unidentified Race	16850	2545.3	90.1	19.3%	31.2%	25.5%	24.0%	49.5%
LEP Status	94114	2422.5	85.3	72.3%	21.7%	4.3%	1.6%	5.9%
IDEA Indicator	75841	2406.5	98.3	75.5%	16.5%	5.1%	2.9%	8.0%
Section 504 Status	9334	2515.1	94.1	31.4%	33.4%	19.7%	15.5%	35.2%
Economic Disadvantage Status	387313	2476.3	93.5	45.9%	33.0%	14.3%	6.8%	21.1%

TABLE 5.17. GRADE 7 MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	743128	2527.0	107.6	33.5%	29.6%	20.7%	16.2%	36.9%
Female	364145	2530.3	103.3	31.7%	31.1%	21.3%	15.9%	37.2%
Male	378929	2523.8	111.5	35.2%	28.1%	20.1%	16.5%	36.6%
American Indian or Alaska Native	9019	2477.9	98.2	51.3%	30.3%	12.9%	5.6%	18.5%
Asian	56725	2608.9	105.8	12.5%	19.1%	24.1%	44.3%	68.4%
Black/African American	43837	2468.3	99.8	55.3%	28.1%	12.0%	4.6%	16.6%
Native Hawaiian or Pacific Islander	7901	2495.4	100.2	43.8%	30.9%	17.7%	7.5%	25.2%
Hispanic/Latino Ethnicity	252164	2485.8	96.9	47.8%	31.6%	14.8%	5.8%	20.6%
White	303719	2555.3	99.5	22.3%	29.8%	26.2%	21.8%	48.0%
Two or More Races	53119	2523.0	104.9	34.3%	30.8%	20.4%	14.4%	34.8%
Unidentified Race	16644	2562.9	96.3	19.4%	29.3%	27.7%	23.6%	51.3%
LEP Status	82161	2428.4	90.1	74.4%	19.1%	4.6%	1.8%	6.4%
IDEA Indicator	72008	2418.0	98.8	76.3%	15.9%	5.3%	2.6%	7.9%
Section 504 Status	10258	2533.5	101.2	30.8%	31.8%	20.9%	16.5%	37.4%
Economic Disadvantage Status	378915	2489.7	98.8	46.2%	31.5%	15.6%	6.7%	22.3%

TABLE 5.18. GRADE 8 MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	743207	2541.9	115.1	37.4%	26.9%	18.0%	17.8%	35.8%
Female	363401	2547.7	110.0	34.8%	28.4%	19.0%	17.9%	36.9%
Male	379779	2536.5	119.4	39.8%	25.5%	17.0%	17.7%	34.7%
American Indian or Alaska Native	8784	2488.3	102.2	57.1%	25.9%	11.0%	6.1%	17.1%
Asian	56876	2632.3	115.1	14.2%	17.9%	20.4%	47.5%	67.9%
Black/African American	44922	2480.2	102.6	59.9%	24.5%	10.3%	5.3%	15.6%
Native Hawaiian or Pacific Islander	7420	2510.2	107.3	46.9%	27.7%	16.0%	9.3%	25.3%
Hispanic/Latino Ethnicity	251308	2499.7	101.6	51.5%	28.4%	13.3%	6.9%	20.2%
White	304833	2569.7	108.8	26.6%	27.6%	22.4%	23.4%	45.8%
Two or More Races	52146	2538.0	113.0	38.6%	27.3%	17.6%	16.5%	34.1%
Unidentified Race	16918	2582.5	104.1	21.7%	27.2%	24.8%	26.3%	51.1%
LEP Status	74020	2440.0	93.8	77.4%	15.9%	4.2%	2.5%	6.7%
IDEA Indicator	70356	2427.3	97.3	80.2%	13.0%	4.2%	2.5%	6.7%
Section 504 Status	11238	2545.0	110.8	36.2%	28.4%	17.8%	17.6%	35.4%
Economic Disadvantage Status	373469	2503.5	105.2	50.4%	27.5%	13.8%	8.3%	22.1%

TABLE 5.19. GRADE 11 MATHEMATICS AVERAGE OVERALL SCALE SCORE AND PERCENTAGE IN EACH ACHIEVEMENT LEVEL BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Achievement Levels				
	N	M	SD	1	2	3	4	3 & 4
Total	558553	2560.2	123.6	45.1%	25.3%	18.6%	11.0%	29.6%
Female	274092	2565.5	116.3	42.7%	27.3%	20.0%	10.0%	30.0%
Male	284439	2555.2	130.0	47.4%	23.4%	17.3%	11.9%	29.2%
American Indian or Alaska Native	5888	2517.1	109.0	60.8%	22.8%	12.5%	3.8%	16.3%
Asian	47818	2660.7	126.6	18.1%	19.1%	26.6%	36.3%	62.9%
Black/African American	32028	2501.9	109.4	64.8%	21.8%	10.5%	2.9%	13.4%
Native Hawaiian or Pacific Islander	5750	2523.9	109.0	56.3%	27.1%	12.4%	4.2%	16.6%
Hispanic/Latino Ethnicity	215900	2525.4	108.4	56.4%	25.9%	13.7%	4.0%	17.7%
White	202762	2583.4	121.9	36.3%	26.4%	23.1%	14.1%	37.2%
Two or More Races	31353	2546.2	120.4	49.8%	25.4%	16.0%	8.9%	24.9%
Unidentified Race	17054	2606.6	110.7	27.3%	27.8%	29.3%	15.6%	44.9%
LEP Status	43310	2455.6	98.0	83.9%	10.7%	3.7%	1.7%	5.4%
IDEA Indicator	47638	2446.0	95.9	84.9%	10.3%	3.4%	1.4%	4.8%
Section 504 Status	8385	2561.7	120.1	44.0%	27.0%	18.4%	10.6%	29.0%
Economic Disadvantage Status	285091	2525.9	111.7	56.5%	25.1%	13.6%	4.8%	18.4%

Claim-Level Results

Students also received achievement level assignments at the claim level. The content-specific claim scores are computed in relationship to the Level 3 cut score, and reported as Above Standard, At/Near Standard, or Below Standard.

- Table 5.20 presents each Smarter Balanced assessment claim.
- Table 5.21 through Table 5.27 presents aggregate student results for the ELA/literacy Claim 1 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.
- Table 5.28 through Table 5.34 presents aggregate student results for the ELA/literacy Claim 2 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.
- Table 5.35 through Table 5.41 presents aggregate student results for the ELA/literacy Claim 3 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.
- Table 5.42 through Table 5.48 presents aggregate student results for the ELA/literacy Claim 4 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.
- Table 5.49 through Table 5.55 presents aggregate student results for the Mathematics Claim 1 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.

- Table 5.56 through Table 5.62 presents aggregate student results for the Mathematics Claim 2/4 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.
- Table 5.63 through Table 5.69 presents aggregate student results for the Mathematics Claim 3 average scale score and for the percentage of students in each reporting category for grades 3 through 8 and 11.

These results are presented at the aggregate level (all submitted Consortium data) and disaggregated by gender, by race/ethnicity, and by various status flags: limited English proficiency, IDEA indicator, Section 504, and economically disadvantaged.

TABLE 5.20. 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 5.21. GRADE 3 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	776842	2408.1	100.3	36.4%	42.0%	21.5%
Female	379753	2418.7	99.1	32.0%	43.4%	24.5%
Male	397058	2397.9	100.4	40.6%	40.7%	18.7%
American Indian or Alaska Native	9536	2371.3	91.3	51.4%	38.9%	9.7%
Asian	55908	2451.9	96.6	20.4%	42.9%	36.7%
Black/African American	44474	2369.8	94.3	52.6%	37.0%	10.5%
Native Hawaiian or Pacific Islander	7516	2379.5	94.3	47.4%	40.5%	12.1%
Hispanic/Latino Ethnicity	275218	2374.7	92.5	49.9%	39.5%	10.6%
White	302323	2436.6	97.1	24.7%	44.6%	30.7%
Two or More Races	65339	2407.5	99.6	36.4%	42.7%	20.9%
Unidentified Race	16528	2435.2	93.3	24.1%	47.4%	28.5%
LEP Status	183892	2357.7	87.0	58.0%	35.6%	6.3%
IDEA Indicator	75412	2349.4	96.2	63.5%	28.3%	8.2%
Section 504 Status	5679	2411.2	100.5	34.6%	42.8%	22.5%
Economic Disadvantage Status	418082	2377.2	93.8	48.7%	39.9%	11.4%

TABLE 5.22. GRADE 4 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	738046	2448.0	108.8	34.3%	44.2%	21.6%
Female	361377	2459.1	107.5	30.3%	45.2%	24.5%
Male	376637	2437.3	109.0	38.0%	43.2%	18.8%
American Indian or Alaska Native	9311	2405.6	102.2	49.5%	40.6%	9.9%
Asian	55038	2499.1	103.6	18.1%	43.7%	38.2%
Black/African American	42197	2402.9	105.8	50.7%	39.0%	10.3%
Native Hawaiian or Pacific Islander	8334	2420.7	103.4	43.1%	44.1%	12.8%
Hispanic/Latino Ethnicity	252264	2410.2	100.6	48.1%	41.7%	10.2%
White	296265	2478.2	104.0	22.9%	46.9%	30.2%
Two or More Races	59189	2445.8	107.2	35.0%	44.5%	20.6%
Unidentified Race	15448	2477.4	100.0	22.7%	49.2%	28.1%
LEP Status	140655	2375.6	89.8	63.6%	32.9%	3.5%
IDEA Indicator	76464	2373.5	103.2	65.8%	27.5%	6.7%
Section 504 Status	6964	2450.6	106.1	33.0%	45.6%	21.4%
Economic Disadvantage Status	390792	2414.3	101.8	46.5%	42.2%	11.3%

TABLE 5.23. GRADE 5 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	761900	2485.5	106.0	33.9%	42.6%	23.5%
Female	372892	2499.1	104.2	28.8%	43.9%	27.3%
Male	388990	2472.5	106.1	38.8%	41.4%	19.9%
American Indian or Alaska Native	9809	2441.0	99.1	50.4%	38.8%	10.8%
Asian	58577	2533.0	102.2	18.3%	41.9%	39.8%
Black/African American	43949	2443.0	99.1	50.9%	37.6%	11.5%
Native Hawaiian or Pacific Islander	8589	2456.8	101.2	44.2%	41.6%	14.2%
Hispanic/Latino Ethnicity	259323	2447.6	97.7	47.9%	40.7%	11.5%
White	307315	2515.9	101.5	22.4%	44.6%	33.0%
Two or More Races	57938	2483.4	104.2	34.1%	43.9%	22.0%
Unidentified Race	16400	2510.0	97.8	22.7%	48.2%	29.0%
LEP Status	122514	2402.9	82.8	68.9%	28.4%	2.6%
IDEA Indicator	80620	2404.0	95.5	69.1%	25.0%	5.9%
Section 504 Status	8635	2489.3	103.7	31.4%	44.9%	23.7%
Economic Disadvantage Status	398182	2450.9	99.2	46.3%	41.3%	12.4%

TABLE 5.24. GRADE 6 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751374	2493.7	114.3	35.0%	47.2%	17.8%
Female	367791	2505.4	111.7	30.5%	49.5%	20.0%
Male	383560	2482.4	115.6	39.4%	44.9%	15.7%
American Indian or Alaska Native	9148	2445.8	108.6	52.5%	40.1%	7.4%
Asian	57880	2548.8	111.2	18.7%	47.1%	34.2%
Black/African American	43562	2447.9	107.3	52.6%	39.6%	7.7%
Native Hawaiian or Pacific Islander	8161	2459.9	110.9	46.3%	43.6%	10.2%
Hispanic/Latino Ethnicity	254212	2457.0	106.6	47.1%	44.4%	8.5%
White	306052	2521.9	109.6	25.2%	50.5%	24.4%
Two or More Races	55601	2487.7	114.7	36.3%	47.6%	16.1%
Unidentified Race	16758	2524.1	107.8	23.5%	52.4%	24.1%
LEP Status	92325	2398.7	90.4	72.0%	26.7%	1.3%
IDEA Indicator	75997	2403.6	101.4	70.0%	26.5%	3.5%
Section 504 Status	9371	2497.0	111.0	32.9%	49.7%	17.4%
Economic Disadvantage Status	386416	2458.6	108.3	46.4%	44.6%	9.0%

TABLE 5.25. GRADE 7 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	739048	2528.6	110.9	32.9%	46.4%	20.8%
Female	362161	2541.2	108.6	28.2%	48.2%	23.6%
Male	376837	2516.4	111.8	37.4%	44.6%	18.0%
American Indian or Alaska Native	8966	2484.7	105.4	47.7%	42.4%	9.9%
Asian	56003	2582.8	108.3	17.3%	43.9%	38.8%
Black/African American	43651	2483.8	103.6	48.9%	41.6%	9.5%
Native Hawaiian or Pacific Islander	7779	2490.3	105.6	46.7%	42.2%	11.1%
Hispanic/Latino Ethnicity	249983	2489.7	102.6	46.4%	43.6%	10.0%
White	303446	2558.4	105.8	22.0%	49.6%	28.4%
Two or More Races	52657	2524.6	109.8	33.7%	47.2%	19.1%
Unidentified Race	16563	2556.6	102.7	22.2%	50.9%	27.0%
LEP Status	79810	2424.2	83.7	75.5%	23.2%	1.3%
IDEA Indicator	71646	2432.9	96.5	71.5%	24.7%	3.8%
Section 504 Status	10247	2534.1	107.6	30.7%	48.0%	21.3%
Economic Disadvantage Status	375934	2492.5	104.3	45.5%	43.7%	10.8%

TABLE 5.26. GRADE 8 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	742001	2553.4	107.5	29.2%	46.2%	24.6%
Female	362892	2567.5	104.5	24.2%	47.5%	28.3%
Male	379084	2540.0	108.6	34.0%	44.9%	21.1%
American Indian or Alaska Native	8813	2510.2	103.5	44.9%	42.8%	12.3%
Asian	56431	2606.8	102.3	14.2%	42.0%	43.8%
Black/African American	44921	2510.3	103.0	44.6%	43.0%	12.3%
Native Hawaiian or Pacific Islander	7407	2518.0	104.3	41.6%	44.5%	13.9%
Hispanic/Latino Ethnicity	249580	2518.9	101.2	40.2%	46.7%	13.1%
White	305957	2579.2	102.8	20.4%	46.7%	32.8%
Two or More Races	52036	2550.0	106.6	30.0%	47.2%	22.8%
Unidentified Race	16856	2583.8	97.0	17.8%	50.2%	32.0%
LEP Status	71747	2451.7	84.4	70.3%	27.7%	2.0%
IDEA Indicator	70143	2458.0	94.7	67.5%	28.1%	4.3%
Section 504 Status	11276	2556.4	104.8	27.6%	47.8%	24.6%
Economic Disadvantage Status	371394	2520.7	103.1	39.8%	46.2%	14.0%

TABLE 5.27. GRADE 11 ELA/LITERACY AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	630625	2597.9	115.3	19.5%	48.0%	32.5%
Female	309799	2607.8	111.7	16.6%	48.2%	35.2%
Male	320803	2588.3	117.9	22.3%	47.9%	29.8%
American Indian or Alaska Native	6970	2563.5	114.8	28.1%	50.2%	21.7%
Asian	53184	2640.6	113.0	11.8%	39.6%	48.6%
Black/African American	35149	2555.2	113.6	31.0%	49.6%	19.5%
Native Hawaiian or Pacific Islander	6360	2556.5	111.4	30.3%	50.5%	19.2%
Hispanic/Latino Ethnicity	214609	2566.5	108.1	25.8%	53.5%	20.7%
White	247703	2623.4	113.3	14.0%	44.4%	41.7%
Two or More Races	49600	2592.9	114.4	20.3%	49.2%	30.5%
Unidentified Race	17050	2622.2	100.9	11.1%	50.1%	38.9%
LEP Status	45401	2474.9	87.9	59.8%	38.1%	2.1%
IDEA Indicator	52742	2497.1	105.3	51.7%	40.9%	7.4%
Section 504 Status	11387	2614.0	112.9	15.5%	47.1%	37.4%
Economic Disadvantage Status	311642	2567.4	110.9	26.3%	52.0%	21.7%

TABLE 5.28. GRADE 3 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	776842	2407.4	101.3	32.9%	47.0%	20.0%
Female	379753	2419.8	100.3	28.5%	47.9%	23.6%
Male	397058	2395.6	100.9	37.2%	46.2%	16.6%
American Indian or Alaska Native	9536	2368.7	93.5	47.8%	44.0%	8.2%
Asian	55908	2457.5	99.4	17.3%	44.4%	38.3%
Black/African American	44474	2370.3	98.8	46.7%	43.4%	10.0%
Native Hawaiian or Pacific Islander	7516	2391.6	95.4	38.4%	47.6%	14.0%
Hispanic/Latino Ethnicity	275218	2371.4	94.8	46.8%	43.7%	9.5%
White	302323	2436.4	94.9	21.3%	50.9%	27.8%
Two or More Races	65339	2407.0	99.5	32.8%	47.7%	19.5%
Unidentified Race	16528	2439.0	94.6	20.8%	49.6%	29.7%
LEP Status	183892	2354.3	91.5	54.6%	39.4%	6.0%
IDEA Indicator	75412	2336.6	102.2	62.8%	30.3%	7.0%
Section 504 Status	5679	2407.6	98.4	32.6%	48.0%	19.5%
Economic Disadvantage Status	418082	2375.8	95.4	45.1%	44.3%	10.7%

TABLE 5.29. GRADE 4 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	738046	2455.8	103.3	30.8%	49.3%	19.8%
Female	361377	2471.8	101.4	25.4%	50.4%	24.2%
Male	376637	2440.5	102.7	36.0%	48.3%	15.7%
American Indian or Alaska Native	9311	2411.6	96.8	47.4%	44.6%	8.0%
Asian	55038	2511.1	101.0	15.0%	44.7%	40.4%
Black/African American	42197	2415.5	99.9	45.2%	45.4%	9.5%
Native Hawaiian or Pacific Islander	8334	2434.3	98.1	38.2%	48.7%	13.2%
Hispanic/Latino Ethnicity	252264	2419.9	96.1	44.1%	46.3%	9.6%
White	296265	2482.6	97.8	20.1%	53.4%	26.4%
Two or More Races	59189	2453.1	100.8	31.8%	49.6%	18.6%
Unidentified Race	15448	2491.6	95.6	18.4%	50.2%	31.4%
LEP Status	140655	2386.7	88.7	59.2%	37.5%	3.3%
IDEA Indicator	76464	2371.9	101.6	66.0%	28.5%	5.5%
Section 504 Status	6964	2452.9	96.5	31.0%	51.4%	17.6%
Economic Disadvantage Status	390792	2423.8	96.4	42.7%	46.7%	10.6%

TABLE 5.30. GRADE 5 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	761900	2494.4	104.9	29.9%	46.2%	23.9%
Female	372892	2512.8	102.4	23.7%	47.0%	29.4%
Male	388990	2476.7	104.3	35.9%	45.4%	18.7%
American Indian or Alaska Native	9809	2443.3	97.4	48.7%	41.6%	9.6%
Asian	58577	2551.4	104.2	14.8%	39.0%	46.2%
Black/African American	43949	2448.6	102.2	45.9%	42.5%	11.6%
Native Hawaiian or Pacific Islander	8589	2473.9	95.4	35.7%	48.9%	15.5%
Hispanic/Latino Ethnicity	259323	2461.4	98.1	41.7%	44.8%	13.5%
White	307315	2518.7	100.0	20.4%	49.1%	30.5%
Two or More Races	57938	2490.7	100.9	30.5%	47.8%	21.8%
Unidentified Race	16400	2533.8	99.0	17.8%	44.3%	37.9%
LEP Status	122514	2416.4	86.4	61.3%	35.2%	3.5%
IDEA Indicator	80620	2401.0	98.5	68.6%	25.9%	5.5%
Section 504 Status	8635	2492.5	99.2	29.9%	48.3%	21.9%
Economic Disadvantage Status	398182	2463.3	97.6	40.8%	45.3%	13.8%

TABLE 5.31. GRADE 6 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751374	2519.1	102.5	28.3%	50.4%	21.3%
Female	367791	2537.2	99.4	22.3%	51.3%	26.4%
Male	383560	2501.7	102.5	34.1%	49.6%	16.3%
American Indian or Alaska Native	9148	2467.0	99.7	46.9%	45.2%	7.9%
Asian	57880	2576.5	98.9	13.2%	42.8%	44.0%
Black/African American	43562	2474.5	104.0	40.0%	49.8%	10.2%
Native Hawaiian or Pacific Islander	8161	2494.4	96.8	38.5%	48.4%	13.1%
Hispanic/Latino Ethnicity	254212	2488.9	94.4	39.9%	49.1%	11.0%
White	306052	2540.4	99.4	19.2%	53.6%	27.2%
Two or More Races	55601	2514.8	100.3	31.1%	49.2%	19.7%
Unidentified Race	16758	2558.7	91.7	16.6%	48.9%	34.5%
LEP Status	92325	2433.8	87.3	65.0%	33.0%	2.0%
IDEA Indicator	75997	2422.7	96.1	71.2%	25.3%	3.5%
Section 504 Status	9371	2515.9	95.6	29.0%	52.5%	18.5%
Economic Disadvantage Status	386416	2489.9	94.9	40.8%	47.5%	11.7%

TABLE 5.32. GRADE 7 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	739048	2546.4	108.4	27.4%	47.4%	25.2%
Female	362161	2566.9	104.2	20.9%	47.9%	31.2%
Male	376837	2526.6	108.7	33.6%	46.9%	19.4%
American Indian or Alaska Native	8966	2494.1	104.8	45.5%	43.3%	11.2%
Asian	56003	2607.5	102.8	12.4%	37.7%	49.9%
Black/African American	43651	2499.4	107.6	42.3%	45.4%	12.4%
Native Hawaiian or Pacific Islander	7779	2512.4	104.6	38.0%	47.2%	14.8%
Hispanic/Latino Ethnicity	249983	2511.8	100.6	38.1%	48.5%	13.4%
White	303446	2571.3	103.7	18.9%	48.8%	32.3%
Two or More Races	52657	2541.9	106.6	28.7%	47.5%	23.8%
Unidentified Race	16563	2587.7	96.1	14.7%	45.7%	39.6%
LEP Status	79810	2446.5	90.9	65.9%	31.7%	2.4%
IDEA Indicator	71646	2438.6	97.4	69.8%	26.4%	3.7%
Section 504 Status	10247	2540.4	101.3	28.6%	49.8%	21.6%
Economic Disadvantage Status	375934	2512.6	101.8	38.2%	47.4%	14.3%

TABLE 5.33. GRADE 8 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	742001	2561.8	107.3	26.3%	51.0%	22.7%
Female	362892	2583.6	102.6	19.4%	52.0%	28.6%
Male	379084	2541.0	107.5	33.0%	49.9%	17.1%
American Indian or Alaska Native	8813	2510.7	102.1	44.1%	46.4%	9.5%
Asian	56431	2621.5	103.9	11.8%	43.1%	45.1%
Black/African American	44921	2513.1	107.1	41.6%	47.6%	10.8%
Native Hawaiian or Pacific Islander	7407	2534.1	100.0	35.1%	50.8%	14.1%
Hispanic/Latino Ethnicity	249580	2529.9	96.8	36.0%	52.2%	11.7%
White	305957	2584.5	105.5	18.6%	52.0%	29.4%
Two or More Races	52036	2559.8	105.7	27.4%	50.5%	22.0%
Unidentified Race	16856	2599.9	93.7	13.7%	52.9%	33.3%
LEP Status	71747	2464.1	85.8	65.3%	33.0%	1.7%
IDEA Indicator	70143	2457.6	93.0	69.2%	27.6%	3.1%
Section 504 Status	11276	2558.2	102.9	27.5%	51.8%	20.7%
Economic Disadvantage Status	371394	2531.4	99.0	36.4%	50.8%	12.9%

TABLE 5.34. GRADE 11 ELA/LITERACY AVERAGE CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	630625	2597.4	123.6	21.1%	45.9%	33.0%
Female	309799	2620.0	115.3	15.1%	45.9%	39.0%
Male	320803	2575.6	127.4	27.0%	45.8%	27.2%
American Indian or Alaska Native	6970	2553.8	120.6	31.6%	49.0%	19.3%
Asian	53184	2660.3	117.7	10.2%	33.9%	55.9%
Black/African American	35149	2544.3	123.5	35.2%	46.5%	18.3%
Native Hawaiian or Pacific Islander	6360	2557.0	119.4	30.4%	50.0%	19.6%
Hispanic/Latino Ethnicity	214609	2563.1	117.7	28.5%	50.5%	21.0%
White	247703	2621.0	119.2	15.4%	43.9%	40.7%
Two or More Races	49600	2592.9	120.3	21.0%	48.5%	30.5%
Unidentified Race	17050	2645.5	107.5	9.7%	41.6%	48.6%
LEP Status	45401	2464.5	101.5	62.9%	34.4%	2.7%
IDEA Indicator	52742	2475.0	111.1	59.4%	35.2%	5.4%
Section 504 Status	11387	2602.8	117.7	18.0%	48.8%	33.1%
Economic Disadvantage Status	311642	2564.3	120.1	28.4%	49.6%	22.0%

TABLE 5.35. GRADE 3 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	776842	2415.9	117.9	20.6%	62.6%	16.8%
Female	379753	2423.0	115.8	18.4%	63.6%	18.0%
Male	397058	2409.2	119.6	22.7%	61.7%	15.6%
American Indian or Alaska Native	9536	2374.2	116.0	31.1%	61.0%	7.8%
Asian	55908	2459.9	109.7	10.4%	61.7%	27.9%
Black/African American	44474	2370.7	120.1	30.6%	61.4%	8.0%
Native Hawaiian or Pacific Islander	7516	2389.8	114.4	27.3%	63.4%	9.4%
Hispanic/Latino Ethnicity	275218	2382.3	113.3	29.8%	62.2%	8.1%
White	302323	2445.8	112.6	12.5%	63.2%	24.3%
Two or More Races	65339	2414.8	117.3	21.3%	62.5%	16.2%
Unidentified Race	16528	2443.9	105.3	12.2%	67.8%	20.0%
LEP Status	183892	2361.4	111.3	36.4%	58.7%	4.9%
IDEA Indicator	75412	2343.6	125.9	46.7%	46.3%	7.0%
Section 504 Status	5679	2424.5	118.5	19.3%	61.2%	19.5%
Economic Disadvantage Status	418082	2386.0	114.1	29.3%	61.8%	8.9%

TABLE 5.36. GRADE 4 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	738046	2454.5	121.4	21.3%	61.7%	17.0%
Female	361377	2459.5	119.0	19.6%	62.7%	17.7%
Male	376637	2449.8	123.5	22.9%	60.8%	16.3%
American Indian or Alaska Native	9311	2407.3	120.0	32.0%	60.5%	7.5%
Asian	55038	2503.3	114.8	11.3%	58.1%	30.7%
Black/African American	42197	2405.1	122.2	35.5%	56.5%	8.0%
Native Hawaiian or Pacific Islander	8334	2424.2	120.7	28.0%	61.7%	10.2%
Hispanic/Latino Ethnicity	252264	2418.2	115.7	29.6%	62.0%	8.3%
White	296265	2484.9	115.7	13.9%	62.5%	23.6%
Two or More Races	59189	2452.2	120.2	21.1%	62.8%	16.1%
Unidentified Race	15448	2480.3	111.8	13.5%	65.3%	21.2%
LEP Status	140655	2383.0	110.0	40.4%	56.3%	3.3%
IDEA Indicator	76464	2376.4	126.1	45.0%	48.9%	6.1%
Section 504 Status	6964	2461.9	120.7	18.8%	62.9%	18.4%
Economic Disadvantage Status	390792	2422.3	117.1	28.2%	62.6%	9.2%

TABLE 5.37. GRADE 5 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	761900	2477.5	130.1	21.1%	62.6%	16.3%
Female	372892	2482.2	128.8	19.7%	63.1%	17.2%
Male	388990	2473.0	131.2	22.5%	62.1%	15.4%
American Indian or Alaska Native	9809	2428.3	123.7	32.6%	60.6%	6.8%
Asian	58577	2528.9	125.5	11.5%	58.3%	30.2%
Black/African American	43949	2424.1	128.6	33.5%	59.1%	7.4%
Native Hawaiian or Pacific Islander	8589	2449.0	127.8	27.7%	62.4%	9.9%
Hispanic/Latino Ethnicity	259323	2440.4	125.1	30.0%	61.6%	8.4%
White	307315	2507.5	124.2	13.6%	64.4%	22.0%
Two or More Races	57938	2477.0	128.5	20.7%	64.1%	15.1%
Unidentified Race	16400	2507.1	121.8	13.9%	64.7%	21.4%
LEP Status	122514	2391.0	113.3	44.8%	52.9%	2.3%
IDEA Indicator	80620	2388.4	126.1	48.2%	47.1%	4.6%
Section 504 Status	8635	2486.7	126.3	18.8%	64.0%	17.2%
Economic Disadvantage Status	398182	2443.8	126.0	29.2%	62.0%	8.8%

TABLE 5.38. GRADE 6 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751374	2525.6	128.0	17.6%	67.9%	14.5%
Female	367791	2536.2	124.8	15.0%	69.0%	16.0%
Male	383560	2515.4	130.2	20.1%	66.8%	13.1%
American Indian or Alaska Native	9148	2475.6	127.9	29.5%	63.4%	7.1%
Asian	57880	2573.8	120.1	9.1%	65.4%	25.5%
Black/African American	43562	2478.7	129.9	28.4%	64.2%	7.4%
Native Hawaiian or Pacific Islander	8161	2499.5	129.0	23.6%	66.6%	9.9%
Hispanic/Latino Ethnicity	254212	2493.5	124.9	24.3%	67.5%	8.2%
White	306052	2550.8	122.7	11.9%	69.2%	19.0%
Two or More Races	55601	2522.5	129.0	18.5%	68.1%	13.4%
Unidentified Race	16758	2555.7	116.0	10.4%	70.8%	18.8%
LEP Status	92325	2427.6	118.4	43.8%	54.1%	2.0%
IDEA Indicator	75997	2420.4	128.2	48.3%	48.3%	3.5%
Section 504 Status	9371	2531.3	122.5	14.9%	70.9%	14.2%
Economic Disadvantage Status	386416	2495.7	126.3	24.1%	67.3%	8.6%

TABLE 5.39. GRADE 7 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	739048	2537.4	124.5	21.8%	64.5%	13.7%
Female	362161	2544.8	122.7	19.7%	65.5%	14.8%
Male	376837	2530.3	125.9	23.8%	63.6%	12.6%
American Indian or Alaska Native	8966	2492.3	122.2	32.6%	60.8%	6.6%
Asian	56003	2583.9	116.7	11.4%	65.4%	23.2%
Black/African American	43651	2488.9	121.8	35.0%	58.5%	6.5%
Native Hawaiian or Pacific Islander	7779	2499.4	120.8	31.5%	61.2%	7.3%
Hispanic/Latino Ethnicity	249983	2501.5	118.9	30.4%	62.4%	7.2%
White	303446	2566.6	120.3	14.5%	66.8%	18.7%
Two or More Races	52657	2536.6	124.7	21.6%	65.3%	13.1%
Unidentified Race	16563	2560.9	114.4	14.2%	69.5%	16.3%
LEP Status	79810	2434.3	103.7	52.9%	45.8%	1.3%
IDEA Indicator	71646	2440.3	115.3	51.6%	45.5%	2.9%
Section 504 Status	10247	2545.8	121.2	18.8%	66.4%	14.8%
Economic Disadvantage Status	375934	2505.0	120.4	29.4%	62.9%	7.7%

TABLE 5.40. GRADE 8 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	742001	2552.0	125.1	20.7%	66.1%	13.2%
Female	362892	2562.8	121.3	17.1%	68.5%	14.4%
Male	379084	2541.7	127.8	24.2%	63.8%	12.1%
American Indian or Alaska Native	8813	2507.4	123.2	32.9%	60.6%	6.5%
Asian	56431	2603.6	117.6	10.0%	65.2%	24.8%
Black/African American	44921	2504.5	124.7	32.5%	61.1%	6.4%
Native Hawaiian or Pacific Islander	7407	2519.7	120.0	29.6%	62.6%	7.8%
Hispanic/Latino Ethnicity	249580	2521.4	117.6	28.3%	64.8%	6.9%
White	305957	2574.6	124.7	14.7%	67.9%	17.4%
Two or More Races	52036	2553.3	123.1	20.5%	66.7%	12.8%
Unidentified Race	16856	2582.6	112.2	12.3%	70.9%	16.8%
LEP Status	71747	2450.7	104.2	53.3%	45.5%	1.1%
IDEA Indicator	70143	2454.7	114.5	54.2%	43.2%	2.7%
Section 504 Status	11276	2559.6	121.6	18.2%	67.9%	13.9%
Economic Disadvantage Status	371394	2523.7	118.9	28.3%	64.3%	7.5%

TABLE 5.41. GRADE 11 ELA/LITERACY AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	630625	2580.0	136.8	19.7%	62.0%	18.3%
Female	309799	2590.5	132.1	16.2%	64.5%	19.3%
Male	320803	2569.9	140.4	23.1%	59.6%	17.4%
American Indian or Alaska Native	6970	2543.6	134.7	28.2%	60.7%	11.2%
Asian	53184	2626.1	132.1	11.5%	58.7%	29.8%
Black/African American	35149	2539.0	135.0	29.8%	59.6%	10.6%
Native Hawaiian or Pacific Islander	6360	2536.2	132.6	29.9%	60.1%	9.9%
Hispanic/Latino Ethnicity	214609	2550.5	130.7	25.3%	63.0%	11.6%
White	247703	2602.1	136.5	15.2%	61.6%	23.1%
Two or More Races	49600	2578.1	136.6	19.6%	63.6%	16.8%
Unidentified Race	17050	2608.0	126.1	11.7%	65.2%	23.1%
LEP Status	45401	2460.7	106.4	54.4%	44.4%	1.2%
IDEA Indicator	52742	2476.8	121.7	49.8%	46.4%	3.8%
Section 504 Status	11387	2592.1	134.8	16.5%	63.7%	19.9%
Economic Disadvantage Status	311642	2550.5	132.9	25.8%	62.4%	11.9%

TABLE 5.42. GRADE 3 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	776842	2400.9	118.6	28.3%	52.5%	19.2%
Female	379753	2410.5	117.3	25.0%	53.6%	21.4%
Male	397058	2391.8	119.2	31.5%	51.4%	17.1%
American Indian or Alaska Native	9536	2357.1	111.3	41.7%	50.4%	7.9%
Asian	55908	2452.0	115.7	15.3%	48.6%	36.1%
Black/African American	44474	2357.7	115.4	41.6%	48.9%	9.5%
Native Hawaiian or Pacific Islander	7516	2377.7	114.8	36.2%	50.3%	13.6%
Hispanic/Latino Ethnicity	275218	2368.1	112.4	38.8%	50.5%	10.7%
White	302323	2427.8	115.2	19.2%	55.5%	25.3%
Two or More Races	65339	2401.3	116.7	28.0%	52.8%	19.2%
Unidentified Race	16528	2433.8	112.5	17.8%	53.6%	28.6%
LEP Status	183892	2350.7	108.4	45.2%	47.8%	7.0%
IDEA Indicator	75412	2340.0	114.8	51.8%	40.0%	8.1%
Section 504 Status	5679	2403.5	117.6	27.2%	52.9%	19.9%
Economic Disadvantage Status	418082	2370.6	112.5	38.0%	50.7%	11.3%

TABLE 5.43. GRADE 4 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	738046	2438.0	124.4	26.8%	54.2%	19.0%
Female	361377	2448.8	123.5	23.8%	54.7%	21.5%
Male	376637	2427.7	124.4	29.8%	53.7%	16.5%
American Indian or Alaska Native	9311	2389.0	117.2	41.4%	51.0%	7.6%
Asian	55038	2497.7	117.8	13.4%	49.2%	37.4%
Black/African American	42197	2387.5	124.6	39.2%	52.0%	8.8%
Native Hawaiian or Pacific Islander	8334	2420.0	119.8	34.4%	51.5%	14.1%
Hispanic/Latino Ethnicity	252264	2405.0	118.2	38.0%	51.3%	10.7%
White	296265	2461.8	121.2	17.9%	58.1%	24.0%
Two or More Races	59189	2440.3	121.9	27.0%	54.0%	19.0%
Unidentified Race	15448	2478.6	114.4	16.8%	53.9%	29.3%
LEP Status	140655	2371.0	110.0	50.4%	45.0%	4.6%
IDEA Indicator	76464	2364.4	118.0	55.8%	37.7%	6.5%
Section 504 Status	6964	2437.0	121.5	26.5%	55.4%	18.1%
Economic Disadvantage Status	390792	2408.3	117.9	37.4%	51.4%	11.2%

TABLE 5.44. GRADE 5 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	761900	2506.8	113.2	19.0%	53.6%	27.4%
Female	372892	2519.5	110.3	15.8%	53.1%	31.1%
Male	388990	2494.7	114.6	22.1%	54.0%	23.9%
American Indian or Alaska Native	9809	2452.3	111.3	33.4%	54.1%	12.4%
Asian	58577	2562.7	105.3	8.8%	42.1%	49.1%
Black/African American	43949	2458.7	114.2	31.2%	54.3%	14.5%
Native Hawaiian or Pacific Islander	8589	2487.6	111.7	23.7%	54.9%	21.4%
Hispanic/Latino Ethnicity	259323	2475.4	110.3	26.8%	55.8%	17.4%
White	307315	2529.9	106.7	12.5%	53.8%	33.7%
Two or More Races	57938	2505.9	111.2	18.7%	54.5%	26.8%
Unidentified Race	16400	2545.7	102.8	10.0%	49.2%	40.8%
LEP Status	122514	2427.2	102.1	42.1%	52.0%	5.9%
IDEA Indicator	80620	2417.2	113.0	48.3%	43.8%	8.0%
Section 504 Status	8635	2509.4	105.9	16.6%	57.0%	26.4%
Economic Disadvantage Status	398182	2477.7	110.0	26.0%	55.9%	18.0%

TABLE 5.45. GRADE 6 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751374	2521.9	114.2	16.8%	60.4%	22.7%
Female	367791	2536.4	110.9	13.5%	59.9%	26.7%
Male	383560	2507.9	115.5	20.1%	61.0%	19.0%
American Indian or Alaska Native	9148	2472.9	112.3	29.1%	61.0%	9.8%
Asian	57880	2578.5	105.0	7.6%	47.8%	44.7%
Black/African American	43562	2477.0	116.6	26.9%	61.2%	11.8%
Native Hawaiian or Pacific Islander	8161	2496.8	112.5	23.0%	60.9%	16.2%
Hispanic/Latino Ethnicity	254212	2492.1	110.6	23.8%	62.0%	14.2%
White	306052	2542.8	109.6	11.2%	61.6%	27.1%
Two or More Races	55601	2520.5	112.2	16.8%	60.5%	22.7%
Unidentified Race	16758	2557.0	102.4	9.5%	55.4%	35.1%
LEP Status	92325	2436.2	104.2	41.8%	54.4%	3.8%
IDEA Indicator	75997	2432.4	109.1	44.9%	50.0%	5.1%
Section 504 Status	9371	2520.7	108.8	15.2%	63.5%	21.4%
Economic Disadvantage Status	386416	2492.9	110.3	23.6%	61.6%	14.8%

TABLE 5.46. GRADE 7 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	739048	2533.6	120.5	21.9%	55.0%	23.1%
Female	362161	2550.5	116.3	17.2%	55.8%	27.0%
Male	376837	2517.3	122.2	26.4%	54.3%	19.4%
American Indian or Alaska Native	8966	2485.1	115.6	33.9%	55.7%	10.3%
Asian	56003	2597.1	110.9	9.6%	44.9%	45.5%
Black/African American	43651	2484.9	118.4	34.4%	54.3%	11.3%
Native Hawaiian or Pacific Islander	7779	2500.7	118.2	31.0%	54.6%	14.4%
Hispanic/Latino Ethnicity	249983	2498.8	115.9	31.3%	55.4%	13.4%
White	303446	2557.1	115.5	14.7%	56.6%	28.6%
Two or More Races	52657	2537.6	118.7	20.9%	55.3%	23.8%
Unidentified Race	16563	2570.6	108.4	12.6%	53.7%	33.7%
LEP Status	79810	2436.5	102.2	53.6%	43.6%	2.8%
IDEA Indicator	71646	2437.8	107.5	54.5%	41.0%	4.5%
Section 504 Status	10247	2533.2	117.4	20.8%	57.2%	22.0%
Economic Disadvantage Status	375934	2501.0	116.7	30.9%	54.9%	14.2%

TABLE 5.47. GRADE 8 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	742001	2553.0	121.0	20.5%	56.9%	22.6%
Female	362892	2571.8	117.0	15.7%	57.2%	27.0%
Male	379084	2534.9	122.1	25.0%	56.5%	18.4%
American Indian or Alaska Native	8813	2503.9	117.1	32.6%	57.0%	10.4%
Asian	56431	2615.3	111.7	9.1%	46.8%	44.2%
Black/African American	44921	2505.7	118.7	31.7%	57.1%	11.2%
Native Hawaiian or Pacific Islander	7407	2521.9	116.8	28.9%	55.8%	15.3%
Hispanic/Latino Ethnicity	249580	2520.4	115.0	28.3%	57.9%	13.7%
White	305957	2574.8	118.5	14.5%	58.1%	27.4%
Two or More Races	52036	2555.4	118.8	20.0%	56.4%	23.6%
Unidentified Race	16856	2587.5	109.9	12.1%	54.8%	33.1%
LEP Status	71747	2455.8	100.5	50.8%	46.4%	2.8%
IDEA Indicator	70143	2454.7	103.8	52.4%	43.6%	4.0%
Section 504 Status	11276	2548.8	118.8	20.7%	58.0%	21.3%
Economic Disadvantage Status	371394	2521.1	116.0	28.6%	56.8%	14.6%

TABLE 5.48. GRADE 11 ELA/LITERACY AVERAGE CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	630625	2596.8	133.5	15.5%	50.7%	33.8%
Female	309799	2615.5	126.6	11.6%	50.0%	38.3%
Male	320803	2578.7	137.5	19.3%	51.4%	29.3%
American Indian or Alaska Native	6970	2551.6	134.9	23.9%	55.0%	21.1%
Asian	53184	2655.9	120.2	7.3%	38.7%	54.0%
Black/African American	35149	2545.4	134.9	25.5%	54.6%	19.9%
Native Hawaiian or Pacific Islander	6360	2552.9	132.1	23.1%	56.4%	20.5%
Hispanic/Latino Ethnicity	214609	2569.7	131.1	19.5%	55.9%	24.6%
White	247703	2614.4	130.8	12.5%	48.0%	39.5%
Two or More Races	49600	2595.5	132.7	15.6%	52.2%	32.2%
Unidentified Race	17050	2640.2	114.5	7.4%	46.5%	46.2%
LEP Status	45401	2473.2	115.9	42.6%	53.2%	4.2%
IDEA Indicator	52742	2481.3	123.6	41.3%	51.5%	7.2%
Section 504 Status	11387	2602.2	130.4	13.9%	51.6%	34.5%
Economic Disadvantage Status	311642	2567.8	133.3	20.4%	55.0%	24.6%

TABLE 5.49. GRADE 3 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	781627	2424.3	83.8	36.4%	36.3%	27.3%
Female	382059	2423.2	80.3	36.8%	37.4%	25.8%
Male	399539	2425.4	87.1	36.0%	35.3%	28.7%
American Indian or Alaska Native	9651	2388.1	78.9	54.0%	33.1%	12.9%
Asian	56753	2480.1	82.6	15.6%	30.2%	54.3%
Black/African American	44612	2383.3	81.0	55.3%	32.9%	11.8%
Native Hawaiian or Pacific Islander	7564	2407.7	80.3	43.3%	36.4%	20.3%
Hispanic/Latino Ethnicity	276667	2396.7	76.2	49.6%	35.7%	14.6%
White	303904	2445.6	79.4	25.2%	38.5%	36.3%
Two or More Races	65864	2422.5	83.9	37.5%	36.0%	26.5%
Unidentified Race	16612	2450.3	75.0	22.2%	39.5%	38.4%
LEP Status	186551	2387.8	76.4	55.5%	32.6%	11.8%
IDEA Indicator	75660	2361.1	94.0	65.2%	23.1%	11.7%
Section 504 Status	5722	2424.4	85.4	37.2%	34.4%	28.4%
Economic Disadvantage Status	420509	2399.2	78.3	48.2%	35.6%	16.2%

TABLE 5.50. GRADE 4 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	765272	2463.8	85.0	42.3%	33.6%	24.1%
Female	374732	2462.1	80.8	43.1%	34.6%	22.3%
Male	390509	2465.5	88.8	41.5%	32.7%	25.9%
American Indian or Alaska Native	9442	2427.0	79.2	60.4%	28.6%	11.0%
Asian	57904	2523.3	85.0	18.7%	29.5%	51.8%
Black/African American	43694	2420.3	81.5	62.9%	27.5%	9.5%
Native Hawaiian or Pacific Islander	8466	2448.7	79.3	48.5%	34.4%	17.0%
Hispanic/Latino Ethnicity	266598	2431.5	75.3	58.7%	30.5%	10.8%
White	302312	2487.3	79.5	29.4%	37.8%	32.8%
Two or More Races	60413	2464.1	84.0	42.0%	34.0%	24.0%
Unidentified Race	16443	2489.6	76.9	28.1%	38.6%	33.3%
LEP Status	149456	2411.5	73.2	70.5%	23.3%	6.2%
IDEA Indicator	79400	2396.2	90.8	72.9%	18.3%	8.8%
Section 504 Status	7138	2469.5	82.7	40.0%	34.4%	25.5%
Economic Disadvantage Status	407588	2436.5	77.9	55.8%	31.3%	12.9%

TABLE 5.51. GRADE 5 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	763194	2491.3	93.8	46.6%	31.8%	21.6%
Female	373516	2490.7	89.4	47.2%	32.8%	20.1%
Male	389661	2491.9	97.8	46.1%	30.9%	23.0%
American Indian or Alaska Native	9821	2447.4	87.0	66.5%	25.1%	8.4%
Asian	59223	2557.1	92.6	21.0%	30.5%	48.5%
Black/African American	43939	2440.6	89.2	68.6%	23.6%	7.8%
Native Hawaiian or Pacific Islander	8604	2475.8	87.3	52.9%	32.0%	15.1%
Hispanic/Latino Ethnicity	260106	2456.3	83.3	63.3%	27.5%	9.2%
White	306903	2515.8	88.3	34.2%	36.8%	29.0%
Two or More Races	58155	2491.7	92.0	46.5%	32.1%	21.3%
Unidentified Race	16443	2520.6	85.5	32.4%	37.1%	30.6%
LEP Status	124573	2427.2	78.7	77.7%	18.4%	3.9%
IDEA Indicator	80450	2413.5	93.1	78.9%	14.8%	6.2%
Section 504 Status	8609	2497.6	90.5	43.9%	33.6%	22.5%
Economic Disadvantage Status	398993	2461.5	85.7	60.6%	28.3%	11.0%

TABLE 5.52. GRADE 6 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751146	2512.9	107.4	44.9%	34.1%	21.0%
Female	367617	2517.9	102.9	42.9%	35.8%	21.3%
Male	383506	2508.2	111.3	46.7%	32.6%	20.7%
American Indian or Alaska Native	9081	2460.0	101.8	65.0%	26.9%	8.0%
Asian	58440	2591.0	104.2	19.6%	31.0%	49.5%
Black/African American	43599	2457.7	102.6	66.4%	25.8%	7.8%
Native Hawaiian or Pacific Islander	8152	2488.8	102.0	53.4%	33.3%	13.3%
Hispanic/Latino Ethnicity	254805	2476.6	99.3	59.7%	30.4%	9.9%
White	304541	2537.9	99.5	33.7%	39.1%	27.2%
Two or More Races	55678	2504.4	106.4	48.1%	33.6%	18.3%
Unidentified Race	16850	2550.4	96.9	29.2%	39.3%	31.5%
LEP Status	94114	2423.0	94.4	81.6%	15.4%	3.0%
IDEA Indicator	75841	2404.2	105.9	82.8%	13.0%	4.2%
Section 504 Status	9334	2516.6	100.5	44.3%	35.5%	20.2%
Economic Disadvantage Status	387313	2478.0	101.0	58.7%	30.7%	10.6%

TABLE 5.53. GRADE 7 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	743128	2528.4	112.8	42.9%	34.1%	23.1%
Female	364145	2531.3	108.2	41.7%	35.6%	22.7%
Male	378929	2525.6	116.9	44.0%	32.5%	23.5%
American Indian or Alaska Native	9019	2478.8	103.5	61.2%	29.2%	9.5%
Asian	56725	2614.1	111.2	17.7%	27.8%	54.5%
Black/African American	43837	2469.0	105.7	65.0%	26.5%	8.4%
Native Hawaiian or Pacific Islander	7901	2496.2	107.1	53.4%	33.1%	13.5%
Hispanic/Latino Ethnicity	252164	2487.7	103.1	58.0%	31.2%	10.7%
White	303719	2555.8	104.2	31.5%	38.5%	30.0%
Two or More Races	53119	2523.3	110.4	44.5%	34.5%	21.1%
Unidentified Race	16644	2566.9	101.9	27.6%	38.4%	34.1%
LEP Status	82161	2428.6	98.0	81.6%	14.9%	3.5%
IDEA Indicator	72008	2416.8	105.4	81.9%	13.9%	4.2%
Section 504 Status	10258	2534.2	106.1	40.8%	35.9%	23.2%
Economic Disadvantage Status	378915	2491.3	104.9	56.3%	31.8%	11.8%

TABLE 5.54. GRADE 8 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	743207	2541.4	121.7	43.7%	33.9%	22.4%
Female	363401	2547.1	116.8	41.4%	35.7%	22.8%
Male	379779	2536.1	126.0	45.8%	32.1%	22.1%
American Indian or Alaska Native	8784	2485.8	108.8	63.4%	27.9%	8.7%
Asian	56876	2636.0	121.5	17.8%	28.2%	53.9%
Black/African American	44922	2478.1	110.0	66.0%	25.9%	8.1%
Native Hawaiian or Pacific Islander	7420	2510.8	115.4	53.2%	32.0%	14.7%
Hispanic/Latino Ethnicity	251308	2500.1	109.3	57.9%	31.4%	10.7%
White	304833	2567.9	115.4	33.3%	38.3%	28.5%
Two or More Races	52146	2537.1	119.8	45.5%	33.5%	21.1%
Unidentified Race	16918	2585.5	111.7	27.3%	38.3%	34.4%
LEP Status	74020	2440.5	102.8	80.2%	15.8%	4.0%
IDEA Indicator	70356	2426.5	104.7	83.0%	13.2%	3.8%
Section 504 Status	11238	2543.8	117.8	42.8%	35.5%	21.6%
Economic Disadvantage Status	373469	2503.3	112.6	56.9%	30.8%	12.2%

TABLE 5.55. GRADE 11 MATHEMATICS AVERAGE CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	558553	2560.2	130.7	49.2%	32.6%	18.1%
Female	274092	2567.5	123.1	46.5%	35.5%	18.1%
Male	284439	2553.1	137.2	51.9%	29.9%	18.2%
American Indian or Alaska Native	5888	2512.9	115.9	65.6%	26.9%	7.6%
Asian	47818	2665.4	133.6	20.7%	30.0%	49.3%
Black/African American	32028	2502.7	118.9	67.3%	25.8%	6.9%
Native Hawaiian or Pacific Islander	5750	2522.0	118.3	61.0%	30.3%	8.6%
Hispanic/Latino Ethnicity	215900	2527.2	116.6	59.4%	31.6%	9.1%
White	202762	2580.6	128.9	42.0%	35.4%	22.6%
Two or More Races	31353	2542.1	127.7	55.8%	29.9%	14.3%
Unidentified Race	17054	2610.1	116.9	30.7%	41.3%	27.9%
LEP Status	43310	2455.4	109.1	83.6%	13.2%	3.2%
IDEA Indicator	47638	2442.3	105.3	86.5%	11.0%	2.4%
Section 504 Status	8385	2560.0	126.7	49.5%	33.3%	17.2%
Economic Disadvantage Status	285091	2526.1	119.7	60.1%	30.2%	9.8%

TABLE 5.56. GRADE 3 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	781627	2416.0	92.3	32.6%	44.1%	23.3%
Female	382059	2414.8	90.5	33.1%	44.4%	22.4%
Male	399539	2417.2	94.1	32.0%	43.7%	24.2%
American Indian or Alaska Native	9651	2383.6	86.0	46.7%	42.2%	11.0%
Asian	56753	2466.6	89.7	15.6%	38.6%	45.8%
Black/African American	44612	2374.3	87.9	50.0%	40.5%	9.4%
Native Hawaiian or Pacific Islander	7564	2391.6	89.2	43.8%	41.3%	14.9%
Hispanic/Latino Ethnicity	276667	2382.2	86.8	46.8%	42.1%	11.0%
White	303904	2444.0	84.8	20.0%	47.3%	32.7%
Two or More Races	65864	2415.2	91.0	33.2%	44.1%	22.7%
Unidentified Race	16612	2440.4	83.6	21.2%	47.5%	31.2%
LEP Status	186551	2370.1	86.1	53.3%	38.4%	8.3%
IDEA Indicator	75660	2356.0	98.3	60.4%	29.4%	10.2%
Section 504 Status	5722	2418.5	93.5	31.2%	44.2%	24.6%
Economic Disadvantage Status	420509	2386.2	87.8	45.2%	42.2%	12.5%

TABLE 5.57. GRADE 4 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	765272	2450.3	103.7	33.5%	47.5%	18.9%
Female	374732	2450.3	101.0	33.2%	48.8%	18.0%
Male	390509	2450.3	106.3	33.8%	46.3%	19.8%
American Indian or Alaska Native	9442	2412.1	100.6	48.1%	43.8%	8.1%
Asian	57904	2507.9	98.6	15.7%	43.0%	41.2%
Black/African American	43694	2396.9	105.8	51.5%	42.0%	6.5%
Native Hawaiian or Pacific Islander	8466	2430.9	97.3	41.3%	47.3%	11.4%
Hispanic/Latino Ethnicity	266598	2414.2	97.0	48.4%	43.8%	7.8%
White	302312	2478.8	96.6	21.2%	52.1%	26.7%
Two or More Races	60413	2451.7	99.8	33.3%	48.5%	18.2%
Unidentified Race	16443	2476.6	90.8	23.0%	52.3%	24.7%
LEP Status	149456	2387.5	95.4	60.5%	35.7%	3.8%
IDEA Indicator	79400	2381.1	106.7	64.0%	29.4%	6.6%
Section 504 Status	7138	2457.0	99.1	31.1%	49.2%	19.7%
Economic Disadvantage Status	407588	2420.7	97.0	46.2%	44.6%	9.3%

TABLE 5.58. GRADE 5 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	763194	2472.4	120.0	41.2%	40.7%	18.2%
Female	373516	2473.8	116.6	40.8%	41.6%	17.5%
Male	389661	2471.1	123.2	41.5%	39.8%	18.7%
American Indian or Alaska Native	9821	2423.2	117.0	58.8%	34.7%	6.5%
Asian	59223	2538.6	111.1	20.4%	38.7%	40.9%
Black/African American	43939	2412.6	118.6	63.5%	30.7%	5.8%
Native Hawaiian or Pacific Islander	8604	2446.9	117.3	50.4%	38.0%	11.6%
Hispanic/Latino Ethnicity	260106	2426.7	115.5	58.3%	34.7%	7.0%
White	306903	2507.5	107.6	27.4%	47.4%	25.3%
Two or More Races	58155	2472.8	117.3	40.7%	41.7%	17.6%
Unidentified Race	16443	2503.2	107.1	28.9%	46.2%	24.9%
LEP Status	124573	2384.1	111.2	75.5%	22.0%	2.5%
IDEA Indicator	80450	2382.0	121.7	73.9%	21.1%	5.0%
Section 504 Status	8609	2483.6	113.1	37.0%	43.7%	19.3%
Economic Disadvantage Status	398993	2434.4	116.3	55.2%	36.1%	8.7%

TABLE 5.59. GRADE 6 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751146	2490.3	129.5	35.9%	47.1%	16.9%
Female	367617	2492.6	125.9	35.0%	48.5%	16.5%
Male	383506	2488.1	132.8	36.7%	45.8%	17.4%
American Indian or Alaska Native	9081	2432.8	127.1	53.7%	40.4%	5.9%
Asian	58440	2566.6	120.6	16.7%	42.6%	40.7%
Black/African American	43599	2421.1	130.7	56.6%	37.9%	5.5%
Native Hawaiian or Pacific Islander	8152	2463.3	123.1	44.6%	46.0%	9.3%
Hispanic/Latino Ethnicity	254805	2450.1	120.0	49.0%	44.5%	6.5%
White	304541	2520.2	123.7	25.4%	51.4%	23.3%
Two or More Races	55678	2486.3	126.9	37.1%	47.5%	15.4%
Unidentified Race	16850	2529.0	110.6	23.4%	53.5%	23.1%
LEP Status	94114	2389.9	116.2	70.6%	27.6%	1.8%
IDEA Indicator	75841	2379.7	124.2	73.0%	23.6%	3.3%
Section 504 Status	9334	2498.5	121.5	33.5%	49.6%	17.0%
Economic Disadvantage Status	387313	2455.1	120.1	47.7%	44.7%	7.5%

TABLE 5.60. GRADE 7 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	743128	2502.7	140.9	30.5%	49.9%	19.6%
Female	364145	2504.0	138.9	29.9%	50.8%	19.3%
Male	378929	2501.5	142.7	31.1%	49.0%	19.9%
American Indian or Alaska Native	9019	2448.3	134.4	43.5%	48.7%	7.8%
Asian	56725	2590.6	131.0	13.0%	40.5%	46.5%
Black/African American	43837	2434.6	134.7	47.5%	46.2%	6.3%
Native Hawaiian or Pacific Islander	7901	2468.5	133.6	38.8%	50.8%	10.4%
Hispanic/Latino Ethnicity	252164	2452.6	132.1	43.7%	48.4%	7.9%
White	303719	2538.4	131.9	20.4%	52.9%	26.6%
Two or More Races	53119	2502.1	137.0	30.4%	51.2%	18.4%
Unidentified Race	16644	2540.2	127.8	20.3%	52.9%	26.8%
LEP Status	82161	2390.4	123.0	64.3%	33.4%	2.3%
IDEA Indicator	72008	2389.3	127.1	65.2%	31.2%	3.6%
Section 504 Status	10258	2513.0	135.1	27.7%	51.7%	20.7%
Economic Disadvantage Status	378915	2459.7	132.7	41.8%	49.0%	9.2%

TABLE 5.61. GRADE 8 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	743207	2522.0	145.3	28.0%	52.3%	19.7%
Female	363401	2524.1	142.8	26.8%	53.8%	19.4%
Male	379779	2520.0	147.5	29.2%	50.8%	20.0%
American Indian or Alaska Native	8784	2469.7	134.1	40.5%	51.5%	8.0%
Asian	56876	2611.4	141.8	12.6%	41.0%	46.4%
Black/African American	44922	2454.4	134.8	43.8%	49.9%	6.3%
Native Hawaiian or Pacific Islander	7420	2482.2	138.7	36.9%	52.3%	10.8%
Hispanic/Latino Ethnicity	251308	2471.5	133.6	39.6%	52.4%	8.0%
White	304833	2558.0	138.0	19.0%	54.4%	26.6%
Two or More Races	52146	2518.4	143.6	28.2%	52.9%	18.8%
Unidentified Race	16918	2557.8	133.7	18.4%	54.8%	26.8%
LEP Status	74020	2405.9	124.2	60.4%	37.1%	2.5%
IDEA Indicator	70356	2405.7	125.3	61.6%	35.2%	3.1%
Section 504 Status	11238	2528.9	140.6	26.3%	53.1%	20.6%
Economic Disadvantage Status	373469	2478.6	135.9	38.2%	52.0%	9.8%

TABLE 5.62. GRADE 11 MATHEMATICS AVERAGE CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	558553	2530.8	158.3	33.8%	50.9%	15.2%
Female	274092	2531.0	153.6	33.0%	52.9%	14.1%
Male	284439	2530.6	162.7	34.6%	49.1%	16.3%
American Indian or Alaska Native	5888	2488.0	149.7	42.1%	50.3%	7.7%
Asian	47818	2633.4	155.7	15.1%	45.2%	39.7%
Black/African American	32028	2461.3	143.0	50.0%	45.3%	4.6%
Native Hawaiian or Pacific Islander	5750	2485.1	150.8	42.7%	49.8%	7.5%
Hispanic/Latino Ethnicity	215900	2488.5	144.3	43.3%	49.9%	6.8%
White	202762	2563.0	156.3	25.9%	53.7%	20.4%
Two or More Races	31353	2520.0	158.8	35.2%	51.4%	13.4%
Unidentified Race	17054	2575.8	146.0	22.4%	56.6%	21.0%
LEP Status	43310	2412.6	129.3	65.3%	32.3%	2.4%
IDEA Indicator	47638	2409.0	128.5	65.1%	32.5%	2.4%
Section 504 Status	8385	2537.9	156.5	31.7%	52.5%	15.7%
Economic Disadvantage Status	285091	2491.2	148.4	42.5%	49.6%	7.9%

TABLE 5.63. GRADE 3 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	781627	2415.7	96.1	24.3%	52.9%	22.8%
Female	382059	2417.9	93.9	23.1%	54.2%	22.7%
Male	399539	2413.7	98.0	25.5%	51.6%	22.8%
American Indian or Alaska Native	9651	2377.1	90.7	37.9%	52.1%	10.1%
Asian	56753	2470.2	95.0	10.9%	42.7%	46.4%
Black/African American	44612	2372.4	93.3	39.5%	50.8%	9.7%
Native Hawaiian or Pacific Islander	7564	2397.3	91.3	31.3%	52.9%	15.8%
Hispanic/Latino Ethnicity	276667	2385.8	88.5	33.8%	54.8%	11.4%
White	303904	2439.5	91.9	16.0%	53.4%	30.7%
Two or More Races	65864	2415.2	95.7	24.1%	53.3%	22.6%
Unidentified Race	16612	2443.1	88.4	15.2%	52.2%	32.5%
LEP Status	186551	2374.7	88.6	37.7%	53.5%	8.8%
IDEA Indicator	75660	2362.3	95.7	48.1%	42.0%	9.9%
Section 504 Status	5722	2419.6	96.5	24.3%	50.9%	24.8%
Economic Disadvantage Status	420509	2389.8	89.5	33.0%	54.0%	13.0%

TABLE 5.64. GRADE 4 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	765272	2455.7	95.1	33.4%	45.7%	20.9%
Female	374732	2455.8	93.0	32.9%	46.8%	20.2%
Male	390509	2455.5	97.2	33.9%	44.5%	21.6%
American Indian or Alaska Native	9442	2416.9	90.7	49.0%	41.7%	9.3%
Asian	57904	2513.9	92.6	14.9%	39.5%	45.6%
Black/African American	43694	2410.1	93.2	48.8%	43.1%	8.0%
Native Hawaiian or Pacific Islander	8466	2440.0	85.6	40.0%	46.1%	13.9%
Hispanic/Latino Ethnicity	266598	2423.2	85.8	47.5%	43.4%	9.2%
White	302312	2480.0	91.4	22.1%	49.2%	28.7%
Two or More Races	60413	2455.4	93.6	33.9%	45.6%	20.5%
Unidentified Race	16443	2481.6	85.4	23.0%	48.5%	28.5%
LEP Status	149456	2402.2	82.7	58.2%	37.1%	4.7%
IDEA Indicator	79400	2396.3	91.0	63.0%	29.7%	7.3%
Section 504 Status	7138	2457.6	93.5	32.5%	45.8%	21.6%
Economic Disadvantage Status	407588	2428.1	86.4	46.1%	43.1%	10.8%

TABLE 5.65. GRADE 5 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	763194	2478.6	109.8	36.3%	47.4%	16.3%
Female	373516	2480.1	107.6	35.6%	48.4%	16.0%
Male	389661	2477.2	111.9	37.0%	46.4%	16.6%
American Indian or Alaska Native	9821	2433.6	102.3	53.0%	41.2%	5.8%
Asian	59223	2543.0	105.5	16.8%	45.6%	37.7%
Black/African American	43939	2429.1	102.4	54.1%	40.5%	5.4%
Native Hawaiian or Pacific Islander	8604	2456.8	103.1	43.7%	46.8%	9.5%
Hispanic/Latino Ethnicity	260106	2439.2	99.6	51.2%	42.7%	6.1%
White	306903	2507.6	105.1	24.8%	52.5%	22.7%
Two or More Races	58155	2477.3	109.6	36.1%	48.3%	15.7%
Unidentified Race	16443	2507.1	101.3	25.6%	51.9%	22.5%
LEP Status	124573	2405.8	93.4	66.0%	32.0%	2.1%
IDEA Indicator	80450	2405.6	102.8	66.9%	28.7%	4.4%
Section 504 Status	8609	2484.4	107.2	33.6%	49.7%	16.7%
Economic Disadvantage Status	398993	2444.9	101.6	48.9%	43.7%	7.4%

TABLE 5.66. GRADE 6 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	751146	2502.0	118.3	31.4%	51.3%	17.3%
Female	367617	2506.5	114.8	29.5%	53.0%	17.5%
Male	383506	2497.7	121.3	33.2%	49.7%	17.2%
American Indian or Alaska Native	9081	2451.0	112.7	47.3%	46.6%	6.2%
Asian	58440	2575.0	115.0	14.2%	44.2%	41.5%
Black/African American	43599	2446.6	114.9	47.4%	46.7%	5.9%
Native Hawaiian or Pacific Islander	8152	2475.4	109.1	40.3%	50.5%	9.2%
Hispanic/Latino Ethnicity	254805	2466.9	106.5	42.7%	49.9%	7.3%
White	304541	2527.1	116.0	22.3%	54.6%	23.1%
Two or More Races	55678	2493.8	117.2	34.1%	50.7%	15.3%
Unidentified Race	16850	2537.9	106.0	20.6%	54.5%	24.9%
LEP Status	94114	2417.9	98.6	62.0%	36.1%	1.9%
IDEA Indicator	75841	2411.3	105.7	66.8%	29.8%	3.3%
Section 504 Status	9334	2504.6	113.6	30.1%	53.5%	16.4%
Economic Disadvantage Status	387313	2468.9	107.5	42.6%	49.3%	8.1%

TABLE 5.67. GRADE 7 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	743128	2508.1	137.7	19.3%	62.2%	18.5%
Female	364145	2515.2	133.5	17.2%	64.0%	18.8%
Male	378929	2501.4	141.2	21.3%	60.5%	18.2%
American Indian or Alaska Native	9019	2449.0	133.3	31.3%	61.7%	7.1%
Asian	56725	2596.0	130.1	7.9%	47.1%	45.0%
Black/African American	43837	2444.8	134.6	32.6%	61.0%	6.5%
Native Hawaiian or Pacific Islander	7901	2477.0	127.3	23.3%	67.0%	9.7%
Hispanic/Latino Ethnicity	252164	2470.5	124.9	24.4%	67.6%	8.0%
White	303719	2533.1	136.7	15.4%	60.2%	24.5%
Two or More Races	53119	2504.1	134.4	18.8%	64.6%	16.6%
Unidentified Race	16644	2550.3	121.4	10.6%	63.4%	26.0%
LEP Status	82161	2417.5	117.3	37.4%	60.2%	2.4%
IDEA Indicator	72008	2410.5	120.0	40.3%	56.6%	3.1%
Section 504 Status	10258	2513.0	133.9	17.9%	63.2%	18.9%
Economic Disadvantage Status	378915	2474.0	125.2	23.5%	67.7%	8.8%

TABLE 5.68. GRADE 8 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	743207	2534.2	131.2	31.1%	51.2%	17.7%
Female	363401	2542.8	126.8	28.1%	53.4%	18.5%
Male	379779	2525.9	134.9	34.0%	49.1%	16.9%
American Indian or Alaska Native	8784	2479.9	122.5	46.0%	47.5%	6.5%
Asian	56876	2626.4	128.4	12.3%	42.1%	45.6%
Black/African American	44922	2477.8	121.7	46.6%	47.3%	6.2%
Native Hawaiian or Pacific Islander	7420	2505.3	120.9	39.3%	50.4%	10.4%
Hispanic/Latino Ethnicity	251308	2495.7	117.4	41.5%	50.6%	7.9%
White	304833	2557.6	130.1	23.7%	53.8%	22.5%
Two or More Races	52146	2529.8	129.3	31.8%	51.5%	16.7%
Unidentified Race	16918	2577.6	117.4	18.7%	54.9%	26.4%
LEP Status	74020	2438.8	111.7	60.4%	36.9%	2.8%
IDEA Indicator	70356	2425.4	110.8	68.3%	29.0%	2.7%
Section 504 Status	11238	2535.5	126.2	30.3%	53.0%	16.7%
Economic Disadvantage Status	373469	2498.7	119.6	41.4%	49.4%	9.2%

TABLE 5.69. GRADE 11 MATHEMATICS AVERAGE CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 ADMINISTRATION

Group	Scale Scores			Reporting Categories		
	N	M	SD	Below Standard	At/Near Standard	Above Standard
Total	558553	2555.3	140.8	30.0%	55.4%	14.5%
Female	274092	2559.2	136.1	28.0%	58.1%	13.9%
Male	284439	2551.5	145.1	32.0%	52.8%	15.2%
American Indian or Alaska Native	5888	2512.0	131.5	40.8%	52.6%	6.5%
Asian	47818	2654.3	141.3	12.1%	47.4%	40.5%
Black/African American	32028	2502.7	126.3	43.3%	51.9%	4.9%
Native Hawaiian or Pacific Islander	5750	2521.9	127.0	38.2%	55.4%	6.5%
Hispanic/Latino Ethnicity	215900	2521.7	127.6	36.5%	56.7%	6.8%
White	202762	2576.6	141.2	25.2%	56.4%	18.4%
Two or More Races	31353	2541.5	139.7	33.5%	54.5%	12.0%
Unidentified Race	17054	2599.5	132.5	18.3%	59.9%	21.8%
LEP Status	43310	2467.1	117.7	51.5%	46.1%	2.4%
IDEA Indicator	47638	2459.3	114.2	57.0%	40.9%	2.1%
Section 504 Status	8385	2553.4	139.4	30.5%	55.8%	13.6%
Economic Disadvantage Status	285091	2522.5	130.4	37.0%	55.4%	7.6%

Percentile Tables for Overall Scale Scores

Table 5.70 through Table 5.76 present the overall ELA/literacy scale score for the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th for grades 3 through 8 and 11. Table 5.77 through Table 5.83 present the decile information 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.

Percentile Tables for Claim-level Scale Scores

- Table 5.84 through Table 5.90 present the decile information for the Claim 1 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5.91 through Table 5.97 present the decile information for the Claim 2 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5.98 through Table 5.104 present the decile information for the Claim 3 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5.105 through Table 5.111 present the decile information for the Claim 4 ELA/literacy scale scores for grades 3 through 8 and 11.
- Table 5.112 through Table 5.118 present the decile information for the Claim 1 mathematics scale scores for grades 3 through 8 and 11.

- Table 5.119 through Table 5.125 present the decile information for the Claim 2/4 mathematics scale scores for grades 3 through 8 and 11.
- Table 5.126 through Table 5.132 present the decile information for the Claim 3 mathematics scale scores for grades 3 through 8 and 11.

TABLE 5.70. GRADE 3 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

		Percentiles								
		10	20	30	40	50	60	70	80	90
Total		2298	2334	2362	2388	2412	2437	2463	2491	2528
Female		2308	2345	2374	2400	2424	2447	2472	2500	2536
Male		2291	2325	2352	2377	2401	2426	2453	2482	2519
American Indian or Alaska Native		2279	2308	2330	2351	2371	2392	2415	2441	2479
Asian		2340	2384	2415	2441	2465	2487	2510	2535	2568
Black/African American		2276	2304	2327	2348	2370	2393	2417	2447	2485
Native Hawaiian or Pacific Islander		2288	2319	2345	2368	2389	2410	2433	2460	2496
Hispanic/Latino Ethnicity		2281	2310	2334	2356	2377	2398	2421	2448	2485
White		2328	2367	2397	2422	2445	2466	2489	2513	2545
Two or More Races		2300	2335	2363	2388	2411	2435	2460	2489	2526
Unidentified Race		2333	2372	2401	2424	2445	2466	2488	2511	2543
LEP Status		2273	2300	2321	2340	2359	2379	2400	2425	2460
IDEA Indicator		2254	2279	2298	2317	2338	2361	2388	2422	2471
Section 504 Status		2303	2337	2367	2392	2415	2439	2464	2492	2529
Economic Disadvantage Status		2283	2313	2337	2359	2380	2402	2425	2453	2489

TABLE 5.71. GRADE 4 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2333	2371	2402	2430	2457	2482	2508	2537	2575
Female	2344	2384	2415	2443	2468	2493	2519	2547	2584
Male	2323	2360	2390	2418	2445	2471	2498	2527	2565
American Indian or Alaska Native	2310	2338	2363	2386	2407	2431	2455	2485	2524
Asian	2380	2428	2464	2491	2515	2538	2561	2587	2621
Black/African American	2305	2335	2360	2383	2407	2432	2459	2489	2528
Native Hawaiian or Pacific Islander	2321	2354	2382	2407	2429	2454	2480	2508	2545
Hispanic/Latino Ethnicity	2314	2345	2370	2394	2417	2440	2465	2493	2531
White	2363	2406	2438	2464	2488	2510	2532	2557	2591
Two or More Races	2335	2372	2401	2428	2453	2479	2503	2533	2571
Unidentified Race	2371	2412	2444	2470	2493	2513	2535	2559	2593
LEP Status	2299	2326	2346	2364	2383	2402	2423	2448	2482
IDEA Indicator	2281	2307	2326	2345	2365	2388	2416	2453	2503
Section 504 Status	2341	2376	2404	2430	2456	2482	2505	2534	2572
Economic Disadvantage Status	2316	2348	2374	2398	2421	2445	2470	2498	2535

TABLE 5.72. GRADE 5 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2372	2410	2442	2471	2498	2524	2550	2579	2616
Female	2387	2426	2458	2486	2512	2537	2562	2590	2627
Male	2360	2397	2427	2456	2483	2510	2537	2566	2604
American Indian or Alaska Native	2344	2374	2398	2420	2442	2465	2491	2521	2560
Asian	2416	2468	2504	2533	2557	2581	2604	2629	2663
Black/African American	2342	2374	2398	2422	2446	2471	2499	2529	2568
Native Hawaiian or Pacific Islander	2361	2393	2422	2447	2472	2497	2522	2548	2585
Hispanic/Latino Ethnicity	2354	2386	2411	2435	2458	2482	2509	2537	2574
White	2402	2445	2477	2504	2528	2550	2572	2598	2631
Two or More Races	2375	2411	2441	2468	2494	2519	2544	2572	2610
Unidentified Race	2409	2454	2485	2510	2533	2555	2578	2602	2635
LEP Status	2332	2359	2379	2396	2413	2430	2450	2475	2509
IDEA Indicator	2313	2340	2360	2378	2397	2417	2443	2478	2528
Section 504 Status	2381	2417	2447	2474	2499	2522	2547	2575	2612
Economic Disadvantage Status	2355	2388	2413	2437	2461	2486	2512	2540	2576

TABLE 5.73. GRADE 6 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2396	2437	2469	2495	2520	2544	2570	2599	2636
Female	2414	2455	2485	2511	2534	2557	2582	2610	2645
Male	2383	2423	2453	2480	2505	2530	2556	2586	2624
American Indian or Alaska Native	2362	2396	2422	2445	2467	2490	2514	2543	2582
Asian	2447	2497	2531	2558	2582	2604	2626	2650	2682
Black/African American	2362	2397	2424	2448	2472	2496	2522	2551	2590
Native Hawaiian or Pacific Islander	2378	2415	2445	2470	2493	2516	2539	2567	2602
Hispanic/Latino Ethnicity	2377	2414	2441	2465	2488	2509	2532	2558	2595
White	2423	2466	2497	2522	2545	2567	2591	2616	2649
Two or More Races	2394	2435	2466	2492	2516	2540	2564	2593	2630
Unidentified Race	2440	2482	2512	2535	2557	2578	2600	2624	2654
LEP Status	2344	2373	2394	2412	2430	2448	2467	2490	2521
IDEA Indicator	2331	2359	2380	2399	2418	2439	2462	2491	2535
Section 504 Status	2408	2445	2473	2496	2518	2539	2563	2591	2629
Economic Disadvantage Status	2378	2414	2442	2466	2489	2511	2534	2561	2598

TABLE 5.74. GRADE 7 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2410	2453	2486	2516	2544	2570	2597	2626	2664
Female	2429	2472	2505	2534	2560	2585	2609	2638	2674
Male	2397	2437	2469	2499	2527	2555	2582	2613	2652
American Indian or Alaska Native	2381	2414	2441	2465	2490	2516	2542	2573	2613
Asian	2464	2518	2556	2585	2609	2632	2654	2678	2710
Black/African American	2379	2413	2440	2466	2491	2517	2545	2576	2616
Native Hawaiian or Pacific Islander	2386	2423	2453	2481	2507	2531	2558	2587	2624
Hispanic/Latino Ethnicity	2392	2427	2454	2479	2504	2529	2555	2583	2620
White	2441	2487	2521	2549	2573	2596	2619	2645	2678
Two or More Races	2412	2453	2486	2515	2542	2567	2592	2621	2659
Unidentified Race	2454	2499	2532	2559	2582	2603	2623	2647	2678
LEP Status	2358	2384	2404	2421	2438	2455	2474	2498	2534
IDEA Indicator	2350	2376	2396	2414	2432	2452	2476	2507	2554
Section 504 Status	2422	2460	2491	2518	2541	2565	2592	2621	2660
Economic Disadvantage Status	2392	2427	2456	2481	2506	2532	2558	2586	2623

TABLE 5.75. GRADE 8 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2433	2475	2507	2535	2562	2589	2616	2644	2681
Female	2455	2496	2527	2554	2580	2605	2629	2656	2692
Male	2417	2458	2489	2517	2544	2572	2600	2630	2669
American Indian or Alaska Native	2399	2434	2461	2486	2510	2535	2561	2593	2633
Asian	2487	2538	2574	2603	2627	2649	2672	2696	2728
Black/African American	2398	2434	2462	2488	2512	2537	2564	2596	2636
Native Hawaiian or Pacific Islander	2412	2449	2478	2504	2528	2551	2578	2610	2644
Hispanic/Latino Ethnicity	2416	2452	2480	2504	2527	2550	2575	2604	2640
White	2458	2503	2536	2565	2591	2614	2637	2662	2696
Two or More Races	2434	2475	2506	2534	2560	2585	2612	2641	2678
Unidentified Race	2478	2522	2553	2578	2600	2620	2641	2663	2694
LEP Status	2377	2405	2425	2443	2460	2477	2495	2517	2551
IDEA Indicator	2368	2396	2416	2434	2453	2472	2495	2523	2569
Section 504 Status	2441	2481	2509	2535	2560	2586	2612	2640	2678
Economic Disadvantage Status	2414	2452	2479	2504	2528	2552	2578	2607	2644

TABLE 5.76. GRADE 11 ELA/LITERACY OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2445	2501	2542	2577	2607	2636	2664	2695	2734
Female	2472	2525	2564	2595	2623	2649	2675	2704	2741
Male	2426	2479	2521	2558	2591	2621	2652	2685	2726
American Indian or Alaska Native	2417	2462	2499	2529	2560	2591	2621	2654	2696
Asian	2499	2565	2608	2641	2669	2694	2719	2747	2783
Black/African American	2407	2451	2488	2521	2553	2584	2615	2648	2690
Native Hawaiian or Pacific Islander	2416	2463	2499	2530	2559	2589	2616	2648	2688
Hispanic/Latino Ethnicity	2427	2475	2513	2544	2573	2600	2627	2657	2696
White	2470	2530	2573	2606	2634	2660	2686	2713	2748
Two or More Races	2448	2500	2540	2573	2602	2629	2657	2687	2726
Unidentified Race	2506	2558	2595	2622	2646	2667	2689	2713	2745
LEP Status	2375	2405	2427	2448	2468	2489	2512	2539	2576
IDEA Indicator	2375	2406	2430	2453	2477	2502	2530	2566	2618
Section 504 Status	2465	2517	2556	2589	2617	2644	2670	2699	2736
Economic Disadvantage Status	2425	2473	2510	2543	2573	2601	2629	2660	2700

TABLE 5.77. GRADE 3 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2322	2357	2382	2404	2424	2444	2465	2489	2523
Female	2326	2359	2382	2403	2423	2443	2463	2486	2519
Male	2318	2356	2382	2405	2426	2446	2467	2492	2526
American Indian or Alaska Native	2295	2329	2352	2371	2390	2407	2427	2448	2480
Asian	2372	2410	2437	2460	2480	2499	2520	2544	2578
Black/African American	2286	2321	2346	2365	2384	2402	2423	2445	2476
Native Hawaiian or Pacific Islander	2307	2344	2368	2388	2405	2424	2444	2468	2500
Hispanic/Latino Ethnicity	2305	2337	2359	2377	2394	2412	2432	2454	2483
White	2350	2386	2411	2431	2449	2467	2485	2507	2537
Two or More Races	2321	2357	2381	2402	2423	2442	2463	2487	2521
Unidentified Race	2358	2390	2414	2434	2451	2468	2486	2506	2535
LEP Status	2297	2329	2350	2367	2383	2400	2419	2442	2473
IDEA Indicator	2246	2288	2316	2339	2360	2382	2408	2438	2478
Section 504 Status	2321	2356	2381	2404	2426	2447	2468	2492	2526
Economic Disadvantage Status	2304	2339	2361	2380	2398	2416	2436	2458	2489

TABLE 5.78. GRADE 4 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2360	2393	2419	2441	2462	2483	2505	2532	2568
Female	2364	2395	2419	2441	2461	2481	2502	2528	2563
Male	2356	2391	2418	2442	2463	2485	2509	2536	2573
American Indian or Alaska Native	2334	2364	2387	2407	2425	2444	2465	2488	2523
Asian	2411	2450	2478	2502	2524	2546	2568	2592	2624
Black/African American	2324	2357	2380	2400	2418	2438	2458	2483	2515
Native Hawaiian or Pacific Islander	2352	2385	2407	2428	2447	2466	2486	2508	2540
Hispanic/Latino Ethnicity	2343	2372	2393	2411	2429	2447	2467	2490	2521
White	2389	2424	2449	2470	2489	2508	2527	2550	2582
Two or More Races	2362	2395	2420	2441	2462	2482	2504	2530	2566
Unidentified Race	2394	2427	2451	2470	2489	2507	2527	2550	2579
LEP Status	2329	2357	2376	2392	2407	2423	2441	2463	2495
IDEA Indicator	2298	2329	2351	2371	2390	2411	2435	2466	2510
Section 504 Status	2369	2401	2424	2445	2465	2487	2508	2533	2569
Economic Disadvantage Status	2344	2374	2396	2415	2434	2453	2473	2496	2529

TABLE 5.79. GRADE 5 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2377	2412	2439	2464	2489	2514	2540	2570	2607
Female	2382	2416	2441	2465	2488	2511	2537	2566	2603
Male	2371	2408	2437	2464	2490	2516	2543	2573	2611
American Indian or Alaska Native	2348	2381	2404	2425	2443	2463	2488	2515	2553
Asian	2430	2475	2509	2536	2561	2583	2606	2631	2666
Black/African American	2338	2372	2396	2416	2436	2457	2481	2509	2549
Native Hawaiian or Pacific Islander	2368	2402	2428	2448	2470	2492	2517	2546	2578
Hispanic/Latino Ethnicity	2358	2389	2411	2430	2449	2470	2493	2520	2557
White	2407	2446	2474	2498	2520	2542	2564	2589	2621
Two or More Races	2379	2414	2441	2465	2489	2513	2538	2568	2605
Unidentified Race	2410	2446	2475	2499	2521	2543	2565	2589	2620
LEP Status	2338	2368	2387	2403	2419	2435	2453	2477	2513
IDEA Indicator	2314	2345	2367	2386	2405	2424	2448	2481	2531
Section 504 Status	2388	2422	2449	2473	2496	2519	2544	2572	2611
Economic Disadvantage Status	2360	2392	2415	2435	2456	2477	2501	2528	2566

TABLE 5.80. GRADE 6 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2376	2428	2463	2490	2515	2540	2566	2596	2638
Female	2388	2436	2469	2495	2519	2542	2567	2596	2636
Male	2366	2420	2457	2486	2512	2537	2565	2596	2639
American Indian or Alaska Native	2336	2378	2412	2440	2466	2490	2514	2540	2577
Asian	2455	2504	2538	2567	2592	2616	2642	2671	2709
Black/African American	2330	2375	2408	2435	2460	2484	2509	2537	2577
Native Hawaiian or Pacific Islander	2358	2408	2442	2470	2493	2515	2538	2565	2601
Hispanic/Latino Ethnicity	2352	2398	2431	2457	2480	2502	2525	2552	2588
White	2414	2464	2495	2520	2543	2565	2588	2615	2651
Two or More Races	2371	2420	2455	2483	2508	2531	2557	2588	2629
Unidentified Race	2428	2475	2505	2529	2550	2572	2595	2621	2656
LEP Status	2310	2351	2379	2403	2425	2446	2467	2492	2527
IDEA Indicator	2281	2321	2350	2375	2400	2426	2455	2488	2537
Section 504 Status	2395	2440	2469	2494	2516	2540	2564	2594	2634
Economic Disadvantage Status	2351	2398	2432	2458	2482	2504	2528	2555	2592

TABLE 5.81. GRADE 7 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2384	2436	2472	2503	2530	2558	2587	2620	2664
Female	2395	2444	2478	2507	2533	2559	2587	2619	2662
Male	2375	2428	2466	2498	2527	2556	2587	2621	2667
American Indian or Alaska Native	2348	2394	2427	2454	2480	2506	2531	2561	2604
Asian	2466	2522	2561	2593	2620	2645	2673	2701	2740
Black/African American	2336	2383	2416	2445	2470	2495	2522	2554	2597
Native Hawaiian or Pacific Islander	2360	2411	2445	2473	2500	2527	2553	2582	2621
Hispanic/Latino Ethnicity	2358	2404	2437	2464	2489	2513	2538	2569	2609
White	2424	2475	2509	2536	2561	2586	2611	2640	2678
Two or More Races	2385	2435	2470	2499	2526	2552	2581	2613	2656
Unidentified Race	2437	2486	2518	2545	2570	2593	2617	2645	2682
LEP Status	2313	2353	2381	2405	2427	2448	2472	2500	2542
IDEA Indicator	2293	2334	2362	2386	2410	2435	2463	2498	2549
Section 504 Status	2403	2449	2481	2509	2534	2561	2588	2621	2664
Economic Disadvantage Status	2359	2406	2440	2467	2493	2517	2543	2574	2615

TABLE 5.82. GRADE 8 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2391	2442	2480	2512	2542	2572	2605	2643	2693
Female	2405	2454	2489	2519	2547	2576	2608	2644	2692
Male	2379	2431	2470	2504	2536	2568	2602	2642	2695
American Indian or Alaska Native	2359	2399	2432	2460	2485	2511	2540	2574	2624
Asian	2472	2534	2577	2613	2645	2674	2704	2737	2782
Black/African American	2348	2391	2424	2453	2479	2504	2532	2567	2615
Native Hawaiian or Pacific Islander	2367	2414	2452	2484	2512	2542	2570	2605	2649
Hispanic/Latino Ethnicity	2367	2412	2445	2474	2500	2526	2553	2586	2632
White	2425	2478	2515	2546	2574	2602	2631	2664	2708
Two or More Races	2391	2440	2477	2508	2536	2567	2600	2637	2687
Unidentified Race	2444	2497	2533	2561	2588	2615	2642	2672	2712
LEP Status	2325	2362	2388	2411	2434	2456	2481	2512	2560
IDEA Indicator	2310	2346	2371	2394	2416	2440	2468	2502	2557
Section 504 Status	2405	2450	2485	2515	2543	2572	2603	2641	2693
Economic Disadvantage Status	2367	2412	2446	2475	2502	2529	2559	2593	2642

TABLE 5.83. GRADE 11 MATHEMATICS OVERALL SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2401	2452	2490	2525	2558	2591	2626	2667	2724
Female	2416	2465	2501	2534	2565	2595	2628	2665	2718
Male	2389	2439	2478	2516	2551	2587	2624	2669	2732
American Indian or Alaska Native	2382	2424	2456	2484	2511	2541	2572	2611	2662
Asian	2484	2553	2600	2638	2673	2706	2739	2776	2822
Black/African American	2363	2407	2439	2469	2497	2527	2559	2597	2647
Native Hawaiian or Pacific Islander	2386	2430	2461	2494	2524	2552	2583	2614	2665
Hispanic/Latino Ethnicity	2387	2432	2466	2495	2524	2553	2583	2618	2666
White	2421	2476	2519	2555	2587	2619	2651	2690	2741
Two or More Races	2395	2441	2477	2510	2542	2574	2608	2648	2708
Unidentified Race	2457	2513	2552	2586	2615	2641	2670	2701	2744
LEP Status	2336	2378	2405	2427	2448	2469	2494	2526	2578
IDEA Indicator	2327	2369	2396	2418	2438	2460	2483	2517	2568
Section 504 Status	2408	2457	2495	2529	2560	2590	2624	2663	2722
Economic Disadvantage Status	2384	2430	2463	2494	2523	2553	2584	2621	2673

TABLE 5.84. GRADE 3 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2284	2325	2354	2381	2408	2436	2465	2496	2537
Female	2295	2335	2366	2393	2421	2447	2475	2505	2545
Male	2276	2315	2344	2370	2396	2424	2454	2486	2528
American Indian or Alaska Native	2262	2301	2326	2347	2369	2391	2416	2446	2489
Asian	2324	2369	2404	2434	2460	2484	2508	2535	2573
Black/African American	2258	2296	2322	2344	2366	2390	2417	2449	2493
Native Hawaiian or Pacific Islander	2267	2306	2333	2355	2376	2401	2427	2459	2501
Hispanic/Latino Ethnicity	2265	2303	2328	2351	2372	2395	2421	2452	2494
White	2311	2355	2388	2417	2443	2468	2492	2519	2557
Two or More Races	2285	2325	2354	2381	2407	2434	2463	2494	2536
Unidentified Race	2314	2357	2389	2416	2441	2464	2488	2515	2553
LEP Status	2255	2293	2317	2337	2356	2376	2399	2427	2468
IDEA Indicator	2236	2279	2303	2323	2343	2364	2390	2425	2478
Section 504 Status	2285	2327	2358	2386	2412	2439	2468	2499	2541
Economic Disadvantage Status	2266	2304	2330	2352	2375	2398	2425	2456	2498

TABLE 5.85. GRADE 4 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2312	2356	2389	2420	2450	2481	2511	2545	2588
Female	2323	2367	2401	2432	2463	2492	2522	2555	2597
Male	2302	2346	2378	2408	2438	2469	2501	2535	2578
American Indian or Alaska Native	2282	2325	2354	2379	2403	2429	2459	2493	2538
Asian	2360	2411	2449	2481	2509	2535	2562	2591	2631
Black/African American	2273	2321	2350	2375	2400	2428	2458	2494	2540
Native Hawaiian or Pacific Islander	2295	2339	2369	2394	2419	2445	2476	2511	2554
Hispanic/Latino Ethnicity	2289	2331	2359	2383	2407	2433	2462	2497	2541
White	2341	2390	2427	2459	2487	2513	2539	2568	2607
Two or More Races	2313	2356	2388	2417	2446	2476	2507	2542	2585
Unidentified Race	2350	2392	2427	2457	2484	2510	2535	2564	2604
LEP Status	2270	2310	2335	2356	2375	2395	2417	2446	2488
IDEA Indicator	2258	2298	2324	2346	2367	2389	2415	2454	2512
Section 504 Status	2315	2361	2393	2422	2452	2483	2512	2545	2588
Economic Disadvantage Status	2292	2333	2362	2386	2411	2438	2468	2502	2547

TABLE 5.86. GRADE 5 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2351	2393	2426	2457	2487	2517	2546	2580	2624
Female	2364	2408	2442	2473	2502	2530	2559	2592	2635
Male	2340	2381	2412	2442	2473	2503	2533	2567	2611
American Indian or Alaska Native	2320	2361	2388	2412	2437	2463	2492	2528	2571
Asian	2393	2445	2484	2515	2542	2567	2594	2624	2663
Black/African American	2322	2361	2389	2412	2437	2465	2495	2529	2575
Native Hawaiian or Pacific Islander	2333	2374	2401	2427	2454	2481	2512	2544	2588
Hispanic/Latino Ethnicity	2328	2367	2395	2419	2444	2470	2499	2532	2575
White	2381	2429	2465	2496	2523	2549	2575	2604	2644
Two or More Races	2353	2393	2425	2455	2484	2513	2542	2575	2620
Unidentified Race	2381	2427	2461	2490	2515	2539	2565	2595	2634
LEP Status	2302	2341	2364	2384	2402	2420	2441	2468	2508
IDEA Indicator	2293	2333	2357	2377	2396	2416	2441	2477	2533
Section 504 Status	2358	2400	2433	2464	2492	2519	2548	2580	2625
Economic Disadvantage Status	2329	2369	2397	2422	2448	2475	2504	2537	2581

TABLE 5.87. GRADE 6 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2345	2397	2433	2466	2497	2527	2560	2596	2641
Female	2359	2411	2448	2480	2510	2539	2570	2604	2648
Male	2334	2384	2420	2452	2483	2515	2548	2586	2634
American Indian or Alaska Native	2302	2356	2390	2418	2445	2473	2503	2541	2586
Asian	2398	2456	2497	2531	2560	2588	2617	2648	2688
Black/African American	2310	2360	2391	2418	2444	2472	2503	2541	2590
Native Hawaiian or Pacific Islander	2320	2368	2402	2431	2459	2488	2521	2555	2606
Hispanic/Latino Ethnicity	2320	2369	2402	2430	2457	2484	2514	2548	2596
White	2377	2430	2468	2500	2529	2558	2587	2618	2659
Two or More Races	2338	2391	2428	2460	2491	2521	2553	2589	2635
Unidentified Race	2381	2435	2473	2502	2531	2558	2587	2618	2661
LEP Status	2273	2327	2356	2380	2401	2422	2444	2471	2511
IDEA Indicator	2268	2322	2353	2379	2401	2423	2448	2482	2535
Section 504 Status	2355	2405	2439	2470	2500	2529	2560	2595	2639
Economic Disadvantage Status	2319	2369	2403	2431	2458	2486	2517	2552	2600

TABLE 5.88. GRADE 7 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2384	2432	2468	2500	2531	2561	2592	2627	2673
Female	2399	2447	2483	2515	2544	2573	2603	2637	2683
Male	2372	2419	2454	2485	2517	2548	2580	2616	2663
American Indian or Alaska Native	2349	2396	2428	2456	2482	2511	2542	2574	2623
Asian	2434	2490	2532	2565	2594	2621	2649	2681	2724
Black/African American	2354	2398	2427	2454	2480	2508	2538	2573	2621
Native Hawaiian or Pacific Islander	2359	2401	2430	2459	2487	2517	2549	2583	2631
Hispanic/Latino Ethnicity	2361	2404	2434	2461	2487	2514	2544	2578	2624
White	2417	2469	2507	2538	2566	2592	2619	2650	2693
Two or More Races	2383	2430	2465	2497	2526	2555	2586	2621	2668
Unidentified Race	2420	2469	2504	2534	2562	2589	2616	2647	2688
LEP Status	2307	2358	2384	2405	2424	2444	2464	2490	2528
IDEA Indicator	2306	2356	2383	2406	2427	2449	2473	2507	2561
Section 504 Status	2397	2441	2475	2506	2536	2565	2595	2629	2674
Economic Disadvantage Status	2361	2405	2436	2463	2490	2518	2548	2583	2629

TABLE 5.89. GRADE 8 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2410	2460	2497	2530	2560	2588	2617	2648	2689
Female	2429	2478	2515	2546	2574	2601	2628	2659	2699
Male	2396	2445	2481	2514	2545	2575	2605	2637	2678
American Indian or Alaska Native	2375	2423	2454	2482	2509	2539	2569	2600	2643
Asian	2468	2523	2561	2592	2619	2644	2669	2696	2732
Black/African American	2376	2423	2454	2482	2510	2538	2568	2601	2644
Native Hawaiian or Pacific Islander	2384	2430	2460	2489	2518	2547	2577	2611	2654
Hispanic/Latino Ethnicity	2386	2434	2466	2494	2521	2548	2576	2607	2649
White	2439	2493	2531	2562	2590	2615	2640	2668	2704
Two or More Races	2410	2459	2494	2526	2555	2583	2611	2644	2685
Unidentified Race	2452	2504	2540	2568	2594	2617	2640	2667	2702
LEP Status	2333	2382	2411	2433	2453	2473	2494	2520	2558
IDEA Indicator	2329	2380	2409	2433	2454	2475	2501	2534	2584
Section 504 Status	2417	2466	2502	2534	2563	2591	2619	2649	2688
Economic Disadvantage Status	2386	2433	2466	2495	2523	2550	2579	2611	2653

TABLE 5.90. GRADE 11 ELA/LITERACY CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2444	2499	2538	2572	2604	2634	2666	2701	2749
Female	2459	2512	2551	2584	2614	2643	2673	2707	2754
Male	2431	2487	2527	2561	2593	2624	2657	2694	2744
American Indian or Alaska Native	2413	2465	2504	2536	2565	2594	2628	2665	2714
Asian	2484	2545	2587	2622	2654	2683	2713	2746	2795
Black/African American	2407	2458	2494	2525	2556	2586	2619	2656	2704
Native Hawaiian or Pacific Islander	2411	2461	2497	2527	2558	2587	2618	2653	2700
Hispanic/Latino Ethnicity	2425	2476	2511	2541	2569	2597	2627	2660	2705
White	2469	2528	2569	2603	2634	2662	2691	2725	2771
Two or More Races	2442	2495	2534	2567	2597	2627	2659	2694	2741
Unidentified Race	2489	2539	2574	2603	2629	2655	2680	2711	2752
LEP Status	2356	2403	2432	2455	2477	2497	2519	2546	2584
IDEA Indicator	2362	2410	2442	2468	2493	2518	2546	2581	2636
Section 504 Status	2463	2519	2558	2590	2621	2650	2681	2715	2764
Economic Disadvantage Status	2423	2474	2510	2541	2570	2599	2630	2664	2711

TABLE 5.91. GRADE 3 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2277	2323	2357	2385	2411	2436	2463	2493	2536
Female	2289	2336	2370	2399	2424	2448	2474	2505	2547
Male	2265	2311	2345	2372	2399	2424	2450	2480	2522
American Indian or Alaska Native	2251	2292	2322	2347	2370	2396	2419	2446	2484
Asian	2327	2376	2411	2439	2464	2488	2514	2544	2587
Black/African American	2247	2289	2321	2348	2372	2399	2425	2453	2493
Native Hawaiian or Pacific Islander	2268	2313	2345	2371	2395	2420	2443	2472	2511
Hispanic/Latino Ethnicity	2251	2294	2324	2350	2374	2397	2422	2451	2490
White	2313	2360	2392	2418	2441	2464	2488	2515	2555
Two or More Races	2279	2325	2358	2385	2410	2435	2461	2490	2533
Unidentified Race	2316	2363	2394	2420	2443	2466	2491	2519	2558
LEP Status	2240	2281	2309	2333	2356	2379	2402	2430	2468
IDEA Indicator	2210	2253	2282	2307	2332	2359	2388	2423	2471
Section 504 Status	2282	2325	2357	2384	2410	2435	2461	2489	2532
Economic Disadvantage Status	2254	2297	2328	2354	2378	2402	2427	2456	2495

TABLE 5.92. GRADE 4 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2325	2371	2404	2432	2458	2483	2511	2543	2589
Female	2343	2388	2421	2448	2473	2499	2526	2558	2606
Male	2310	2356	2388	2416	2442	2468	2495	2526	2571
American Indian or Alaska Native	2294	2334	2363	2388	2411	2436	2462	2491	2532
Asian	2378	2427	2462	2491	2517	2542	2570	2604	2651
Black/African American	2292	2336	2365	2391	2416	2440	2467	2499	2542
Native Hawaiian or Pacific Islander	2308	2354	2385	2413	2438	2461	2486	2515	2556
Hispanic/Latino Ethnicity	2299	2342	2372	2398	2422	2446	2471	2499	2539
White	2357	2403	2435	2461	2485	2509	2535	2565	2610
Two or More Races	2326	2371	2402	2430	2455	2479	2505	2536	2583
Unidentified Race	2367	2413	2444	2470	2495	2518	2544	2574	2619
LEP Status	2276	2317	2345	2368	2389	2410	2433	2460	2495
IDEA Indicator	2248	2291	2320	2344	2368	2392	2420	2455	2504
Section 504 Status	2333	2374	2403	2429	2452	2476	2502	2533	2577
Economic Disadvantage Status	2302	2345	2376	2402	2426	2450	2475	2504	2544

TABLE 5.93. GRADE 5 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2361	2405	2438	2468	2496	2523	2552	2585	2633
Female	2381	2425	2459	2488	2515	2541	2569	2602	2650
Male	2345	2389	2421	2449	2477	2504	2533	2566	2612
American Indian or Alaska Native	2323	2365	2394	2417	2441	2465	2492	2524	2570
Asian	2410	2463	2501	2532	2560	2587	2615	2648	2697
Black/African American	2319	2366	2395	2421	2447	2474	2502	2535	2580
Native Hawaiian or Pacific Islander	2352	2393	2425	2451	2476	2500	2526	2554	2593
Hispanic/Latino Ethnicity	2339	2380	2409	2435	2461	2486	2514	2545	2588
White	2390	2435	2468	2496	2522	2547	2573	2604	2650
Two or More Races	2364	2406	2437	2465	2491	2516	2544	2576	2623
Unidentified Race	2403	2448	2484	2512	2539	2564	2591	2622	2666
LEP Status	2308	2348	2374	2396	2416	2437	2460	2488	2524
IDEA Indicator	2279	2323	2351	2374	2395	2418	2444	2479	2529
Section 504 Status	2368	2408	2439	2467	2493	2518	2544	2576	2622
Economic Disadvantage Status	2341	2382	2412	2438	2463	2489	2516	2546	2588

TABLE 5.94. GRADE 6 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2389	2437	2471	2498	2523	2547	2573	2605	2649
Female	2411	2458	2491	2517	2541	2564	2590	2621	2665
Male	2372	2420	2453	2481	2506	2530	2556	2587	2630
American Indian or Alaska Native	2339	2389	2421	2447	2471	2495	2520	2548	2591
Asian	2446	2500	2534	2560	2585	2609	2634	2664	2706
Black/African American	2337	2393	2427	2454	2480	2505	2530	2560	2602
Native Hawaiian or Pacific Islander	2371	2416	2450	2476	2500	2522	2546	2573	2613
Hispanic/Latino Ethnicity	2369	2415	2445	2471	2494	2515	2538	2566	2605
White	2415	2462	2496	2522	2546	2569	2594	2623	2666
Two or More Races	2386	2433	2466	2495	2519	2542	2567	2598	2641
Unidentified Race	2439	2486	2517	2542	2564	2586	2610	2636	2675
LEP Status	2319	2366	2395	2419	2439	2460	2482	2505	2538
IDEA Indicator	2299	2347	2377	2401	2424	2446	2470	2500	2541
Section 504 Status	2398	2441	2470	2496	2517	2539	2564	2594	2638
Economic Disadvantage Status	2369	2414	2445	2471	2495	2516	2540	2567	2607

TABLE 5.95. GRADE 7 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2403	2456	2493	2524	2551	2579	2608	2639	2685
Female	2429	2481	2516	2545	2573	2599	2625	2658	2703
Male	2382	2435	2472	2503	2531	2559	2587	2620	2666
American Indian or Alaska Native	2357	2406	2439	2467	2495	2522	2549	2583	2629
Asian	2467	2526	2565	2595	2621	2645	2671	2701	2743
Black/African American	2356	2409	2444	2474	2502	2530	2559	2591	2635
Native Hawaiian or Pacific Islander	2371	2423	2461	2492	2518	2545	2572	2602	2641
Hispanic/Latino Ethnicity	2379	2428	2463	2491	2517	2541	2568	2597	2637
White	2433	2487	2523	2551	2578	2604	2629	2662	2705
Two or More Races	2401	2453	2489	2520	2547	2573	2601	2633	2677
Unidentified Race	2458	2512	2546	2573	2596	2619	2642	2670	2710
LEP Status	2324	2370	2401	2426	2449	2472	2496	2523	2560
IDEA Indicator	2309	2355	2387	2413	2437	2461	2487	2518	2564
Section 504 Status	2410	2456	2489	2517	2542	2568	2594	2626	2672
Economic Disadvantage Status	2378	2428	2462	2491	2517	2543	2569	2599	2639

TABLE 5.96. GRADE 8 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2422	2473	2508	2538	2564	2591	2620	2654	2702
Female	2451	2499	2532	2560	2586	2613	2640	2673	2719
Male	2400	2451	2487	2516	2543	2569	2598	2632	2680
American Indian or Alaska Native	2379	2424	2457	2485	2510	2537	2565	2596	2642
Asian	2484	2538	2574	2603	2630	2656	2685	2717	2762
Black/African American	2370	2423	2458	2489	2516	2542	2570	2604	2650
Native Hawaiian or Pacific Islander	2402	2453	2485	2513	2537	2561	2587	2619	2661
Hispanic/Latino Ethnicity	2404	2451	2483	2509	2533	2556	2581	2610	2652
White	2445	2498	2534	2563	2590	2616	2644	2676	2720
Two or More Races	2422	2472	2507	2536	2562	2588	2617	2650	2697
Unidentified Race	2479	2525	2556	2582	2604	2627	2651	2680	2721
LEP Status	2346	2392	2423	2447	2468	2488	2509	2534	2569
IDEA Indicator	2332	2379	2409	2434	2456	2479	2503	2532	2575
Section 504 Status	2426	2473	2506	2533	2559	2584	2612	2647	2694
Economic Disadvantage Status	2402	2450	2483	2509	2534	2558	2584	2614	2656

TABLE 5.97. GRADE 11 ELA/LITERACY CLAIM 2 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2425	2493	2539	2576	2608	2639	2671	2707	2757
Female	2464	2526	2567	2600	2629	2658	2688	2723	2772
Male	2395	2464	2512	2551	2586	2618	2651	2688	2740
American Indian or Alaska Native	2387	2450	2492	2528	2561	2592	2622	2658	2707
Asian	2496	2568	2613	2647	2679	2709	2741	2778	2795
Black/African American	2370	2434	2479	2516	2550	2583	2617	2654	2702
Native Hawaiian or Pacific Islander	2389	2453	2497	2534	2566	2597	2627	2660	2705
Hispanic/Latino Ethnicity	2399	2462	2506	2541	2572	2601	2631	2665	2712
White	2456	2525	2570	2605	2634	2663	2692	2726	2773
Two or More Races	2426	2493	2538	2573	2604	2632	2662	2696	2743
Unidentified Race	2502	2561	2599	2630	2657	2684	2712	2743	2789
LEP Status	2317	2370	2406	2435	2464	2492	2521	2553	2596
IDEA Indicator	2316	2372	2410	2442	2472	2501	2533	2570	2622
Section 504 Status	2443	2507	2549	2583	2612	2641	2670	2704	2753
Economic Disadvantage Status	2396	2461	2505	2541	2573	2603	2633	2668	2716

TABLE 5.98. GRADE 3 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2266	2325	2366	2396	2423	2450	2480	2514	2565
Female	2277	2336	2374	2403	2429	2456	2485	2520	2570
Male	2258	2318	2359	2388	2416	2444	2473	2510	2561
American Indian or Alaska Native	2210	2288	2324	2357	2383	2408	2436	2468	2514
Asian	2323	2377	2411	2440	2466	2493	2523	2557	2608
Black/African American	2198	2282	2319	2355	2381	2408	2435	2468	2516
Native Hawaiian or Pacific Islander	2238	2301	2343	2374	2399	2424	2449	2482	2528
Hispanic/Latino Ethnicity	2219	2296	2335	2366	2391	2416	2442	2473	2519
White	2307	2362	2397	2426	2452	2480	2508	2541	2591
Two or More Races	2265	2326	2364	2394	2421	2448	2477	2513	2563
Unidentified Race	2315	2368	2399	2424	2448	2473	2500	2531	2580
LEP Status	2203	2276	2315	2345	2371	2396	2421	2451	2494
IDEA Indicator	2181	2229	2280	2315	2347	2377	2410	2449	2506
Section 504 Status	2270	2334	2372	2404	2430	2460	2490	2527	2575
Economic Disadvantage Status	2224	2298	2338	2369	2394	2419	2446	2478	2525

TABLE 5.99. GRADE 4 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2297	2358	2399	2431	2462	2491	2523	2558	2609
Female	2305	2365	2405	2437	2466	2495	2526	2561	2611
Male	2287	2351	2393	2426	2457	2487	2519	2555	2607
American Indian or Alaska Native	2253	2310	2353	2384	2414	2443	2471	2505	2553
Asian	2353	2413	2453	2485	2513	2543	2573	2606	2659
Black/African American	2246	2307	2350	2380	2411	2440	2471	2508	2557
Native Hawaiian or Pacific Islander	2264	2330	2370	2402	2432	2459	2489	2526	2573
Hispanic/Latino Ethnicity	2268	2326	2365	2396	2424	2451	2480	2514	2562
White	2336	2396	2434	2465	2494	2521	2550	2586	2636
Two or More Races	2298	2358	2397	2429	2458	2487	2518	2554	2604
Unidentified Race	2336	2393	2430	2462	2490	2516	2544	2575	2622
LEP Status	2234	2296	2333	2363	2390	2415	2442	2473	2517
IDEA Indicator	2205	2270	2314	2347	2378	2409	2442	2482	2540
Section 504 Status	2306	2367	2408	2439	2469	2498	2530	2566	2617
Economic Disadvantage Status	2269	2330	2369	2400	2428	2456	2485	2520	2568

TABLE 5.100. GRADE 5 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2293	2364	2410	2449	2484	2520	2556	2595	2647
Female	2302	2371	2415	2454	2489	2525	2560	2597	2649
Male	2282	2357	2404	2444	2479	2515	2552	2591	2645
American Indian or Alaska Native	2251	2329	2364	2397	2429	2460	2494	2533	2591
Asian	2352	2424	2472	2512	2546	2578	2609	2643	2693
Black/African American	2232	2313	2357	2393	2424	2458	2495	2537	2596
Native Hawaiian or Pacific Islander	2261	2339	2386	2417	2452	2486	2522	2562	2617
Hispanic/Latino Ethnicity	2263	2333	2373	2409	2442	2475	2511	2552	2606
White	2337	2405	2449	2486	2519	2551	2582	2618	2667
Two or More Races	2298	2368	2411	2448	2482	2517	2552	2591	2645
Unidentified Race	2339	2405	2450	2486	2518	2550	2582	2617	2664
LEP Status	2232	2279	2333	2360	2393	2419	2449	2485	2538
IDEA Indicator	2201	2259	2319	2347	2386	2416	2451	2495	2561
Section 504 Status	2320	2377	2422	2458	2493	2526	2562	2601	2652
Economic Disadvantage Status	2264	2335	2378	2412	2446	2479	2515	2556	2610

TABLE 5.101. GRADE 6 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2351	2422	2462	2501	2533	2565	2599	2639	2700
Female	2370	2436	2477	2511	2543	2573	2608	2648	2707
Male	2335	2408	2452	2491	2521	2556	2588	2628	2693
American Indian or Alaska Native	2308	2368	2412	2448	2481	2511	2546	2582	2640
Asian	2416	2479	2520	2555	2584	2618	2650	2694	2724
Black/African American	2306	2370	2415	2450	2486	2515	2554	2591	2647
Native Hawaiian or Pacific Islander	2325	2391	2439	2474	2506	2539	2571	2612	2669
Hispanic/Latino Ethnicity	2325	2388	2435	2466	2500	2530	2562	2602	2657
White	2387	2452	2495	2527	2558	2588	2621	2663	2723
Two or More Races	2350	2418	2459	2498	2528	2560	2594	2632	2695
Unidentified Race	2404	2461	2501	2533	2563	2592	2624	2664	2720
LEP Status	2270	2325	2369	2402	2432	2457	2494	2527	2578
IDEA Indicator	2210	2308	2346	2384	2419	2449	2488	2528	2589
Section 504 Status	2370	2435	2471	2506	2537	2566	2598	2638	2699
Economic Disadvantage Status	2326	2389	2437	2469	2502	2532	2565	2604	2659

TABLE 5.102. GRADE 7 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2369	2431	2472	2508	2540	2574	2608	2650	2709
Female	2380	2440	2481	2516	2547	2581	2615	2656	2716
Male	2359	2422	2463	2499	2535	2568	2601	2644	2704
American Indian or Alaska Native	2324	2386	2427	2461	2493	2522	2556	2597	2655
Asian	2425	2485	2529	2562	2593	2625	2655	2698	2745
Black/African American	2319	2380	2422	2457	2486	2518	2553	2595	2653
Native Hawaiian or Pacific Islander	2346	2400	2437	2467	2499	2529	2564	2603	2660
Hispanic/Latino Ethnicity	2344	2402	2438	2471	2500	2532	2566	2604	2660
White	2407	2465	2508	2541	2574	2604	2638	2676	2733
Two or More Races	2369	2433	2474	2508	2539	2572	2606	2647	2705
Unidentified Race	2414	2464	2504	2535	2567	2594	2628	2664	2719
LEP Status	2294	2345	2378	2411	2434	2460	2485	2518	2568
IDEA Indicator	2270	2338	2375	2410	2437	2463	2495	2535	2593
Section 504 Status	2382	2443	2484	2518	2548	2579	2612	2655	2714
Economic Disadvantage Status	2346	2405	2441	2475	2505	2535	2571	2609	2664

TABLE 5.103. GRADE 8 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2382	2446	2490	2524	2556	2587	2621	2661	2722
Female	2400	2463	2503	2536	2566	2596	2629	2668	2727
Male	2365	2431	2476	2512	2545	2578	2612	2654	2715
American Indian or Alaska Native	2331	2397	2441	2476	2509	2540	2573	2611	2668
Asian	2447	2508	2548	2581	2611	2642	2674	2714	2769
Black/African American	2314	2392	2438	2475	2508	2539	2573	2612	2667
Native Hawaiian or Pacific Islander	2352	2417	2459	2493	2523	2551	2582	2618	2681
Hispanic/Latino Ethnicity	2357	2421	2462	2495	2525	2554	2584	2621	2674
White	2404	2472	2514	2550	2581	2611	2646	2685	2746
Two or More Races	2389	2451	2492	2525	2556	2586	2619	2658	2717
Unidentified Race	2437	2492	2530	2559	2587	2616	2646	2681	2736
LEP Status	2290	2351	2392	2423	2451	2478	2505	2537	2583
IDEA Indicator	2288	2342	2388	2419	2449	2479	2510	2549	2607
Section 504 Status	2395	2459	2500	2534	2562	2594	2625	2664	2724
Economic Disadvantage Status	2358	2422	2464	2497	2527	2556	2586	2624	2678

TABLE 5.104. GRADE 11 ELA/LITERACY CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2389	2456	2506	2548	2586	2623	2663	2709	2771
Female	2407	2475	2522	2562	2597	2632	2669	2713	2774
Male	2374	2440	2490	2533	2573	2613	2655	2704	2768
American Indian or Alaska Native	2359	2418	2467	2505	2544	2579	2619	2664	2728
Asian	2439	2512	2562	2604	2641	2677	2715	2757	2795
Black/African American	2352	2413	2460	2500	2538	2576	2615	2662	2725
Native Hawaiian or Pacific Islander	2355	2414	2459	2497	2535	2573	2610	2656	2720
Hispanic/Latino Ethnicity	2372	2432	2478	2516	2552	2587	2625	2669	2729
White	2408	2481	2533	2576	2613	2650	2687	2730	2794
Two or More Races	2391	2457	2504	2545	2582	2618	2657	2705	2768
Unidentified Race	2431	2500	2545	2584	2618	2652	2686	2725	2783
LEP Status	2313	2367	2400	2428	2455	2482	2512	2548	2601
IDEA Indicator	2312	2368	2404	2435	2465	2497	2533	2577	2645
Section 504 Status	2402	2474	2524	2562	2599	2636	2673	2719	2783
Economic Disadvantage Status	2369	2430	2476	2515	2551	2587	2626	2670	2732

TABLE 5.105. GRADE 3 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2249	2305	2340	2374	2407	2438	2471	2506	2551
Female	2258	2314	2352	2386	2418	2449	2480	2514	2558
Male	2242	2295	2331	2364	2395	2428	2461	2498	2543
American Indian or Alaska Native	2221	2270	2305	2333	2359	2386	2417	2452	2498
Asian	2297	2356	2400	2436	2467	2496	2524	2555	2594
Black/African American	2221	2266	2302	2330	2357	2387	2419	2458	2507
Native Hawaiian or Pacific Islander	2237	2285	2321	2350	2378	2409	2441	2479	2526
Hispanic/Latino Ethnicity	2232	2279	2314	2341	2369	2397	2429	2465	2512
White	2277	2334	2375	2409	2439	2467	2496	2527	2569
Two or More Races	2251	2307	2342	2375	2406	2438	2469	2504	2549
Unidentified Race	2286	2343	2382	2416	2446	2474	2501	2530	2570
LEP Status	2221	2268	2301	2327	2352	2377	2406	2441	2489
IDEA Indicator	2191	2250	2283	2312	2336	2363	2396	2437	2493
Section 504 Status	2250	2306	2344	2377	2409	2441	2471	2507	2551
Economic Disadvantage Status	2232	2281	2316	2344	2372	2401	2432	2468	2514

TABLE 5.106. GRADE 4 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2273	2336	2377	2413	2447	2480	2513	2547	2592
Female	2284	2346	2390	2426	2460	2492	2523	2557	2601
Male	2265	2328	2364	2401	2435	2469	2502	2537	2583
American Indian or Alaska Native	2252	2296	2334	2361	2392	2422	2453	2489	2535
Asian	2337	2403	2449	2486	2516	2543	2570	2600	2642
Black/African American	2242	2290	2328	2357	2391	2423	2457	2496	2545
Native Hawaiian or Pacific Islander	2265	2325	2358	2391	2426	2458	2490	2526	2571
Hispanic/Latino Ethnicity	2259	2312	2345	2378	2408	2439	2472	2508	2555
White	2309	2362	2410	2445	2477	2505	2533	2564	2606
Two or More Races	2278	2341	2381	2416	2449	2480	2512	2547	2591
Unidentified Race	2326	2384	2428	2463	2493	2521	2547	2577	2617
LEP Status	2250	2280	2325	2346	2373	2399	2428	2462	2510
IDEA Indicator	2240	2266	2306	2336	2359	2389	2421	2463	2521
Section 504 Status	2275	2339	2379	2414	2446	2477	2509	2543	2585
Economic Disadvantage Status	2260	2315	2348	2382	2411	2443	2475	2512	2558

TABLE 5.107. GRADE 5 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2358	2415	2454	2485	2515	2544	2572	2604	2649
Female	2376	2431	2468	2499	2527	2554	2583	2615	2661
Male	2345	2404	2440	2472	2502	2531	2560	2593	2637
American Indian or Alaska Native	2312	2364	2402	2430	2456	2482	2512	2547	2593
Asian	2421	2480	2520	2550	2576	2601	2627	2658	2701
Black/African American	2314	2370	2406	2437	2465	2492	2522	2556	2602
Native Hawaiian or Pacific Islander	2342	2398	2436	2468	2497	2523	2552	2582	2626
Hispanic/Latino Ethnicity	2337	2389	2423	2453	2480	2508	2537	2569	2613
White	2393	2445	2482	2513	2540	2564	2590	2621	2666
Two or More Races	2362	2416	2454	2485	2513	2541	2569	2601	2645
Unidentified Race	2412	2464	2501	2531	2555	2581	2606	2635	2678
LEP Status	2293	2347	2383	2409	2432	2456	2481	2510	2552
IDEA Indicator	2266	2334	2361	2393	2417	2444	2473	2510	2563
Section 504 Status	2373	2424	2460	2489	2516	2542	2568	2598	2642
Economic Disadvantage Status	2339	2392	2426	2456	2483	2511	2539	2571	2614

TABLE 5.108. GRADE 6 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2370	2431	2471	2504	2533	2561	2589	2619	2661
Female	2390	2449	2489	2520	2548	2575	2601	2631	2670
Male	2355	2416	2456	2489	2518	2546	2576	2607	2649
American Indian or Alaska Native	2325	2383	2421	2451	2478	2506	2534	2570	2611
Asian	2437	2499	2539	2570	2595	2618	2642	2668	2705
Black/African American	2311	2381	2424	2457	2486	2515	2544	2577	2620
Native Hawaiian or Pacific Islander	2351	2409	2447	2478	2508	2534	2563	2593	2632
Hispanic/Latino Ethnicity	2347	2404	2442	2473	2501	2528	2556	2587	2627
White	2398	2459	2497	2528	2555	2580	2606	2634	2676
Two or More Races	2372	2433	2471	2503	2532	2559	2586	2616	2655
Unidentified Race	2420	2478	2515	2545	2571	2594	2617	2643	2679
LEP Status	2289	2355	2390	2418	2443	2467	2493	2523	2563
IDEA Indicator	2283	2348	2381	2412	2436	2463	2490	2522	2569
Section 504 Status	2377	2437	2474	2503	2529	2556	2583	2612	2651
Economic Disadvantage Status	2349	2406	2443	2474	2502	2529	2557	2587	2627

TABLE 5.109. GRADE 7 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2366	2430	2474	2512	2545	2578	2607	2640	2683
Female	2389	2454	2496	2534	2564	2594	2621	2651	2692
Male	2349	2410	2454	2490	2527	2559	2593	2627	2671
American Indian or Alaska Native	2314	2386	2426	2458	2487	2519	2553	2589	2633
Asian	2441	2511	2556	2590	2616	2640	2665	2693	2733
Black/African American	2312	2380	2422	2457	2485	2522	2554	2593	2637
Native Hawaiian or Pacific Islander	2335	2397	2440	2475	2507	2540	2573	2607	2649
Hispanic/Latino Ethnicity	2335	2398	2439	2473	2504	2536	2569	2603	2645
White	2396	2463	2506	2541	2573	2599	2627	2656	2696
Two or More Races	2372	2436	2482	2518	2551	2580	2609	2640	2682
Unidentified Race	2419	2483	2527	2560	2587	2612	2635	2662	2699
LEP Status	2291	2346	2384	2413	2438	2463	2489	2523	2568
IDEA Indicator	2288	2341	2381	2412	2436	2462	2491	2527	2580
Section 504 Status	2372	2434	2475	2511	2545	2574	2602	2635	2679
Economic Disadvantage Status	2336	2401	2441	2475	2507	2539	2572	2605	2648

TABLE 5.110. GRADE 8 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2385	2447	2492	2530	2563	2596	2626	2660	2705
Female	2411	2473	2517	2553	2585	2613	2642	2673	2717
Male	2361	2427	2470	2506	2541	2575	2608	2644	2691
American Indian or Alaska Native	2325	2400	2442	2475	2506	2538	2569	2605	2657
Asian	2457	2528	2573	2606	2633	2658	2683	2711	2752
Black/African American	2325	2401	2442	2476	2506	2540	2575	2612	2660
Native Hawaiian or Pacific Islander	2358	2420	2460	2494	2528	2560	2593	2628	2669
Hispanic/Latino Ethnicity	2358	2421	2461	2494	2526	2557	2589	2623	2667
White	2411	2475	2521	2558	2589	2617	2646	2677	2721
Two or More Races	2391	2453	2497	2534	2567	2597	2627	2660	2703
Unidentified Race	2435	2498	2542	2575	2603	2628	2653	2681	2719
LEP Status	2312	2365	2401	2430	2457	2481	2507	2540	2587
IDEA Indicator	2309	2361	2397	2427	2453	2478	2505	2540	2593
Section 504 Status	2385	2446	2490	2526	2559	2590	2620	2653	2698
Economic Disadvantage Status	2358	2421	2461	2494	2526	2558	2590	2624	2669

TABLE 5.111. GRADE 11 ELA/LITERACY CLAIM 4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2401	2482	2536	2580	2616	2649	2681	2715	2762
Female	2434	2513	2564	2602	2635	2664	2693	2726	2772
Male	2373	2456	2510	2556	2596	2632	2666	2703	2750
American Indian or Alaska Native	2348	2430	2480	2522	2562	2598	2635	2673	2723
Asian	2486	2567	2616	2651	2680	2706	2733	2764	2795
Black/African American	2338	2421	2473	2516	2555	2593	2630	2669	2717
Native Hawaiian or Pacific Islander	2355	2433	2486	2527	2563	2599	2634	2672	2721
Hispanic/Latino Ethnicity	2376	2454	2506	2549	2585	2619	2651	2686	2732
White	2425	2506	2561	2602	2636	2667	2696	2728	2774
Two or More Races	2404	2483	2536	2578	2614	2646	2677	2711	2758
Unidentified Race	2482	2555	2602	2633	2659	2683	2708	2738	2780
LEP Status	2299	2357	2407	2442	2475	2504	2537	2575	2626
IDEA Indicator	2299	2357	2410	2446	2480	2511	2546	2589	2648
Section 504 Status	2414	2495	2546	2587	2620	2651	2683	2717	2762
Economic Disadvantage Status	2370	2449	2501	2545	2582	2617	2651	2687	2734

TABLE 5.112. GRADE 3 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2320	2359	2384	2405	2425	2445	2467	2493	2530
Female	2324	2360	2384	2404	2423	2443	2464	2489	2524
Male	2315	2357	2384	2406	2427	2448	2471	2497	2535
American Indian or Alaska Native	2287	2326	2351	2373	2391	2409	2428	2451	2486
Asian	2375	2412	2438	2461	2482	2503	2526	2552	2592
Black/African American	2276	2319	2347	2369	2388	2406	2426	2449	2481
Native Hawaiian or Pacific Islander	2306	2347	2372	2392	2410	2428	2449	2474	2507
Hispanic/Latino Ethnicity	2301	2339	2362	2381	2399	2416	2435	2458	2490
White	2347	2384	2409	2429	2448	2467	2487	2510	2544
Two or More Races	2317	2356	2382	2403	2423	2443	2465	2491	2529
Unidentified Race	2360	2392	2414	2434	2452	2470	2489	2511	2542
LEP Status	2293	2330	2353	2372	2389	2406	2424	2448	2481
IDEA Indicator	2227	2279	2313	2340	2363	2385	2410	2439	2481
Section 504 Status	2317	2355	2381	2404	2425	2446	2470	2495	2532
Economic Disadvantage Status	2301	2339	2364	2383	2401	2419	2439	2462	2496

TABLE 5.113. GRADE 4 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2359	2395	2420	2442	2463	2484	2507	2535	2574
Female	2363	2396	2420	2441	2460	2481	2503	2530	2568
Male	2354	2393	2421	2444	2465	2488	2512	2540	2581
American Indian or Alaska Native	2329	2363	2388	2408	2428	2446	2466	2491	2528
Asian	2412	2451	2479	2504	2527	2550	2574	2600	2637
Black/African American	2317	2357	2383	2403	2422	2441	2462	2487	2521
Native Hawaiian or Pacific Islander	2350	2385	2409	2430	2450	2468	2489	2514	2547
Hispanic/Latino Ethnicity	2339	2372	2394	2413	2431	2449	2469	2492	2526
White	2388	2424	2448	2469	2488	2508	2529	2554	2589
Two or More Races	2360	2396	2421	2443	2463	2484	2507	2534	2573
Unidentified Race	2395	2428	2451	2471	2490	2509	2531	2556	2587
LEP Status	2323	2356	2377	2394	2410	2427	2446	2468	2502
IDEA Indicator	2281	2323	2349	2372	2392	2414	2439	2469	2515
Section 504 Status	2368	2402	2427	2447	2467	2488	2511	2539	2579
Economic Disadvantage Status	2341	2375	2398	2418	2436	2455	2475	2499	2535

TABLE 5.114. GRADE 5 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2372	2413	2441	2467	2491	2516	2543	2573	2612
Female	2379	2416	2443	2467	2490	2513	2539	2568	2607
Male	2366	2409	2440	2467	2493	2520	2547	2577	2617
American Indian or Alaska Native	2336	2378	2405	2427	2447	2468	2491	2519	2558
Asian	2433	2480	2512	2539	2564	2587	2611	2640	2681
Black/African American	2326	2367	2396	2419	2440	2462	2486	2515	2556
Native Hawaiian or Pacific Islander	2364	2404	2432	2455	2477	2499	2524	2551	2586
Hispanic/Latino Ethnicity	2352	2389	2414	2435	2455	2476	2499	2526	2564
White	2402	2442	2472	2497	2520	2543	2565	2591	2626
Two or More Races	2375	2414	2443	2468	2491	2516	2542	2572	2610
Unidentified Race	2410	2448	2478	2501	2523	2546	2569	2594	2627
LEP Status	2329	2366	2389	2409	2427	2444	2464	2489	2526
IDEA Indicator	2297	2337	2365	2388	2409	2431	2455	2488	2537
Section 504 Status	2382	2421	2448	2474	2496	2521	2547	2576	2615
Economic Disadvantage Status	2354	2392	2418	2440	2461	2482	2506	2534	2572

TABLE 5.115. GRADE 6 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2370	2426	2463	2492	2518	2543	2570	2602	2647
Female	2382	2435	2470	2497	2522	2547	2573	2603	2646
Male	2359	2417	2456	2487	2513	2539	2568	2601	2648
American Indian or Alaska Native	2324	2372	2408	2439	2467	2492	2518	2546	2587
Asian	2456	2507	2542	2571	2597	2624	2652	2684	2731
Black/African American	2319	2372	2406	2436	2462	2488	2514	2544	2586
Native Hawaiian or Pacific Islander	2349	2405	2443	2472	2497	2520	2544	2575	2614
Hispanic/Latino Ethnicity	2343	2395	2430	2459	2483	2506	2530	2559	2599
White	2409	2462	2494	2520	2543	2566	2591	2620	2660
Two or More Races	2363	2417	2455	2484	2509	2534	2561	2593	2637
Unidentified Race	2426	2476	2507	2533	2555	2578	2603	2632	2671
LEP Status	2295	2343	2375	2402	2426	2449	2473	2500	2539
IDEA Indicator	2257	2309	2343	2372	2399	2427	2458	2493	2544
Section 504 Status	2390	2438	2469	2494	2518	2542	2568	2599	2644
Economic Disadvantage Status	2342	2395	2431	2460	2485	2508	2533	2562	2602

TABLE 5.116. GRADE 7 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2380	2435	2473	2504	2532	2560	2590	2625	2671
Female	2391	2443	2479	2508	2534	2560	2590	2623	2668
Male	2369	2427	2467	2500	2530	2560	2591	2626	2674
American Indian or Alaska Native	2341	2391	2426	2456	2482	2509	2535	2566	2610
Asian	2465	2524	2564	2597	2626	2652	2680	2713	2759
Black/African American	2325	2379	2415	2445	2472	2498	2527	2559	2603
Native Hawaiian or Pacific Islander	2350	2407	2444	2474	2502	2528	2557	2589	2629
Hispanic/Latino Ethnicity	2350	2403	2438	2466	2492	2517	2543	2574	2618
White	2420	2473	2508	2536	2562	2587	2613	2644	2684
Two or More Races	2379	2432	2469	2499	2527	2554	2583	2617	2662
Unidentified Race	2435	2487	2519	2547	2574	2599	2624	2652	2692
LEP Status	2294	2345	2378	2405	2429	2452	2477	2507	2550
IDEA Indicator	2272	2323	2357	2385	2411	2438	2467	2503	2557
Section 504 Status	2396	2446	2482	2510	2536	2562	2591	2625	2670
Economic Disadvantage Status	2351	2404	2440	2469	2496	2521	2548	2579	2623

TABLE 5.117. GRADE 8 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2383	2437	2477	2510	2542	2573	2607	2647	2701
Female	2396	2448	2486	2518	2547	2578	2610	2648	2699
Male	2371	2427	2467	2503	2536	2569	2604	2646	2703
American Indian or Alaska Native	2344	2393	2428	2457	2485	2512	2541	2576	2628
Asian	2469	2533	2579	2615	2649	2680	2712	2750	2802
Black/African American	2330	2385	2419	2449	2476	2504	2534	2570	2623
Native Hawaiian or Pacific Islander	2357	2409	2449	2483	2513	2543	2575	2612	2660
Hispanic/Latino Ethnicity	2357	2407	2443	2473	2501	2528	2558	2593	2641
White	2417	2472	2510	2542	2572	2601	2632	2667	2715
Two or More Races	2382	2435	2473	2506	2536	2568	2601	2640	2695
Unidentified Race	2438	2493	2532	2564	2592	2620	2649	2681	2726
LEP Status	2305	2354	2385	2411	2436	2460	2487	2521	2572
IDEA Indicator	2286	2336	2368	2394	2418	2443	2472	2509	2565
Section 504 Status	2394	2445	2482	2513	2541	2572	2605	2645	2700
Economic Disadvantage Status	2357	2408	2443	2474	2503	2532	2562	2599	2650

TABLE 5.118. GRADE 11 MATHEMATICS CLAIM 1 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2388	2449	2492	2528	2562	2594	2628	2670	2732
Female	2407	2465	2505	2539	2570	2600	2631	2670	2727
Male	2371	2434	2478	2516	2552	2587	2623	2669	2739
American Indian or Alaska Native	2361	2418	2452	2482	2512	2541	2572	2608	2660
Asian	2482	2555	2601	2639	2677	2713	2750	2790	2840
Black/African American	2339	2398	2438	2472	2503	2534	2569	2605	2655
Native Hawaiian or Pacific Islander	2360	2423	2461	2495	2527	2556	2585	2619	2669
Hispanic/Latino Ethnicity	2369	2428	2468	2501	2531	2561	2591	2625	2673
White	2409	2472	2516	2552	2584	2616	2650	2691	2748
Two or More Races	2377	2434	2474	2509	2540	2571	2604	2646	2712
Unidentified Race	2454	2516	2557	2590	2618	2645	2674	2708	2756
LEP Status	2305	2361	2397	2426	2452	2477	2505	2539	2592
IDEA Indicator	2294	2348	2384	2412	2438	2463	2491	2525	2576
Section 504 Status	2398	2455	2494	2529	2561	2590	2623	2664	2728
Economic Disadvantage Status	2366	2424	2464	2497	2528	2559	2590	2625	2678

TABLE 5.119. GRADE 3 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2296	2343	2374	2399	2422	2444	2467	2493	2528
Female	2299	2343	2373	2397	2420	2442	2465	2490	2525
Male	2293	2343	2375	2402	2425	2447	2469	2495	2531
American Indian or Alaska Native	2269	2315	2345	2368	2389	2410	2431	2455	2488
Asian	2351	2397	2429	2453	2475	2495	2516	2540	2577
Black/African American	2255	2305	2338	2360	2380	2401	2423	2448	2481
Native Hawaiian or Pacific Islander	2276	2321	2351	2375	2397	2418	2441	2465	2501
Hispanic/Latino Ethnicity	2269	2315	2344	2367	2388	2408	2430	2455	2488
White	2337	2381	2409	2431	2451	2469	2489	2512	2543
Two or More Races	2298	2344	2373	2398	2421	2442	2464	2490	2525
Unidentified Race	2335	2377	2406	2427	2448	2466	2486	2507	2538
LEP Status	2257	2304	2332	2354	2374	2394	2416	2441	2475
IDEA Indicator	2189	2273	2306	2332	2356	2381	2409	2441	2484
Section 504 Status	2296	2345	2378	2402	2425	2446	2469	2495	2532
Economic Disadvantage Status	2271	2318	2347	2370	2392	2413	2435	2459	2493

TABLE 5.120. GRADE 4 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2321	2376	2410	2436	2460	2483	2507	2534	2573
Female	2327	2379	2411	2436	2460	2482	2505	2531	2568
Male	2315	2374	2408	2435	2460	2484	2509	2537	2576
American Indian or Alaska Native	2204	2342	2376	2402	2422	2443	2468	2494	2529
Asian	2387	2437	2470	2496	2518	2540	2562	2589	2629
Black/African American	2204	2317	2360	2392	2412	2433	2458	2484	2519
Native Hawaiian or Pacific Islander	2308	2361	2394	2419	2441	2463	2484	2509	2542
Hispanic/Latino Ethnicity	2280	2346	2378	2403	2425	2446	2468	2493	2527
White	2364	2413	2443	2468	2489	2509	2530	2555	2591
Two or More Races	2332	2380	2412	2436	2459	2481	2505	2532	2570
Unidentified Race	2365	2413	2442	2465	2486	2504	2525	2548	2582
LEP Status	2204	2321	2353	2377	2399	2419	2440	2464	2499
IDEA Indicator	2204	2297	2338	2362	2387	2411	2437	2469	2514
Section 504 Status	2338	2386	2418	2442	2463	2485	2510	2536	2576
Economic Disadvantage Status	2294	2352	2383	2408	2430	2452	2474	2499	2534

TABLE 5.121. GRADE 5 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2219	2392	2429	2460	2487	2514	2542	2572	2612
Female	2219	2396	2432	2461	2487	2513	2541	2570	2609
Male	2219	2387	2426	2458	2487	2514	2543	2574	2615
American Indian or Alaska Native	2219	2339	2388	2417	2441	2466	2491	2519	2558
Asian	2402	2460	2499	2531	2557	2580	2604	2629	2664
Black/African American	2219	2219	2377	2406	2430	2454	2481	2511	2553
Native Hawaiian or Pacific Islander	2219	2370	2408	2437	2464	2488	2516	2546	2583
Hispanic/Latino Ethnicity	2219	2353	2390	2418	2443	2467	2493	2522	2562
White	2381	2438	2471	2498	2521	2543	2567	2593	2628
Two or More Races	2219	2395	2432	2461	2488	2513	2540	2569	2608
Unidentified Race	2376	2431	2466	2493	2518	2541	2565	2590	2623
LEP Status	2219	2219	2354	2380	2404	2426	2449	2475	2513
IDEA Indicator	2219	2219	2328	2371	2397	2422	2451	2485	2536
Section 504 Status	2348	2408	2444	2472	2497	2522	2547	2575	2613
Economic Disadvantage Status	2219	2359	2396	2425	2451	2475	2501	2531	2570

TABLE 5.122. GRADE 6 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2235	2394	2445	2481	2511	2538	2565	2597	2639
Female	2235	2401	2450	2484	2512	2537	2564	2595	2637
Male	2235	2387	2440	2478	2510	2538	2566	2598	2642
American Indian or Alaska Native	2235	2235	2380	2421	2456	2486	2515	2543	2582
Asian	2415	2486	2526	2556	2583	2609	2635	2665	2705
Black/African American	2235	2235	2356	2405	2445	2476	2506	2538	2578
Native Hawaiian or Pacific Islander	2235	2370	2419	2453	2483	2510	2536	2566	2604
Hispanic/Latino Ethnicity	2235	2357	2405	2441	2471	2497	2522	2550	2588
White	2336	2441	2485	2515	2541	2565	2590	2618	2656
Two or More Races	2235	2394	2442	2476	2505	2531	2559	2590	2632
Unidentified Race	2390	2454	2492	2520	2543	2565	2589	2617	2653
LEP Status	2235	2235	2313	2377	2408	2436	2464	2493	2532
IDEA Indicator	2235	2235	2235	2349	2388	2421	2455	2493	2543
Section 504 Status	2302	2413	2457	2487	2516	2541	2567	2596	2636
Economic Disadvantage Status	2235	2364	2410	2445	2474	2500	2526	2555	2594

TABLE 5.123. GRADE 7 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2250	2395	2451	2489	2523	2554	2588	2623	2668
Female	2250	2401	2454	2491	2524	2555	2587	2621	2666
Male	2250	2389	2448	2487	2522	2554	2588	2624	2671
American Indian or Alaska Native	2250	2250	2388	2438	2471	2503	2532	2565	2610
Asian	2425	2501	2547	2583	2613	2641	2668	2697	2739
Black/African American	2250	2250	2359	2424	2458	2488	2520	2553	2600
Native Hawaiian or Pacific Islander	2250	2327	2417	2458	2491	2521	2551	2585	2625
Hispanic/Latino Ethnicity	2250	2250	2403	2443	2474	2503	2533	2567	2611
White	2250	2450	2496	2528	2559	2588	2614	2645	2686
Two or More Races	2250	2398	2453	2490	2522	2552	2583	2617	2662
Unidentified Race	2359	2454	2498	2531	2560	2587	2614	2644	2684
LEP Status	2250	2250	2250	2359	2408	2440	2469	2502	2547
IDEA Indicator	2250	2250	2250	2345	2400	2434	2466	2503	2557
Section 504 Status	2250	2419	2465	2501	2531	2561	2591	2626	2671
Economic Disadvantage Status	2250	2250	2410	2449	2480	2510	2540	2574	2619

TABLE 5.124. GRADE 8 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2265	2414	2463	2500	2538	2573	2609	2648	2698
Female	2265	2422	2467	2503	2540	2574	2609	2648	2696
Male	2265	2403	2459	2496	2535	2573	2608	2648	2701
American Indian or Alaska Native	2265	2265	2417	2454	2484	2514	2547	2584	2637
Asian	2429	2505	2562	2603	2638	2667	2697	2731	2779
Black/African American	2265	2265	2382	2444	2474	2501	2535	2573	2623
Native Hawaiian or Pacific Islander	2265	2318	2426	2463	2498	2534	2569	2605	2652
Hispanic/Latino Ethnicity	2265	2265	2425	2457	2487	2518	2550	2587	2636
White	2284	2460	2503	2544	2579	2608	2641	2673	2718
Two or More Races	2265	2411	2459	2497	2534	2569	2604	2643	2692
Unidentified Race	2350	2461	2507	2544	2577	2608	2637	2669	2710
LEP Status	2265	2265	2265	2361	2425	2455	2483	2516	2563
IDEA Indicator	2265	2265	2265	2357	2420	2450	2479	2515	2568
Section 504 Status	2265	2430	2469	2507	2542	2578	2614	2650	2700
Economic Disadvantage Status	2265	2323	2429	2461	2492	2525	2558	2596	2646

TABLE 5.125. GRADE 11 MATHEMATICS CLAIM 2/4 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2280	2358	2463	2505	2546	2587	2628	2672	2727
Female	2280	2377	2468	2508	2547	2585	2624	2666	2719
Male	2280	2338	2457	2503	2545	2588	2631	2677	2735
American Indian or Alaska Native	2280	2280	2390	2464	2504	2540	2579	2625	2680
Asian	2398	2516	2578	2624	2661	2694	2727	2765	2816
Black/African American	2280	2280	2342	2429	2477	2509	2547	2591	2650
Native Hawaiian or Pacific Islander	2280	2280	2374	2462	2505	2542	2583	2621	2678
Hispanic/Latino Ethnicity	2280	2280	2418	2471	2503	2537	2574	2617	2672
White	2280	2431	2501	2547	2587	2623	2659	2697	2747
Two or More Races	2280	2322	2443	2497	2539	2578	2618	2661	2718
Unidentified Race	2280	2468	2518	2561	2599	2633	2664	2699	2742
LEP Status	2280	2280	2280	2321	2415	2461	2491	2524	2581
IDEA Indicator	2280	2280	2280	2319	2402	2452	2487	2522	2581
Section 504 Status	2280	2380	2473	2516	2555	2594	2631	2674	2731
Economic Disadvantage Status	2280	2280	2413	2471	2506	2541	2580	2624	2680

TABLE 5.126. GRADE 3 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2290	2341	2371	2398	2421	2445	2469	2496	2533
Female	2297	2345	2375	2401	2424	2446	2470	2495	2532
Male	2284	2337	2367	2395	2418	2443	2468	2496	2535
American Indian or Alaska Native	2253	2307	2338	2360	2381	2404	2427	2452	2488
Asian	2347	2396	2429	2456	2479	2501	2524	2551	2591
Black/African American	2229	2301	2333	2354	2377	2400	2424	2451	2487
Native Hawaiian or Pacific Islander	2275	2324	2356	2380	2403	2425	2448	2474	2508
Hispanic/Latino Ethnicity	2266	2317	2346	2370	2391	2412	2435	2460	2493
White	2323	2368	2400	2425	2447	2468	2490	2514	2550
Two or More Races	2290	2341	2371	2397	2421	2444	2468	2495	2532
Unidentified Race	2327	2375	2406	2430	2452	2472	2493	2515	2548
LEP Status	2255	2306	2336	2359	2380	2400	2422	2447	2482
IDEA Indicator	2223	2281	2315	2340	2362	2385	2411	2442	2487
Section 504 Status	2294	2343	2374	2401	2423	2448	2473	2500	2537
Economic Disadvantage Status	2269	2319	2349	2373	2395	2417	2439	2465	2499

TABLE 5.127. GRADE 4 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2337	2382	2411	2436	2460	2484	2508	2535	2573
Female	2340	2385	2412	2437	2460	2483	2507	2533	2570
Male	2333	2379	2409	2436	2460	2485	2509	2537	2575
American Indian or Alaska Native	2299	2348	2377	2398	2420	2442	2466	2492	2529
Asian	2394	2441	2473	2499	2522	2544	2567	2593	2630
Black/African American	2282	2344	2374	2392	2415	2437	2459	2486	2522
Native Hawaiian or Pacific Islander	2330	2374	2401	2423	2444	2465	2486	2513	2546
Hispanic/Latino Ethnicity	2315	2358	2385	2407	2427	2447	2469	2494	2528
White	2366	2410	2442	2466	2488	2509	2530	2554	2587
Two or More Races	2338	2382	2412	2437	2460	2483	2506	2533	2570
Unidentified Race	2372	2414	2443	2467	2488	2508	2529	2553	2584
LEP Status	2297	2342	2368	2388	2407	2425	2444	2468	2501
IDEA Indicator	2280	2325	2353	2375	2395	2415	2439	2469	2514
Section 504 Status	2339	2385	2415	2440	2462	2485	2510	2536	2571
Economic Disadvantage Status	2319	2361	2388	2411	2431	2452	2474	2500	2535

TABLE 5.128. GRADE 5 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2338	2392	2426	2455	2482	2510	2540	2574	2617
Female	2343	2396	2429	2457	2483	2510	2539	2573	2616
Male	2334	2387	2423	2453	2481	2510	2541	2575	2618
American Indian or Alaska Native	2307	2354	2388	2415	2438	2461	2487	2518	2562
Asian	2404	2458	2497	2528	2555	2581	2606	2634	2675
Black/African American	2303	2350	2382	2411	2435	2457	2482	2512	2557
Native Hawaiian or Pacific Islander	2326	2376	2411	2438	2462	2487	2515	2545	2584
Hispanic/Latino Ethnicity	2318	2361	2395	2420	2443	2466	2491	2522	2564
White	2371	2425	2460	2489	2515	2541	2567	2597	2635
Two or More Races	2337	2393	2427	2456	2482	2509	2538	2570	2614
Unidentified Race	2376	2427	2461	2490	2515	2540	2565	2593	2629
LEP Status	2227	2335	2366	2392	2413	2434	2454	2480	2517
IDEA Indicator	2219	2328	2358	2386	2409	2430	2454	2485	2536
Section 504 Status	2346	2402	2435	2462	2488	2516	2545	2575	2618
Economic Disadvantage Status	2319	2365	2400	2425	2449	2473	2499	2530	2572

TABLE 5.129. GRADE 6 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2344	2408	2448	2482	2510	2539	2569	2603	2647
Female	2354	2415	2455	2487	2514	2543	2571	2604	2646
Male	2333	2400	2441	2476	2506	2536	2567	2602	2647
American Indian or Alaska Native	2268	2357	2401	2433	2462	2488	2514	2546	2588
Asian	2421	2485	2526	2560	2589	2616	2643	2673	2717
Black/African American	2235	2351	2398	2427	2459	2486	2511	2544	2587
Native Hawaiian or Pacific Islander	2331	2389	2425	2455	2483	2509	2537	2568	2609
Hispanic/Latino Ethnicity	2325	2380	2417	2447	2474	2499	2526	2557	2598
White	2371	2440	2484	2512	2541	2567	2593	2622	2662
Two or More Races	2338	2400	2439	2472	2501	2529	2560	2593	2639
Unidentified Race	2399	2452	2489	2521	2548	2574	2599	2628	2666
LEP Status	2235	2339	2372	2401	2424	2448	2472	2499	2537
IDEA Indicator	2235	2322	2358	2387	2412	2437	2464	2496	2545
Section 504 Status	2356	2416	2454	2485	2512	2538	2567	2599	2643
Economic Disadvantage Status	2326	2381	2418	2448	2475	2501	2529	2560	2602

TABLE 5.130. GRADE 7 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2250	2400	2453	2494	2529	2559	2589	2623	2671
Female	2250	2415	2465	2503	2534	2563	2592	2625	2671
Male	2250	2385	2441	2485	2522	2554	2585	2622	2670
American Indian or Alaska Native	2250	2250	2385	2430	2467	2501	2533	2568	2611
Asian	2426	2507	2552	2586	2615	2643	2671	2705	2751
Black/African American	2250	2250	2379	2428	2465	2501	2531	2564	2608
Native Hawaiian or Pacific Islander	2250	2373	2423	2464	2496	2525	2555	2585	2627
Hispanic/Latino Ethnicity	2250	2371	2421	2456	2487	2516	2545	2576	2618
White	2250	2432	2490	2529	2559	2585	2614	2645	2687
Two or More Races	2250	2396	2450	2490	2523	2553	2582	2616	2662
Unidentified Race	2394	2463	2505	2540	2568	2593	2618	2648	2690
LEP Status	2250	2250	2363	2403	2433	2460	2488	2519	2560
IDEA Indicator	2250	2250	2346	2391	2422	2450	2478	2512	2560
Section 504 Status	2250	2410	2462	2501	2532	2560	2589	2624	2668
Economic Disadvantage Status	2250	2375	2423	2459	2490	2519	2548	2580	2622

TABLE 5.131. GRADE 8 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2350	2424	2472	2509	2540	2573	2608	2647	2702
Female	2370	2440	2485	2515	2548	2580	2613	2652	2703
Male	2332	2409	2459	2499	2530	2565	2601	2644	2700
American Indian or Alaska Native	2285	2370	2416	2454	2489	2515	2548	2583	2634
Asian	2455	2520	2569	2609	2642	2675	2707	2742	2793
Black/African American	2271	2367	2417	2455	2491	2515	2545	2581	2631
Native Hawaiian or Pacific Islander	2335	2402	2446	2482	2513	2542	2573	2609	2659
Hispanic/Latino Ethnicity	2328	2394	2438	2472	2503	2531	2560	2595	2644
White	2376	2456	2504	2535	2568	2600	2632	2669	2718
Two or More Races	2350	2420	2467	2504	2535	2567	2602	2641	2694
Unidentified Race	2422	2483	2524	2557	2587	2615	2645	2678	2722
LEP Status	2265	2332	2375	2410	2441	2470	2499	2531	2581
IDEA Indicator	2265	2320	2360	2392	2421	2449	2481	2519	2571
Section 504 Status	2364	2429	2474	2510	2540	2571	2604	2642	2698
Economic Disadvantage Status	2331	2395	2438	2472	2503	2533	2564	2600	2651

TABLE 5.132. GRADE 11 MATHEMATICS CLAIM 3 SCALE SCORE BY SELECTED DEMOGRAPHIC GROUPS, 2014-2015 PERCENTILES

	Percentiles								
	10	20	30	40	50	60	70	80	90
Total	2351	2439	2485	2524	2559	2595	2634	2677	2735
Female	2365	2448	2494	2531	2565	2599	2636	2676	2729
Male	2340	2430	2477	2516	2552	2590	2631	2678	2742
American Indian or Alaska Native	2305	2398	2449	2483	2517	2548	2584	2625	2679
Asian	2461	2539	2593	2636	2673	2706	2740	2778	2833
Black/African American	2304	2393	2442	2476	2508	2538	2571	2610	2664
Native Hawaiian or Pacific Islander	2332	2414	2462	2496	2527	2557	2592	2632	2681
Hispanic/Latino Ethnicity	2325	2417	2461	2496	2528	2559	2592	2631	2682
White	2379	2460	2508	2548	2584	2621	2658	2699	2754
Two or More Races	2340	2423	2471	2509	2544	2579	2617	2660	2721
Unidentified Race	2419	2493	2541	2579	2614	2647	2677	2712	2757
LEP Status	2280	2351	2410	2444	2473	2501	2530	2563	2612
IDEA Indicator	2280	2346	2402	2437	2465	2492	2520	2552	2599
Section 504 Status	2354	2439	2484	2522	2556	2593	2631	2672	2732
Economic Disadvantage Status	2324	2414	2460	2495	2528	2559	2593	2634	2688

Modes of Administration

Table 5.133 through Table 5.146 present counts of summative assessment administrations by mode⁹. These counts 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 5.133 COUNT OF GRADE 3 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	782806	717957	.917	61581	.079	3268	.004
Female	382293	350938	.918	30105	.079	1250	.003
Male	399802	366991	.918	31476	.079	1335	.003
American Indian or Alaska Native	9670	8641	.894	843	.087	186	.019
Asian	56792	54282	.956	2406	.042	104	.002
Black/African American	44683	37730	.844	6856	.153	97	.002
Native Hawaiian or Pacific Islander	7567	7474	.988	49	.006	44	.006
Hispanic/Latino Ethnicity	277094	268836	.970	8061	.029	197	.001
White	304433	261285	.858	41390	.136	1758	.006
Two or More Races	65928	63087	.957	1976	.030	865	.013
Unidentified Race	16639	16622	.999	0	.000	17	.001
LEP Status	186674	180483	.967	5721	.031	470	.003
IDEA Indicator	75751	75396	.995	0	.000	355	.005
Section 504 Status	5730	5489	.958	203	.035	38	.007
Economic Disadvantage Status	421057	419517	.996	0	.000	1540	.004

⁹ Data used for mode counts provided by the following Consortium members: Delaware, Hawaii, Idaho, Maine, Oregon, South Dakota, US Virgin Islands, Vermont, Washington, West Virginia, California, Montana, Nevada, and North Dakota.

TABLE 5.134 COUNT OF GRADE 4 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	766356	702471	.917	61162	.080	2723	.004
Female	374942	343959	.917	29961	.080	1022	.003
Male	390741	358482	.917	31201	.080	1058	.003
American Indian or Alaska Native	9457	8437	.892	835	.088	185	.020
Asian	57946	55469	.957	2356	.041	121	.002
Black/African American	43775	37015	.846	6636	.152	124	.003
Native Hawaiian or Pacific Islander	8476	8365	.987	65	.008	46	.005
Hispanic/Latino Ethnicity	266953	259357	.972	7426	.028	170	.001
White	302819	259314	.856	42144	.139	1361	.004
Two or More Races	60467	58059	.960	1700	.028	708	.012
Unidentified Race	16463	16455	1.000	0	.000	8	.000
LEP Status	149567	144563	.967	4636	.031	368	.002
IDEA Indicator	79487	79170	.996	0	.000	317	.004
Section 504 Status	7140	6807	.953	311	.044	22	.003
Economic Disadvantage Status	408073	406691	.997	0	.000	1382	.003

TABLE 5.135 COUNT OF GRADE 5 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	764217	700103	.916	61633	.081	2481	.003
Female	373717	342701	.917	30060	.080	956	.003
Male	389937	357385	.917	31573	.081	979	.003
American Indian or Alaska Native	9836	8825	.897	821	.083	190	.019
Asian	59252	56817	.959	2347	.040	88	.001
Black/African American	44022	37427	.850	6484	.147	111	.003
Native Hawaiian or Pacific Islander	8611	8533	.991	40	.005	38	.004
Hispanic/Latino Ethnicity	260467	252973	.971	7358	.028	136	.001
White	307357	263132	.856	42938	.140	1287	.004
Two or More Races	58203	55943	.961	1645	.028	615	.011
Unidentified Race	16469	16453	.999	0	.000	16	.001
LEP Status	124678	121034	.971	3347	.027	297	.002
IDEA Indicator	80527	80226	.996	0	.000	301	.004
Section 504 Status	8619	8262	.959	335	.039	22	.003
Economic Disadvantage Status	399483	398210	.997	0	.000	1273	.003

TABLE 5.136 COUNT OF GRADE 6 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	752233	689794	.917	60704	.081	1735	.002
Female	367982	337575	.917	29663	.081	744	.002
Male	384001	352196	.917	31041	.081	764	.002
American Indian or Alaska Native	9097	8157	.897	791	.087	149	.016
Asian	58488	56194	.961	2211	.038	83	.001
Black/African American	43737	37350	.854	6286	.144	101	.002
Native Hawaiian or Pacific Islander	8157	8090	.992	47	.006	20	.002
Hispanic/Latino Ethnicity	255316	248158	.972	7078	.028	80	.000
White	304844	261172	.857	42794	.140	878	.003
Two or More Races	55717	53802	.966	1497	.027	418	.008
Unidentified Race	16877	16871	1.000	0	.000	6	.000
LEP Status	94276	91489	.970	2652	.028	135	.001
IDEA Indicator	75975	75781	.997	0	.000	194	.003
Section 504 Status	9348	8894	.951	431	.046	23	.002
Economic Disadvantage Status	387945	387108	.998	0	.000	837	.002

TABLE 5.137 COUNT OF GRADE 7 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	744522	682535	.917	60424	.081	1563	.002
Female	364756	334357	.917	29689	.081	710	.002
Male	379607	348124	.917	30735	.081	748	.002
American Indian or Alaska Native	9028	8067	.894	820	.091	141	.016
Asian	56771	54559	.961	2135	.038	77	.001
Black/African American	44033	37716	.857	6224	.141	93	.002
Native Hawaiian or Pacific Islander	7906	7834	.991	37	.005	35	.004
Hispanic/Latino Ethnicity	252924	246050	.973	6841	.027	33	.000
White	304022	260345	.856	42897	.141	780	.003
Two or More Races	53154	51281	.965	1470	.028	403	.008
Unidentified Race	16684	16683	1.000	0	.000	1	.000
LEP Status	82424	79595	.966	2715	.033	114	.001
IDEA Indicator	72214	72059	.998	0	.000	155	.002
Section 504 Status	10277	9787	.952	452	.044	38	.004
Economic Disadvantage Status	379803	379038	.998	0	.000	765	.002

TABLE 5.138 COUNT OF GRADE 8 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	744513	682316	.916	60691	.082	1506	.002
Female	364031	333758	.917	29584	.081	689	.002
Male	380376	348531	.916	31107	.082	738	.002
American Indian or Alaska Native	8799	7872	.895	787	.089	140	.016
Asian	56926	54663	.960	2178	.038	85	.001
Black/African American	45057	38843	.862	6139	.136	75	.002
Native Hawaiian or Pacific Islander	7424	7366	.992	37	.005	21	.003
Hispanic/Latino Ethnicity	252043	245258	.973	6751	.027	34	.000
White	305124	260928	.855	43456	.142	740	.002
Two or More Races	52184	50433	.966	1343	.026	408	.008
Unidentified Race	16956	16953	1.000	0	.000	3	.000
LEP Status	74240	71158	.958	2948	.040	134	.002
IDEA Indicator	70544	70371	.998	0	.000	173	.002
Section 504 Status	11256	10767	.957	453	.040	36	.003
Economic Disadvantage Status	374273	373546	.998	0	.000	727	.002

TABLE 5.139 COUNT OF GRADE 11 MATHEMATICS ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	560409	559225	.998	0	.000	1184	.002
Female	275010	274453	.998	0	.000	557	.002
Male	285360	284751	.998	0	.000	609	.002
American Indian or Alaska Native	5900	5743	.973	0	.000	157	.027
Asian	47957	47901	.999	0	.000	56	.001
Black/African American	32179	32111	.998	0	.000	68	.002
Native Hawaiian or Pacific Islander	5761	5754	.999	0	.000	7	.001
Hispanic/Latino Ethnicity	216943	216934	1.000	0	.000	9	.000
White	203135	202474	.997	0	.000	661	.003
Two or More Races	31400	31175	.993	0	.000	225	.007
Unidentified Race	17134	17133	1.000	0	.000	1	.000
LEP Status	43528	43466	.999	0	.000	62	.001
IDEA Indicator	47863	47705	.997	0	.000	158	.003
Section 504 Status	8421	8383	.995	0	.000	38	.005
Economic Disadvantage Status	286187	285749	.998	0	.000	438	.002

TABLE 5.140 COUNT OF GRADE 3 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	779399	712543	.914	61568	.079	5288	.007
Female	380682	348329	.915	30110	.079	2243	.006
Male	398016	364184	.915	31458	.079	2374	.006
American Indian or Alaska Native	9566	8526	.891	842	.088	198	.021
Asian	56005	53452	.954	2373	.042	180	.003
Black/African American	44691	37639	.842	6904	.154	148	.003
Native Hawaiian or Pacific Islander	7525	7398	.983	47	.006	80	.011
Hispanic/Latino Ethnicity	276364	268171	.970	8006	.029	187	.001
White	303194	258812	.854	41422	.137	2960	.010
Two or More Races	65457	61968	.947	1974	.030	1515	.023
Unidentified Race	16597	16577	.999	0	.000	20	.001
LEP Status	184455	178006	.965	5612	.030	837	.005
IDEA Indicator	75693	75148	.993	0	.000	545	.007
Section 504 Status	5698	5429	.953	203	.036	66	.012
Economic Disadvantage Status	419490	416920	.994	0	.000	2570	.006

TABLE 5.141 COUNT OF GRADE 4 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	764523	700645	.916	61168	.080	2710	.004
Female	374175	343183	.917	29964	.080	1028	.003
Male	389690	357431	.917	31204	.080	1055	.003
American Indian or Alaska Native	9465	8434	.891	836	.088	195	.021
Asian	57252	54802	.957	2331	.041	119	.002
Black/African American	43843	37033	.845	6686	.152	124	.003
Native Hawaiian or Pacific Islander	8467	8357	.987	65	.008	45	.005
Hispanic/Latino Ethnicity	266212	258685	.972	7371	.028	156	.001
White	302834	259282	.856	42172	.139	1380	.005
Two or More Races	60047	57657	.960	1707	.028	683	.011
Unidentified Race	16403	16395	1.000	0	.000	8	.000
LEP Status	147346	142465	.967	4528	.031	353	.002
IDEA Indicator	79361	79068	.996	0	.000	293	.004
Section 504 Status	7142	6808	.953	313	.044	21	.003
Economic Disadvantage Status	406855	405484	.997	0	.000	1371	.003

TABLE 5.142 COUNT OF GRADE 5 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	763552	699414	.916	61645	.081	2493	.003
Female	373374	342349	.917	30061	.081	964	.003
Male	389622	357047	.916	31584	.081	991	.003
American Indian or Alaska Native	9839	8819	.896	826	.084	194	.020
Asian	58635	56229	.959	2318	.040	88	.002
Black/African American	44113	37472	.849	6527	.148	114	.003
Native Hawaiian or Pacific Islander	8596	8519	.991	40	.005	37	.004
Hispanic/Latino Ethnicity	260010	252575	.971	7305	.028	130	.000
White	307906	263610	.856	42981	.140	1315	.004
Two or More Races	58007	55760	.961	1648	.028	599	.010
Unidentified Race	16446	16430	.999	0	.000	16	.001
LEP Status	122794	119258	.971	3252	.026	284	.002
IDEA Indicator	80874	80580	.996	0	.000	294	.004
Section 504 Status	8650	8291	.958	337	.039	22	.003
Economic Disadvantage Status	399128	397856	.997	0	.000	1272	.003

TABLE 5.143 COUNT OF GRADE 6 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	753808	691255	.917	60910	.081	1643	.002
Female	368799	338342	.917	29759	.081	698	.002
Male	384770	352890	.917	31151	.081	729	.002
American Indian or Alaska Native	9179	8235	.897	798	.087	146	.016
Asian	58007	55740	.961	2186	.038	81	.001
Black/African American	43882	37409	.852	6373	.145	100	.002
Native Hawaiian or Pacific Islander	8170	8107	.992	47	.006	16	.002
Hispanic/Latino Ethnicity	255474	248264	.972	7143	.028	67	.000
White	306586	262880	.857	42864	.140	842	.003
Two or More Races	55676	53792	.966	1499	.027	385	.007
Unidentified Race	16834	16828	1.000	0	.000	6	.000
LEP Status	92748	90008	.970	2619	.028	121	.001
IDEA Indicator	76373	76192	.998	0	.000	181	.002
Section 504 Status	9392	8936	.951	436	.046	20	.002
Economic Disadvantage Status	387972	387198	.998	0	.000	774	.002

TABLE 5.144 COUNT OF GRADE 7 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	741949	679806	.916	60670	.082	1473	.002
Female	363518	333038	.916	29803	.082	677	.002
Male	378284	346718	.917	30867	.082	699	.002
American Indian or Alaska Native	8985	8022	.893	820	.091	143	.016
Asian	56107	53903	.961	2128	.038	76	.001
Black/African American	44094	37668	.854	6334	.144	92	.002
Native Hawaiian or Pacific Islander	7798	7727	.991	37	.005	34	.004
Hispanic/Latino Ethnicity	251628	244720	.973	6886	.027	22	.000
White	303987	260259	.856	42992	.141	736	.002
Two or More Races	52716	50873	.965	1473	.028	370	.007
Unidentified Race	16634	16634	1.000	0	.000	0	.000
LEP Status	80315	77525	.965	2691	.034	99	.001
IDEA Indicator	72160	72007	.998	0	.000	153	.002
Section 504 Status	10274	9778	.952	456	.044	40	.004
Economic Disadvantage Status	741949	679806	.916	60670	.082	1473	.002

TABLE 5.145 COUNT OF GRADE 8 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	745158	682723	.916	60981	.082	1454	.002
Female	364442	334068	.917	29704	.082	670	.002
Male	380622	348630	.916	31277	.082	715	.002
American Indian or Alaska Native	8833	7899	.894	791	.090	143	.016
Asian	56587	54348	.960	2158	.038	81	.001
Black/African American	45284	38933	.860	6266	.138	85	.002
Native Hawaiian or Pacific Islander	7422	7362	.992	38	.005	22	.003
Hispanic/Latino Ethnicity	251419	244618	.973	6775	.027	26	.000
White	306562	262244	.855	43592	.142	726	.002
Two or More Races	52110	50381	.967	1361	.026	368	.007
Unidentified Race	16941	16938	1.000	0	.000	3	.000
LEP Status	72251	69252	.958	2881	.040	118	.002
IDEA Indicator	70603	70431	.998	0	.000	172	.002
Section 504 Status	11310	10820	.957	454	.040	36	.003
Economic Disadvantage Status	373405	372725	.998	0	.000	680	.002

TABLE 5.146 COUNT OF GRADE 11 ELA/LITERACY ADMINISTRATIONS BY MODE FOR SELECTED DEMOGRAPHIC GROUPS

Group	Total	Online Adaptive		Online Fixed		Paper-Pencil	
	N	N	P	N	P	N	P
Total	635110	615224	.969	0	.000	19886	.031
Female	312011	302238	.969	0	.000	9773	.031
Male	323056	312963	.969	0	.000	10093	.031
American Indian or Alaska Native	6999	6571	.939	0	.000	428	.061
Asian	53438	51345	.961	0	.000	2093	.039
Black/African American	35533	34312	.966	0	.000	1221	.034
Native Hawaiian or Pacific Islander	6397	6244	.976	0	.000	153	.024
Hispanic/Latino Ethnicity	217241	217231	1.000	0	.000	10	.000
White	248540	237069	.954	0	.000	11471	.046
Two or More Races	49703	45196	.909	0	.000	4507	.091
Unidentified Race	17259	17256	1.000	0	.000	3	.000
LEP Status	46001	45208	.983	0	.000	793	.017
IDEA Indicator	53299	51707	.970	0	.000	1592	.030
Section 504 Status	11452	10639	.929	0	.000	813	.071
Economic Disadvantage Status	314432	307474	.978	0	.000	6958	.022

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., & Schultz, 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 5 Addendum

Item Drift Analysis

Differential, or systematic, change in item parameter values over time is known as item parameter drift. Item drift is problematic because item invariance is a key assumption in item response theory (IRT) scoring. CRESST conducted a study of item parameter drift appearing in the 2014-2015 operational summative tests.

Items were tested for invariance between the initial item calibration study (2013-2014 field test) and the 2014-2015 summative assessment. Only items with at least 1,000 observed scores in the 2014-2015 online administration were evaluated. If the number of observed scores for a particular item exceeded 10,000, 10,000 cases were randomly sampled. For each item that was evaluated, two IRT calibrations were performed. In the first calibration, the parameters of all items in the pool were fixed to their previous estimates. In the second calibration, the parameters of the focal item were freely estimated (while the parameters of all other items in the pool remained fixed). Because items are selected according to an adaptive algorithm, it was assumed that the ability distributions would differ across the groups of students administered different items. Accordingly, the population mean and variance were freely estimated in both calibrations.

Likelihood ratio tests provided a formal evaluation of the null hypothesis that item parameters were exactly equal to their previous estimates. However, this test tends to be quite sensitive (rejecting the null hypothesis for the majority items), and small differences in item parameters may have a negligible impact on item scores. Accordingly, the weighted absolute area between the expected score curves (wABC; see, e.g., Stucky, Edelen, & Chandra, 2015) was used as the primary criterion for judging the severity of item drift across two operational administrations. Expected score curves were computed from the previous item parameter estimates and the parameters estimated from the 2014-2015 data (the second calibration above).

Overall results, and results across claims and across item types (per grade & subject), are summarized in Table 5.147 to Table 5.160. Across grades and subjects, no items showed $wABC > 0.20$. In all the grades and subjects except for ELA grade 11, the percentage of items with $wABC < 0.05$ was above 90%. ELA grade 11 had around 74% of the items with $wABC$ less than 0.05, and 24% of the items with $wABC$ between 0.05 and 0.10. Across item response formats, the short answer (“sa”) and writing extended response (“wer”) items tended to show higher values of $wABC$ than other item types. However, even for these formats, the average $wABC$ values were quite small.

These results suggest that items functioned quite similarly across the two test administrations (2013-2014 vs. 2014-2015), with only minimal differences in item parameters and the resulting expected score functions.

TABLE 5.147 GRADE 3 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	613	523	.853	.022	.019	566	.926	39	.064	5	.008	1	.002	0	.000
Claim 1	189	154	.815	.018	.015	183	.968	5	.026	1	.005	0	.000	0	.000
Claim 2	190	168	.884	.020	.019	180	.952	7	.037	1	.005	1	.005	0	.000
Claim 3	104	78	.750	.017	.010	103	.990	1	.010	0	.000	0	.000	0	.000
Claim 4	130	123	.946	.032	.026	100	.775	26	.202	3	.023	0	.000	0	.000
Item Type = HTQ	80	72	.900	.018	.019	76	.962	2	.025	1	.013	0	.000	0	.000
Item Type = MC	278	229	.824	.025	.022	245	.884	29	.105	2	.007	1	.004	0	.000
Item Type = EBSR	57	44	.772	.015	.009	57	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MS	117	102	.872	.015	.010	117	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = SA	42	41	.976	.027	.018	37	.881	5	.119	0	.000	0	.000	0	.000
Item Type = MI	20	16	.800	.025	.026	18	.900	1	.050	1	.050	0	.000	0	.000
Item Type = WER	19	19	1.000	.036	.023	16	.842	2	.105	1	.053	0	.000	0	.000

TABLE 5.148 GRADE 4 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	633	552	.872	.024	.021	570	.903	52	.082	7	.011	2	.003	0	.000
Claim 1	152	131	.862	.017	.012	149	.980	3	.020	0	.000	0	.000	0	.000
Claim 2	207	185	.894	.023	.017	195	.942	9	.043	3	.014	0	.000	0	.000
Claim 3	110	83	.755	.016	.010	109	.991	1	.009	0	.000	0	.000	0	.000
Claim 4	164	153	.933	.039	.030	117	.722	39	.241	4	.025	2	.012	0	.000
Item Type = HTQ	121	106	.876	.021	.021	109	.901	11	.091	0	.000	1	.008	0	.000
Item Type = MC	63	63	1.000	.031	.020	52	.825	11	.175	0	.000	0	.000	0	.000
Item Type = EBSR	78	71	.910	.025	.024	68	.872	8	.103	2	.026	0	.000	0	.000
Item Type = MS	268	224	.836	.025	.023	242	.910	18	.068	5	.019	1	.004	0	.000
Item Type = SA	21	16	.762	.012	.010	21	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MI	49	39	.796	.016	.009	49	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = WER	33	33	1.000	.032	.014	29	.879	4	.121	0	.000	0	.000	0	.000

TABLE 5.149 GRADE 5 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	639	569	.890	.023	.019	585	.915	50	.078	3	.005	1	.002	0	.000
Claim 1	171	155	.906	.024	.026	151	.883	17	.099	2	.012	1	.006	0	.000
Claim 2	215	195	.907	.022	.015	202	.940	13	.060	0	.000	0	.000	0	.000
Claim 3	108	84	.778	.020	.012	106	.981	2	.019	0	.000	0	.000	0	.000
Claim 4	145	135	.931	.025	.020	126	.869	18	.124	1	.007	0	.000	0	.000
Item Type = HTQ	131	116	.885	.018	.013	127	.969	4	.031	0	.000	0	.000	0	.000
Item Type = MC	233	204	.876	.023	.015	219	.940	14	.060	0	.000	0	.000	0	.000
Item Type = EBSR	24	21	.875	.014	.009	24	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MS	66	66	1.000	.044	.034	41	.621	22	.333	2	.030	1	.015	0	.000
Item Type = SA	81	69	.852	.019	.019	78	.963	2	.025	1	.012	0	.000	0	.000
Item Type = MI	64	53	.828	.019	.014	59	.922	5	.078	0	.000	0	.000	0	.000
Item Type = WER	40	40	1.000	.023	.015	37	.925	3	.075	0	.000	0	.000	0	.000

TABLE 5.150 GRADE 6 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	599	503	.840	.020	.015	570	.952	29	.048	0	.000	0	.000	0	.000
Claim 1	145	123	.848	.019	.014	139	.959	6	.041	0	.000	0	.000	0	.000
Claim 2	203	167	.823	.020	.016	188	.926	15	.074	0	.000	0	.000	0	.000
Claim 3	116	88	.759	.018	.013	112	.966	4	.034	0	.000	0	.000	0	.000
Claim 4	135	125	.926	.024	.013	131	.970	4	.030	0	.000	0	.000	0	.000
Item Type = HTQ	222	180	.811	.020	.013	216	.973	6	.027	0	.000	0	.000	0	.000
Item Type = MC	89	73	.820	.018	.016	84	.944	5	.056	0	.000	0	.000	0	.000
Item Type = EBSR	135	111	.822	.018	.013	132	.978	3	.022	0	.000	0	.000	0	.000
Item Type = MS	50	50	1.000	.026	.017	45	.900	5	.100	0	.000	0	.000	0	.000
Item Type = SA	57	45	.789	.016	.010	57	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MI	19	17	.895	.023	.020	17	.895	2	.105	0	.000	0	.000	0	.000
Item Type = WER	27	27	1.000	.039	.017	19	.704	8	.296	0	.000	0	.000	0	.000

TABLE 5.151 GRADE 7 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	574	472	.822	.021	.017	534	.930	38	.066	2	.003	0	.000	0	.000
Claim 1	153	125	.817	.020	.018	143	.935	9	.059	1	.007	0	.000	0	.000
Claim 2	196	167	.852	.022	.018	180	.918	15	.077	1	.005	0	.000	0	.000
Claim 3	117	82	.701	.016	.011	114	.974	3	.026	0	.000	0	.000	0	.000
Claim 4	108	98	.907	.025	.019	97	.898	11	.102	0	.000	0	.000	0	.000
Item Type = HTQ	192	156	.813	.020	.015	184	.958	8	.042	0	.000	0	.000	0	.000
Item Type = MC	113	93	.823	.016	.011	111	.982	2	.018	0	.000	0	.000	0	.000
Item Type = EBSR	108	82	.759	.019	.018	104	.963	3	.028	1	.009	0	.000	0	.000
Item Type = MS	13	8	.615	.012	.009	13	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = SA	59	44	.746	.016	.012	57	.966	2	.034	0	.000	0	.000	0	.000
Item Type = MI	63	63	1.000	.028	.019	51	.810	12	.190	0	.000	0	.000	0	.000
Item Type = WER	26	26	1.000	.051	.029	14	.538	11	.423	1	.038	0	.000	0	.000

TABLE 5.152 GRADE 8 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	586	493	.843	.021	.018	547	.935	37	.063	1	.002	0	.000	0	.000
Claim 1	149	132	.886	.017	.013	145	.973	4	.027	0	.000	0	.000	0	.000
Claim 2	195	162	.835	.024	.022	177	.912	17	.088	0	.000	0	.000	0	.000
Claim 3	127	94	.740	.019	.013	123	.969	4	.031	0	.000	0	.000	0	.000
Claim 4	115	105	.913	.024	.019	102	.887	12	.104	1	.009	0	.000	0	.000
Item Type = HTQ	221	183	.828	.020	.013	212	.959	9	.041	0	.000	0	.000	0	.000
Item Type = MC	111	79	.712	.016	.013	108	.973	3	.027	0	.000	0	.000	0	.000
Item Type = EBSR	89	76	.864	.017	.013	85	.966	3	.034	0	.000	0	.000	0	.000
Item Type = MS	44	38	.864	.017	.016	42	.955	2	.045	0	.000	0	.000	0	.000
Item Type = SA	11	10	.909	.015	.013	11	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MI	72	69	.958	.030	.022	62	.861	9	.125	1	.014	0	.000	0	.000
Item Type = WER	38	38	1.000	.044	.029	27	.711	11	.289	0	.000	0	.000	0	.000

TABLE 5.153 GRADE 11 ELA ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	929	865	.931	.036	.026	686	.738	223	.240	19	.020	1	.001	0	.000
Claim 1	164	143	.872	.022	.018	152	.927	11	.067	1	.006	0	.000	0	.000
Claim 2	411	388	.944	.036	.023	305	.742	102	.248	4	.010	0	.000	0	.000
Claim 3	66	59	.894	.033	.019	56	.848	10	.152	0	.000	0	.000	0	.000
Claim 4	288	275	.955	.045	.031	173	.601	100	.347	14	.049	1	.003	0	.000
Item Type = HTQ	341	313	.918	.041	.028	226	.663	104	.305	11	.032	0	.000	0	.000
Item Type = MC	223	208	.933	.025	.017	204	.915	19	.085	0	.000	0	.000	0	.000
Item Type = EBSR	188	177	.941	.031	.021	154	.819	34	.181	0	.000	0	.000	0	.000
Item Type = MS	14	12	.857	.009	.008	14	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = SA	39	31	.795	.024	.022	37	.949	1	.026	1	.026	0	.000	0	.000
Item Type = MI	76	76	1.000	.050	.029	41	.539	30	.395	4	.053	1	.013	0	.000
Item Type = WER	48	48	1.000	.065	.022	10	.208	35	.729	3	.063	0	.000	0	.000

TABLE 5.154 GRADE 3 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	739	681	.922	0.019	.014	719	.974	18	.024	0	.000	1	.001	0	.000
Claim 1	371	342	.922	0.020	.012	363	.981	7	.019	0	.000	0	.000	0	.000
Claim 2	116	108	.931	0.022	.020	109	.940	6	.052	0	.000	1	.009	0	.000
Claim 3	142	133	.937	0.017	.013	138	.972	4	.028	0	.000	0	.000	0	.000
Claim 4	110	98	.891	0.013	.008	109	.991	1	.009	0	.000	0	.000	0	.000
Item Type = HTQ	388	358	.923	0.018	.011	379	.979	8	.021	0	.000	0	.000	0	.000
Item Type = MC	118	107	.907	0.018	.014	115	.975	3	.025	0	.000	0	.000	0	.000
Item Type = EBSR	83	75	.904	0.024	.016	77	.928	6	.072	0	.000	0	.000	0	.000
Item Type = MS	27	23	.852	0.018	.014	26	.963	1	.037	0	.000	0	.000	0	.000
Item Type = SA	31	28	.903	0.011	.005	31	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MI	44	42	.955	0.016	.025	43	.977	0	.000	0	.000	1	.023	0	.000
Item Type = WER	48	48	1.000	0.023	.011	48	1.000	0	.000	0	.000	0	.000	0	.000

TABLE 5.155 GRADE 4 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	901	809	.898	0.019	.014	866	.962	34	.038	0	.000	0	.000	0	.000
Claim 1	527	479	.909	0.020	.015	499	.947	28	.053	0	.000	0	.000	0	.000
Claim 2	127	108	.850	0.016	.012	124	.984	2	.016	0	.000	0	.000	0	.000
Claim 3	127	112	.882	0.018	.012	124	.976	3	.024	0	.000	0	.000	0	.000
Claim 4	120	110	.917	0.016	.011	119	.992	1	.008	0	.000	0	.000	0	.000
Item Type = HTQ	403	362	.898	0.018	.013	392	.975	10	.025	0	.000	0	.000	0	.000
Item Type = MC	131	123	.939	0.020	.014	126	.962	5	.038	0	.000	0	.000	0	.000
Item Type = EBSR	18	16	.889	0.018	.010	18	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MS	132	118	.894	0.016	.010	131	.992	1	.008	0	.000	0	.000	0	.000
Item Type = SA	151	127	.841	0.025	.017	135	.894	16	.106	0	.000	0	.000	0	.000
Item Type = MI	40	40	1.000	0.018	.014	38	.950	2	.050	0	.000	0	.000	0	.000
Item Type = WER	26	23	.885	0.012	.006	26	1.000	0	.000	0	.000	0	.000	0	.000

TABLE 5.156 GRADE 5 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	877	784	.894	0.021	.016	817	.932	60	.068	0	.000	0	.000	0	.000
Claim 1	482	440	.913	0.024	.018	431	.894	51	.106	0	.000	0	.000	0	.000
Claim 2	108	92	.852	0.013	.009	108	1.000	0	.000	0	.000	0	.000	0	.000
Claim 3	157	145	.924	0.019	.014	151	.962	6	.038	0	.000	0	.000	0	.000
Claim 4	130	107	.823	0.016	.012	127	.977	3	.023	0	.000	0	.000	0	.000
Item Type = HTQ	411	360	.876	0.018	.015	387	.942	24	.058	0	.000	0	.000	0	.000
Item Type = MC	124	111	.895	0.019	.014	121	.976	3	.024	0	.000	0	.000	0	.000
Item Type = EBSR	203	187	.921	0.028	.018	171	.842	32	.158	0	.000	0	.000	0	.000
Item Type = MS	63	56	.889	0.022	.013	62	.984	1	.016	0	.000	0	.000	0	.000
Item Type = SA	19	19	1.000	0.017	.009	19	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MI	17	12	.706	0.010	.010	17	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = WER	40	39	.975	0.017	.009	40	1.000	0	.000	0	.000	0	.000	0	.000

TABLE 5.157 GRADE 6 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	862	717	.832	0.015	.012	843	.978	19	.022	0	.000	0	.000	0	.000
Claim 1	529	440	.832	0.016	.013	514	.972	15	.028	0	.000	0	.000	0	.000
Claim 2	106	91	.858	0.013	.009	105	.991	1	.009	0	.000	0	.000	0	.000
Claim 3	135	118	.874	0.015	.011	134	.993	1	.007	0	.000	0	.000	0	.000
Claim 4	92	68	.739	0.011	.012	90	.978	2	.022	0	.000	0	.000	0	.000
Item Type = HTQ	106	93	.877	0.015	.012	103	.972	3	.028	0	.000	0	.000	0	.000
Item Type = MC	412	330	.801	0.013	.011	406	.985	6	.015	0	.000	0	.000	0	.000
Item Type = EBSR	103	89	.864	0.017	.011	100	.971	3	.029	0	.000	0	.000	0	.000
Item Type = MS	94	76	.809	0.017	.011	93	.989	1	.011	0	.000	0	.000	0	.000
Item Type = SA	48	44	.917	0.020	.015	46	.958	2	.042	0	.000	0	.000	0	.000
Item Type = MI	19	17	.895	0.017	.018	18	.947	1	.053	0	.000	0	.000	0	.000
Item Type = WER	80	68	.850	0.018	.014	77	.963	3	.038	0	.000	0	.000	0	.000

TABLE 5.158 GRADE 7 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	762	608	.798	0.011	.008	761	.999	1	.001	0	.000	0	.000	0	.000
Claim 1	461	359	.779	0.012	.009	460	.998	1	.002	0	.000	0	.000	0	.000
Claim 2	100	84	.840	0.010	.006	100	1.000	0	.000	0	.000	0	.000	0	.000
Claim 3	116	95	.819	0.011	.008	116	1.000	0	.000	0	.000	0	.000	0	.000
Claim 4	85	70	.824	0.009	.006	85	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = HTQ	81	67	.827	0.010	.008	81	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MC	390	302	.774	0.009	.007	390	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = EBSR	83	68	.819	0.015	.008	83	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MS	112	88	.786	0.012	.009	111	.991	1	.009	0	.000	0	.000	0	.000
Item Type = SA	30	30	1.000	0.017	.010	30	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MI	4	3	.750	0.008	.004	4	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = WER	62	50	.806	0.013	.008	62	1.000	0	.000	0	.000	0	.000	0	.000

TABLE 5.159 GRADE 8 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	710	599	.844	0.012	.011	705	.993	4	.006	1	.001	0	.000	0	.000
Claim 1	417	353	.847	0.013	.011	413	.990	4	.010	0	.000	0	.000	0	.000
Claim 2	76	68	.895	0.010	.008	76	1.000	0	.000	0	.000	0	.000	0	.000
Claim 3	135	106	.785	0.012	.013	134	.993	0	.000	1	.007	0	.000	0	.000
Claim 4	82	72	.878	0.010	.006	82	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = HTQ	297	261	.879	0.010	.008	296	.997	1	.003	0	.000	0	.000	0	.000
Item Type = MC	14	11	.786	0.006	.004	14	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = EBSR	105	83	.790	0.010	.008	105	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = MS	73	58	.795	0.009	.005	73	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = SA	130	110	.846	0.019	.013	127	.977	3	.023	0	.000	0	.000	0	.000
Item Type = MI	52	37	.712	0.011	.009	52	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = WER	39	39	1.000	0.021	.019	38	.974	0	.000	1	.026	0	.000	0	.000

TABLE 5.160 GRADE 11 MATH ITEM DRIFT ANALYSIS (2014-15 ADMINISTRATION)

	N items tested	LRT		weighted absolute area between expected score curves (wABC)											
		p < .01		Mean	SD	(.00,.05)		(.05,.10)		(.10,.15)		(.15,.20)		(.20,1.00)	
		N	P			N	P	N	P	N	P	N	P	N	P
All Items	1361	1138	.837	0.020	.016	1283	.944	76	.056	0	.000	0	.000	0	.000
Claim 1	800	655	.820	0.021	.016	747	.936	51	.064	0	.000	0	.000	0	.000
Claim 2	97	86	.887	0.017	.011	96	.990	1	.010	0	.000	0	.000	0	.000
Claim 3	349	294	.842	0.018	.014	339	.971	10	.029	0	.000	0	.000	0	.000
Claim 4	115	103	.896	0.020	.020	101	.878	14	.122	0	.000	0	.000	0	.000
Item Type = HTQ	265	222	.841	0.019	.016	248	.939	16	.061	0	.000	0	.000	0	.000
Item Type = MC	365	303	.830	0.016	.014	350	.959	15	.041	0	.000	0	.000	0	.000
Item Type = EBSR	392	338	.862	0.026	.016	360	.921	31	.079	0	.000	0	.000	0	.000
Item Type = MS	168	131	.780	0.017	.012	164	.976	4	.024	0	.000	0	.000	0	.000
Item Type = SA	110	85	.773	0.017	.015	104	.945	6	.055	0	.000	0	.000	0	.000
Item Type = MI	14	13	.929	0.014	.010	14	1.000	0	.000	0	.000	0	.000	0	.000
Item Type = WER	47	46	.979	0.021	.022	43	.915	4	.085	0	.000	0	.000	0	.000

References

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.

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 2015 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. (Smarter Balanced, 2014c, p. 25)

Session Time

Table 2.1 presents the estimated testing times. These were provided within the *Online Test Administration Manual* (Smarter Balanced, 2014c, pp. 25-26). The estimated times for each session of each content area test provides sufficient time for students to attempt all items.

TABLE 6.1: 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	Classroom Activity (administered prior to the PT)* hrs : mins	Total hrs : mins
English Language Arts/Literacy	3–5	1: 30	2:00	3:30	:30	4:00
	6–8	1:30	2:00	3:30	: 30	4:00
	HS	2:00	2:00	4:00	: 30	4:30
Mathematics	3–5	1:30	1:00	2:30	: 30	3:00
	6–8	2:00	1:00	3:00	: 30	3:30
	HS	2:00	1:30	3:30	: 30	4:00
Both	3–5	3:00	3:00	6:00	1:00	7:00
	6–8	3:30	3:00	6:30	1:00	7:30
	HS	4:00	3:30	7:30	1:00	8:30

* Classroom Activities are designed to fit into a thirty-minute window; however, the time within the window will vary due to the complexity of the topic and individual student needs.

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* (2014i), the *Online Test Administration Manual* (TAM; 2014c), the *Paper Pencil Test Administration Manual for ELA* (2014g), and the *Pencil Paper Test Administration Manual for Mathematics* (2014h). 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* (2014e).

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 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, 2014c, pp. 37-45).

In addition, Smarter Balanced provides tutorials and practice tests¹⁰ 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 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, 2014c, pp. 15-20).

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 (2014e, p.4). 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.

¹⁰ <http://www.smarterbalanced.org/assessments/practice-and-training-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*. 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 all completed, online, adaptive tests for which item data were available. The exposure rate for a given item is the proportion of tests (in the grade and content area) on which the item appeared.

Table 2.2 and Table 2.3 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. Table 2.2 shows that, on average, the same item appeared in 6% the Grade 3 tests, or, in other words, 6% of Grade 3 examinees saw the same item. 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 2.2 shows that the mean and median exposure rates for ELA/literacy items are well below 25%. Table 2.3 shows that the mean and median exposure rates for mathematics items are also well below 25%.

Table 2.4 and Table 2.5 provide further information about the exposure rate by showing the number of items in the operational pool (N) and proportion of items with exposure rates falling into certain ranges (bins with a width of 0.1), including those that were completely unexposed (Unused). Table 2.4 shows the majority of ELA/literacy CAT items had item exposure rates between 0 and 10%. About 60% of the PT items had item exposure rates between 0 and 10%, while the rest were unused. Table 2.5 shows the majority of mathematics CAT items had item exposure rates between 0 and 10%, and

all of the mathematics PT items had item exposure rates between 0 and 10%, while the rest were unused.

In both content areas, there were a handful of items with high exposure rates. This occurred when there were few items measuring elements in the blueprint. There were also items in both content areas that were unused. The psychometric qualities of these items will be further investigated.

TABLE 6.2. SUMMARY OF ELA/LITERACY ITEM EXPOSURE RATES BY GRADE AND COMPONENT (2014-2015 ADMINISTRATION)

Grade	Type	N	Mean	SD	Min	Max	Median
3	CAT	618	.06	.10	.00	.91	.02
4	CAT	598	.07	.10	.00	.84	.03
5	CAT	594	.07	.12	.00	.83	.03
6	CAT	588	.07	.12	.00	.75	.03
7	CAT	545	.08	.12	.00	.77	.02
8	CAT	541	.08	.10	.00	.58	.04
11	CAT	1475	.03	.06	.00	.41	.01
3	PT	65	.04	.04	.00	.08	.06
4	PT	92	.03	.03	.00	.06	.05
5	PT	100	.03	.02	.00	.06	.05
6	PT	68	.04	.03	.00	.08	.04
7	PT	91	.03	.02	.00	.06	.02
8	PT	105	.02	.02	.00	.07	.03
11	PT	120	.02	.02	.00	.04	.01

TABLE 6.3. SUMMARY OF MATHEMATICS ITEM EXPOSURE RATES BY GRADE AND COMPONENT, 2014-2015 ADMINISTRATION

Grade	Type	N	Mean	SD	Min	Max	Median
3	CAT	900	.04	.06	.00	.49	.02
4	CAT	885	.04	.06	.00	.46	.02
5	CAT	837	.04	.05	.00	.37	.03
6	CAT	770	.04	.06	.00	.44	.02
7	CAT	687	.05	.08	.00	.46	.01
8	CAT	655	.05	.07	.00	.42	.02
11	CAT	1781	.02	.08	.00	.99	.00
3	PT	106	.05	.01	.00	.06	.05
4	PT	99	.05	.01	.00	.06	.05
5	PT	88	.06	.01	.00	.08	.07
6	PT	108	.05	.01	.02	.06	.05
7	PT	90	.06	.01	.01	.08	.06
8	PT	94	.05	.01	.00	.06	.05
11	PT	92	.06	.01	.00	.08	.06

TABLE 6.4. PROPORTION OF ELA/LITERACY ITEMS BY EXPOSURE RATES, 2014-2015 ADMINISTRATION

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	900	.02	.78	.12	.04	.03	.01	.00	.00	.00	.00	.00
4	CAT	885	.01	.78	.12	.05	.02	.01	.01	.00	.00	.00	.00
5	CAT	837	.02	.77	.10	.05	.02	.02	.01	.00	.00	.00	.00
6	CAT	770	.04	.76	.10	.04	.02	.01	.01	.00	.01	.00	.00
7	CAT	687	.01	.75	.13	.04	.04	.01	.01	.00	.01	.00	.00
8	CAT	655	.02	.72	.13	.09	.04	.00	.01	.00	.00	.00	.00
11	CAT	1781	.03	.89	.06	.01	.01	.00	.00	.00	.00	.00	.00
3	PT	106	.43	.57	.00	.00	.00	.00	.00	.00	.00	.00	.00
4	PT	99	.42	.58	.00	.00	.00	.00	.00	.00	.00	.00	.00
5	PT	88	.40	.60	.00	.00	.00	.00	.00	.00	.00	.00	.00
6	PT	108	.41	.59	.00	.00	.00	.00	.00	.00	.00	.00	.00
7	PT	90	.42	.58	.00	.00	.00	.00	.00	.00	.00	.00	.00
8	PT	94	.40	.60	.00	.00	.00	.00	.00	.00	.00	.00	.00
11	PT	92	.40	.60	.00	.00	.00	.00	.00	.00	.00	.00	.00

TABLE 6.5. PROPORTION OF MATHEMATICS ITEMS BY EXPOSURE RATES, 2014-2015 ADMINISTRATION

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	900	.03	.84	.11	.02	.00	.00	.00	.00	.00	.00	.00
4	CAT	885	.02	.88	.08	.01	.00	.01	.00	.00	.00	.00	.00
5	CAT	837	.03	.88	.07	.02	.00	.00	.00	.00	.00	.00	.00
6	CAT	770	.02	.87	.08	.02	.01	.00	.00	.00	.00	.00	.00
7	CAT	687	.01	.82	.10	.05	.02	.01	.00	.00	.00	.00	.00
8	CAT	655	.06	.77	.12	.03	.02	.00	.00	.00	.00	.00	.00
11	CAT	1781	.04	.92	.03	.00	.00	.00	.00	.00	.00	.00	.00
3	PT	106	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
4	PT	99	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
5	PT	88	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
6	PT	108	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
7	PT	90	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
8	PT	94	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
11	PT	92	.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Blueprint Fidelity

A key design document of the summative assessments is the test blueprint, which specifies the number and nature of items to be administered. In June 2015, Smarter Balanced conducted a simulation study to examine the blueprint fulfillment for the simulated test. The results of the simulation study are discussed in Chapter 2. A blueprint fidelity study is planned using operational data from the 2016 administration.

Simulations were conducted for both ELA/literacy and mathematics and in all the tested grade levels (3-8 and high school). At each grade level, two ELA/literacy item pools were evaluated: the general pool and the Braille pool. For mathematics, the evaluation in each grade level included general and Braille pools, as well as a Spanish-translated pool. The simulated tests included both the computerized adaptive test (CAT) and performance task (PT) components, thus mimicking the operational summative tests.

For the vast majority of simulees, the CAT engine used in this simulation study was able to satisfy the requirements of the operational blueprints for the CAT component. Satisfying the blueprint with respect to the PT portion, however, was at times more challenging due to a lack of correspondence between stimuli (and the component items) and the blueprint.

Spring 2015 Embedded Field Test Results

Field test items are embedded into the operational administration during the CAT session. This section presents the analyses of the machine-scored (e.g., multiple choice, equation response) and

hand-scored (e.g., short text) field test items that were embedded during the 2015 test administration.

Machine-Scored Items

Nearly 10,000 machine-scored field test items were administered to students during the 2015 test administration. Table 2.6 presents the number of machine-scored field test items administered at each grade level.

TABLE 6.6. NUMBER OF MACHINE-SCORED, FIELD TEST ITEMS ADMINISTERED BY GRADE, SPRING 2015

Grade	ELA	Math
3	656	564
4	638	661
5	646	617
6	639	676
7	639	681
8	616	695
HS	2145	920
Total	5979	4814

Classical Test Analyses

Using the procedures detailed in Chapter 5, classical item statistics were calculated for all field test items. Table 2.7 and Table 2.8 present the average item score (p-value) and the average item-total correlation (point biserial) for all grades in ELA/literacy and mathematics. Table 2.7 shows that average ELA/literacy item score were in the mid-0.3 range for all grade levels, indicating that, on average, the items were difficult for the students. Table 2.7 also shows that the average item-total correlations tended to be in the 0.4 range, indicating that the items tend to differentiate between high and low ability study.

Table 2.8 presents the same information for the mathematics items. The average mathematics item score ranged from a low of 0.19 in Grade 11 to a high of 0.40 in Grade 3, indicating that the items tended to be difficult for students. The mean item-total correlations ranged from 0.49 in Grade 8 to 0.62 in Grade 11.

TABLE 6.7. CLASSICAL ITEM STATISTICS FOR ELA/LITERACY FIELD TEST ITEMS, SPRING 2015

Grade	Number of Items	Average Item Score		Item-Total* Correlation	
		Mean	SD	Mean	SD
3	656	0.35	0.19	0.48	0.16
4	638	0.34	0.20	0.45	0.16
5	646	0.38	0.22	0.45	0.17
6	639	0.34	0.20	0.41	0.17
7	639	0.31	0.20	0.41	0.18
8	616	0.35	0.21	0.41	0.17
11	2145	0.36	0.20	0.42	0.18

*Total is scale score computed from operational items

TABLE 6.8. 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	564	0.40	0.22	0.54	0.17
4	661	0.38	0.21	0.58	0.15
5	617	0.36	0.18	0.56	0.17
6	676	0.29	0.20	0.53	0.19
7	681	0.23	0.17	0.54	0.21
8	695	0.27	0.20	0.49	0.22
11	920	0.19	0.17	0.62	0.20

*Total is scale score computed from operational items

Item Review

The item-level statistics for machine-scored field test items were examined and flagged for data review with the following criteria:

Flags based on item difficulty and score distribution

- low average item score (less than .10)

- high average item score (greater than .95)
- proportion obtaining any score category <0.03

Flags based on item discrimination

- low item-total correlation (less than .30)
- higher mean criterion score for students in a lower score-point category

Flags for multiple choice items

- among higher ability students (top 20% on overall score), more select a distractor than select the key
- higher criterion score mean for students choosing a distractor than the mean for those choosing the key
- positive correlation between distractor and total score

Table 2.9 and Table 2.10 show the number of items flagged based on difficulty and score distribution for ELA/literacy and mathematics, respectively. For ELA/literacy, about 33% of items (28-39% across the grade levels) were flagged based on at least one criterion. For mathematics, about 35% of items (17-50% across the grade levels) were flagged based on at least one criterion.

TABLE 6.9. NUMBER OF ELA/LITERACY FIELD TEST ITEMS FLAGGED BASED ON DIFFICULTY AND SCORE DISTRIBUTION, SPRING 2015

Grade	Number of Items	Average Item Score		Item-total $r < 0.30$	$<3\%$ in any score level
		<0.10	>0.95		
3	656	67	0	90	35
4	638	74	0	101	35
5	646	59	0	107	29
6	639	66	0	143	38
7	639	90	0	169	37
8	616	70	0	136	28
11	2145	235	0	479	69

TABLE 6.10. NUMBER OF MATHEMATICS FIELD TEST ITEMS FLAGGED BASED ON DIFFICULTY AND SCORE DISTRIBUTION, SPRING 2015

Grade	Number of Items	Average Item Score		Item-total $r < 0.30$	<3% in any score level
		<0.10	>0.95		
3	564	39	0	65	8
4	661	54	0	37	8
5	617	45	1	50	6
6	676	117	0	94	32
7	681	173	0	112	49
8	695	166	0	140	54
11	920	359	0	79	148

Differential Item Functioning

Using the procedures detailed in Chapter 3, DIF statistics were computed for all field test items. 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 2.11 and Table 2.12 summarizes 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). Table 2.11 shows that just about 5% of ELA/literacy items were flagged for moderate to large DIF. Table 2.12 shows that right around 5% of mathematics items were flagged for moderate to large DIF.

TABLE 6.11. 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	656	17	A	630	592	608	621	597	628	621	641
			B	23	57	45	31	23	25	33	14
			C	3	7	2	4	5	0	2	1
4	638	17	A	610	578	587	600	585	596	580	620
			B	25	55	50	34	21	41	51	17
			C	3	5	0	4	1	0	6	1
5	646	35	A	600	570	588	602	586	597	580	624
			B	34	68	57	40	26	45	57	21
			C	12	7	0	4	6	3	9	1
6	639	38	A	594	578	574	596	580	602	539	617
			B	34	53	60	36	24	32	84	21
			C	11	6	2	7	6	2	11	1
7	639	27	A	581	595	593	597	578	593	565	616
			B	42	40	43	34	21	42	64	21
			C	15	3	0	7	3	0	3	1
8	616	34	A	551	551	572	579	559	570	529	599
			B	50	58	42	32	18	43	73	15
			C	14	6	0	5	3	0	11	1
11	2145	134	A	1886	1897	1968	1914	1592	1949	1802	2007
			B	215	197	105	204	32	109	203	129
			C	40	24	10	17	13	12	29	5

TABLE 6.12. 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	564	26	A	530	495	514	520	496	523	538	542
			B	30	67	44	40	15	40	25	21
			C	4	2	3	4	12	1	1	1
4	661	18	A	626	571	600	620	579	605	617	628
			B	34	84	54	40	22	53	41	33
			C	1	6	3	1	3	1	3	0
5	617	26	A	579	498	566	571	543	547	571	602
			B	34	109	46	43	17	64	37	13
			C	3	8	1	2	4	4	4	0
6	676	29	A	624	593	603	616	517	595	607	639
			B	51	74	39	55	12	57	40	31
			C	0	8	1	1	7	5	8	2
7	681	42	A	622	584	588	595	451	589	581	622
			B	57	80	29	72	8	41	40	53
			C	2	13	7	10	5	2	8	1
8	695	29	A	655	613	614	635	485	613	604	660
			B	37	72	39	51	13	46	51	32
			C	3	8	2	3	4	3	5	2
11	920	61	A	808	791	722	798	512	679	663	851
			B	88	87	33	80	12	50	47	47
			C	11	24	3	3	12	6	7	3

Hand-Scored Items

Approximately 400 hand-scored items were administered and scored for the Spring 2015 embedded field test. This section provides an overview of the scoring procedures as well as the results of the scoring process.

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 PT (full write), 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 PTs and short text 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 includes a detailed description of the hand scoring process, the rater qualifications, quality monitoring procedures, and rater training information.

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.

Polytomous items are flagged for elimination if any of the following conditions occur:

- Adjacent agreement < 0.80
- Exact agreement < 0.60
- Quadratic weighted Kappa < 0.20

Dichotomous items are flagged for elimination if any of the following conditions occur:

- Exact agreement < 0.80
- Quadratic weighted Kappa < 0.20

Table 13 shows the number of items flagged by subject and grades. There were 10 items flagged across all grades in the two content areas.

TABLE 6.13. NUMBER OF HAND-SCORED FIELD TEST ITEMS FLAGGED BY SUBJECT AREA AND GRADE, SPRING 2015

Subject	Grade	Number of Flagged Items
ELA/Literacy	3	1
ELA/Literacy	6	1
Mathematics	7	4
Mathematics	8	2
Mathematics	11	2

Spring 2015 Embedded Field Test Results

A large proportion of items were successfully field tested and met statistical criteria. With the exception of grade 11 mathematics, where about half of the items were flagged, a majority of items are eligible for use in operational pools without additional review. The flagged items will undergo editorial and data review by panels composed of both content and bias/sensitivity experts. Based on the recommendation of the reviewers, flagged items will either be approved for operational use, rejected, or revised and moved back to the field test pool.

Smarter item pools tend to be difficult. The consortium is undertaking a project to examine items at all levels of difficulty to assess differences associated with difficulty.

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. (2014c, August 25). *Online, summative, test administration manual: Test of English language arts/literacy and mathematics*. Los Angeles, CA: Author.
- Smarter Balanced Assessment Consortium. (2014e, November 5). *Usability, Accessibility, and Accommodations Guidelines*. Los Angeles, CA: Author.
- Smarter Balanced Assessment Consortium. (2014g, December). *English language arts summative assessment: Paper-pencil test administration manual*. Los Angeles, CA: Author.
- Smarter Balanced Assessment Consortium. (2014h, December). *Mathematics summative assessment: Paper-pencil test administration manual*. Los Angeles, CA: Author.
- Smarter Balanced Assessment Consortium. (2014i). *Test administrator user guide*. Los Angeles, 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 sub-scores corresponding to broad areas within the content area domains. The consortium provides a set of reports based on these scores and sub-scores 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, 2014). 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 sub-scores each have an associated standard error of measurement (SEM) that indicates the reliability of the score. (For a fuller explanation of the SEM, refer to Chapter 2.)

Overall Test Scores

Scale scores are the basic units of overall reporting. These scores fall along a continuous vertical scale (from approximately 2000 to 3000) that increases across grade levels and are used to describe an individual student's current level of achievement. (They will also be used to track growth over time, but in this first implementation year, there are no growth reports.) When aggregated, scale scores are used to describe achievement for different groups of students. The method for setting achievement level criteria so that cut scores delineate proficiency levels is explained in Chapter 5.

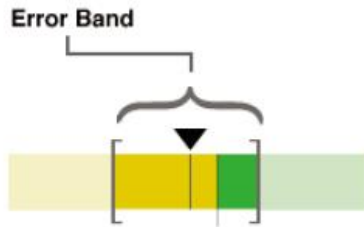
The Smarter Balanced reporting system communicates an overall scale score in relation to Achievement Levels using graphics similar to Figure 7.1. By default, the system uses generic terms for the achievement levels, Level 1, Level 2, Level 3, and Level 4, but members may customize them using terms such 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 based on the SEM. In Figure 7.1, the overall score is 2475, which is in Level 2, and the score's error band encompasses Level 3. 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.



Smarter Balanced tests provide the most precise scores possible within a reasonable time limit, but no test can be 100 percent accurate. The **error band** indicates the range of scores that a student would likely achieve if they were to take the test multiple times. It is similar to the "margin of error" that newspapers report for public opinion surveys.

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.

Smarter Balanced has developed a set of optional Reporting Achievement Level Descriptors (ALDs) 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.

Sub-scores

Sub-scores are scores on important domain areas within each content area. In most case, sub-scores correspond to Claims, but in mathematics, Claims 2 and 4 are so intertwined that they are reported as a single sub-score. The Claims and reporting categories (sub-scores) are primary structural elements in test blueprints and item development. Tables 7.1 and 7.2 provide the claims or sub-score reporting categories for ELA/literacy and mathematics.

TABLE 7.1 ENGLISH LANGUAGE ARTS/LITERACY CLAIMS

<p><i>Claim #1- Reading</i></p> <ul style="list-style-type: none"> Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.
<p><i>Claim #2- Writing</i></p> <ul style="list-style-type: none"> Students can produce effect and well-grounded writing for a range of purposes and audiences.
<p><i>Claim #3- Speaking and Listening</i></p> <ul style="list-style-type: none"> Students can employ effective speaking and listening skills for a range of purposes and audiences. At this time, only listening is assessed.
<p><i>Claim #4- Research</i></p> <ul style="list-style-type: none"> 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

<p><i>Claim #1- Concepts and Procedures</i></p> <ul style="list-style-type: none"> Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency.
<p><i>Claim #2- Problem Solving/ Claim #4- Modeling and Data Analysis</i></p> <ul style="list-style-type: none"> 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
<p><i>Claim #3- Communicating Reasoning</i></p> <ul style="list-style-type: none"> Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.

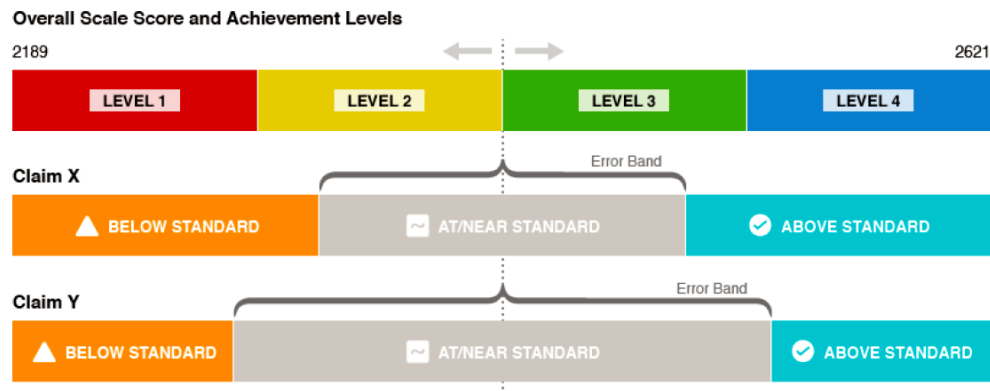
Achievement levels for claims are not established, so sub-scores are not portrayed in achievement levels. In addition, SEMs at the claim/sub-score level are fairly large. Consequently sub-scores are characterized by an indication of whether they are “Below Standard”, “At or Near Standard”, or “Above Standard”. These designations are based on the SEM of the sub-score and the distance of the sub-score from the cut score between levels 2 and 3, which is an indicator of being on-track for college or career readiness. If the Level 2/3 cut score falls within a 1.5 SEM error band, it is designated as “At or Near Standard”. If the Level 2/3 cut score is above the error band, the sub-score is designated as “Below Standard”; if the cut score is below the error band, the sub-score is “Above Standard”.

TABLE 7.3 SUB-SCORE CATEGORIES

Above Standard	Score is > 1.5 SEMs above the Level 2/3 cut score
At or Near Standard	The Level 2/3 cut score falls within an error band of +/- 1.5 SEMs around the sub-score
Below Standard	Score is > 1.5 SEMs below the Level 2/3 cut score

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 SUB-SCORE REPORTING. FROM *REPORTING SYSTEM USER GUIDE*, PP.116-117.



Although sub-scores are portrayed in Smarter Balanced reports by the three-level system above (also called “traffic-light” indicators) sub-score scale scores and SEMs 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/sub-score results along with claim level ALDs. The scores and descriptions provide context for understanding what the assessment has measured and how to interpret the scores and sub-scores. Teachers, students and parents use this report to understand a student’s achievement and progress toward mastery of the CCSS. The report may be part of a larger set of information to provide context for instructional focus. In addition to the overall score displays, sub-scores are reported as shown in Figure 7.4 below.

FIGURE 7.4 ILLUSTRATION OF SUB-SCORE REPORTING ON INDIVIDUAL STUDENT REPORTS

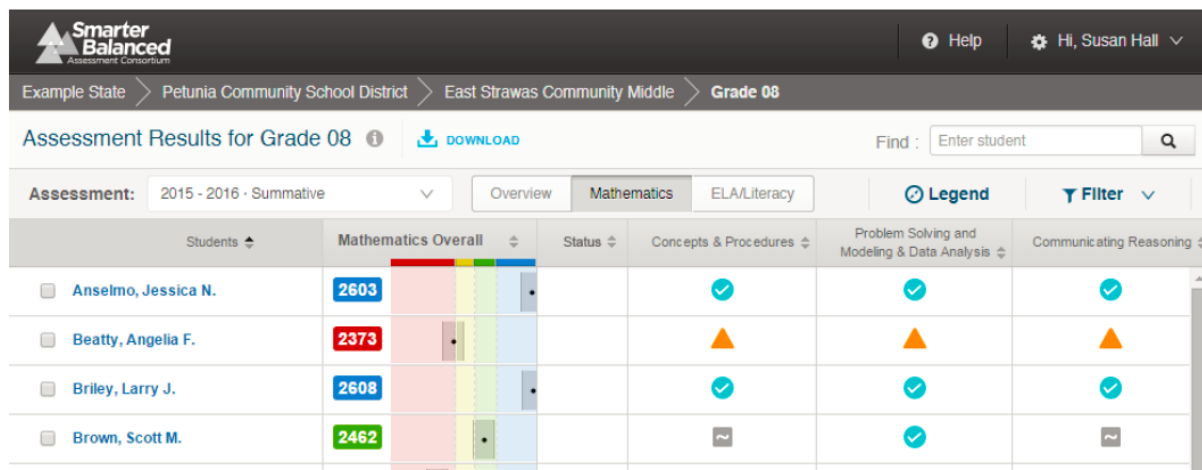


Individual Student Reports can be downloaded as PDF files for easy printing and distribution to parents.

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 sub-score 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 LIST



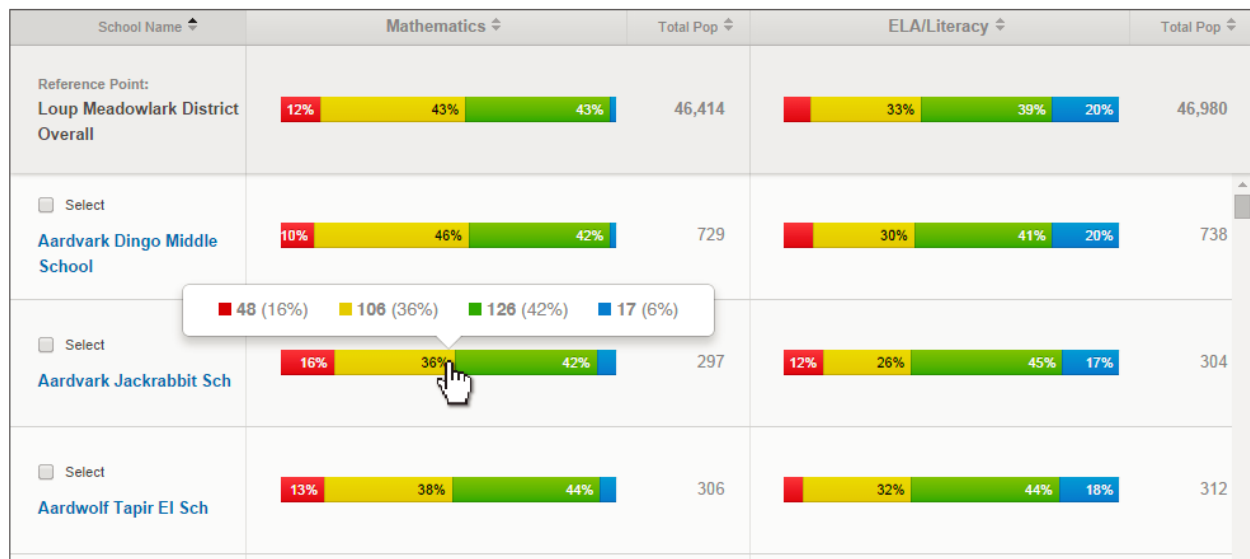
Students	Mathematics Overall	Status	Concepts & Procedures	Problem Solving and Modeling & Data Analysis	Communicating Reasoning
<input type="checkbox"/> Anselmo, Jessica N.	2603		✓	✓	✓
<input type="checkbox"/> Beatty, Angelia F.	2373		▲	▲	▲
<input type="checkbox"/> Briley, Larry J.	2608		✓	✓	✓
<input type="checkbox"/> Brown, Scott M.	2462		~	✓	~

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 7.7 DATA 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-assessment. 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.

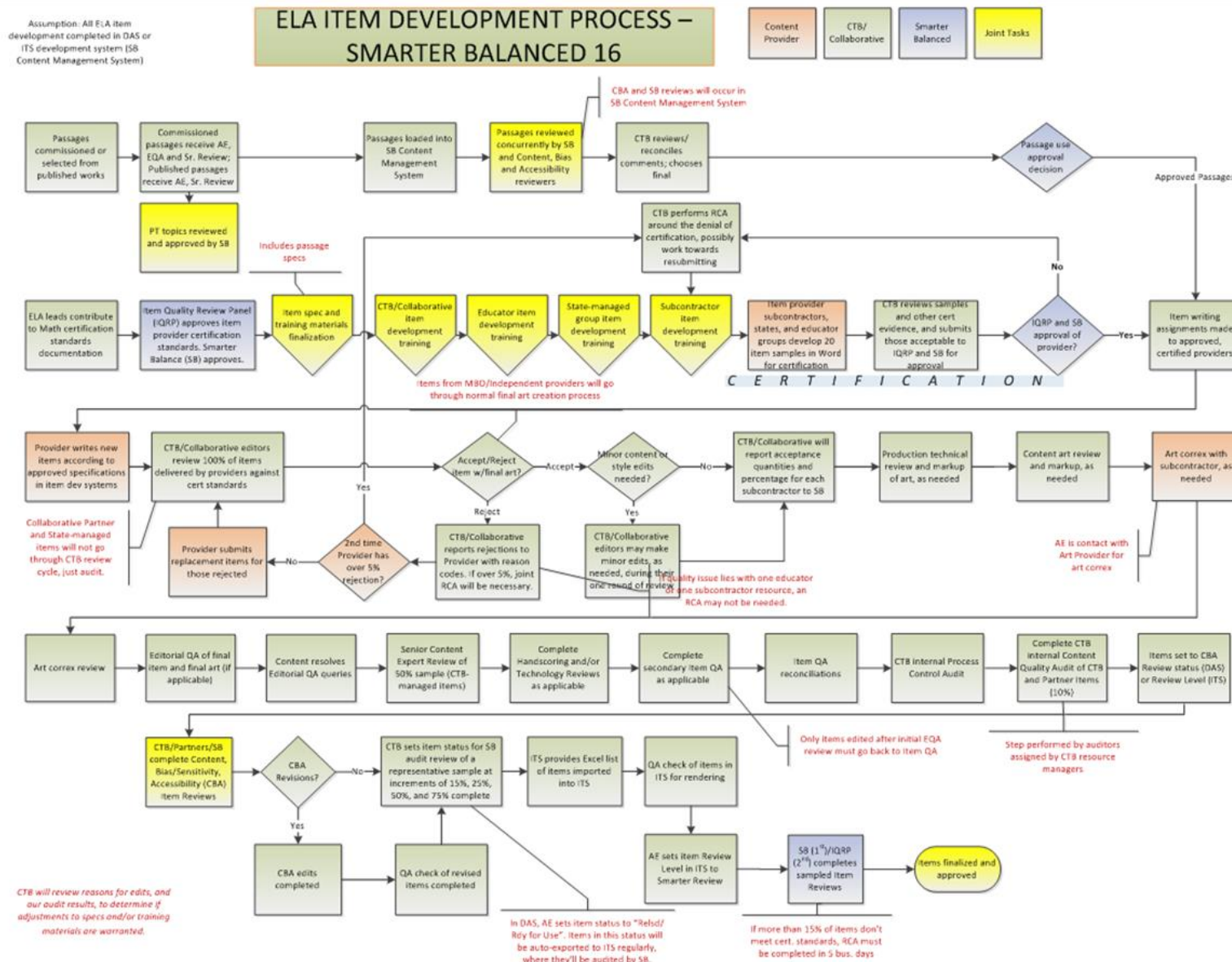
The dynamic nature of the reports, allowing users to sort and filter to get custom information and the provision of customized download data for any kind of analysis, gives Consortium members a rich flexible set of results. By providing capability for multiple reports and downloads, the Smarter Balanced system affords members with a dynamic and flexible system.

References

- Smarter Balanced Assessment Consortium. (2014f, November 14). *Interpretation and Use of Scores and Achievement Levels*. Los Angeles, CA: Author. Retrieved from <https://www.smarterbalanced.org/wp-content/uploads/2015/08/Interpretation-and-Use-of-Scores.pdf>
- Smarter Balanced Assessment Consortium. (2014, September 7). *Reporting System User Guide*. Los Angeles, CA: Author. Retrieved from <http://www.smarterapp.org/manuals/Reporting-UserGuide.html>

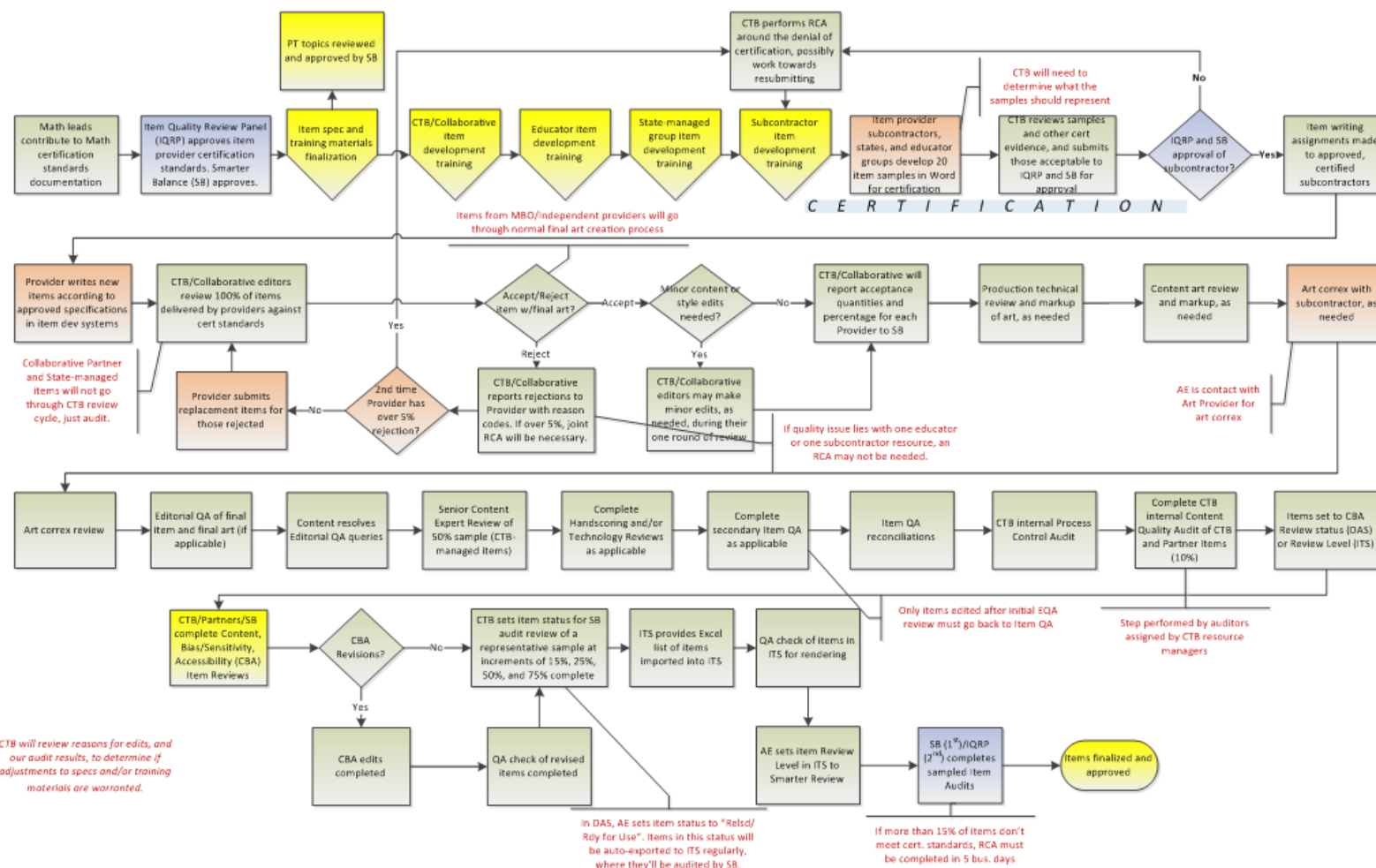
Appendix A: Item Development Process

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.



Assumption: All MATH item development completed in DAS or ITS item development system (SB Content Management System)

MATH ITEM DEVELOPMENT PROCESS – SMARTER BALANCED 16



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 level descriptors 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 Level Descriptors

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 scale 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 Level Descriptors

Descriptors for Smarter Balanced achievement levels provided to Consortium members for the 2014-2015 test administration. Please note that members may choose to alter the descriptors or name the four achievement levels.

Enclosed are the Achievement Level Descriptors 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 Level Descriptors

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	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	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
high school.	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 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 Level Descriptors

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	Level 3 The student has met the achievement standard and demonstrates progress	Level 3 The student has met the achievement standard and demonstrates progress toward mastery of

High School	Grades 6–8	Grades 3–5
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.	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.	the knowledge and skills in English language arts/literacy needed for likely success in future coursework.
<p>Level 2</p> <p>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.</p>	<p>Level 2</p> <p>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.</p>	<p>Level 2</p> <p>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.</p>
<p>Level 1</p> <p>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.</p>	<p>Level 1</p> <p>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.</p>	<p>Level 1</p> <p>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.</p>