PARCC

Final Technical Report for 2018 Administration

Pearson

February 28, 2019

Table of Contents

List of Tables

List of Figures

Executive Summary

What's New for the 2017–2018 PARCC Technical Report

The purpose of this technical report is to describe the fourth operational administration of the Partnership for Assessment of Readiness for College and Careers (PARCC) assessments in the 2017–2018 academic year. The structure and format of this technical report is consistent with the prior technical reports with the following changes or updates:

- **Section 5 – Test Taker Characteristics:** Participation included students from Bureau of Indian Education, District of Columbia, Department of Defense Education Activity, Illinois, Maryland, New Jersey, and New Mexico.
- **Section 9 – Validity:** Section 9.6 includes a summary of a series of four component studies evaluating the usability and effect of a drawing tool for online mathematics items for grades 3 through 5.

Overview

PARCC is a state-led consortium creating next-generation assessments that, compared to traditional K–12 assessments, more accurately measures student progress toward college and career readiness. The PARCC assessments are aligned to the Common Core State Standards (CCSS) and were administered operationally for the first time in the 2014–2015 academic year. PARCC comprises assessments in both English language arts/literacy (ELA/L) and mathematics in grades 3 through 8 and high school.

The information provided in this technical report is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014). The purpose of this technical report is to describe the fourth operational administration of the PARCC assessments in the 2017–2018 academic year and includes the following topics:

- background and purpose of the assessments;
- test development of items and forms;
- test administration, security, and scoring;
- test taker characteristics;
- classical item analyses and differential item functioning;
- reliability and validity of scores;
- item response theory (IRT) calibration and scaling;
- performance level setting;
- development of the score reporting scales and student performance;
- student growth measures; and
- quality control procedures.

Background and Purpose

Assessments for the first operational administration were constructed in 2014. Eleven states and the District of Columbia participated in the first administration of the PARCC assessments during the 2014–2015 school year. A small subset of students was tested in fall 2014. ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered in the fall; these assessments were administered on paper only. The majority of students tested during the spring 2015 window when all grades and content areas were administered online and on paper. Seven states, the Bureau of Indian Education, and District of Columbia participated in the second administration in school year 2015–2016. Not all participating states had students testing in all grades.

In fall 2015 ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2016 window when all grades and content areas were administered online and on paper. Six states, the Bureau of Indian Education, and District of Columbia participated in the third administration in school year 2016–2017. Not all participating states had students testing in all grades.

In fall 2016 ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2017 window when all grades and content areas were administered online and on paper. Four states, the Bureau of Indian Education, District of Columbia, and the Department of Defense Education Activity participated in the fourth administration in school year 2017–2018. Not all participating states had students testing in all grades.

In fall 2017 ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2018 window when all grades and content areas were administered online and on paper.

The PARCC assessments are designed to achieve several purposes. First, the tests are intended to provide evidence to determine whether students are on track for college- and career-readiness. Second, the tests are structured to access the full range of CCSS and measure the total breadth of student performance. Finally, the tests are designed to provide data to help inform classroom instruction, student interventions, and professional development.

Item Types

The tests contain selected response, brief and extended constructed response, technology-enabled, and technology-enhanced items (TEI), as well as performance tasks. Technology-enabled items are single-response or constructed-response items that involve some type of digital stimulus or open-ended response box with which the students engage in answering questions. Technology-enhanced items involve specialized student interactions for collecting performance data. Therefore, the act of performing the task is the way in which data are collected. Students may be asked, among other tasks, to categorize information, organize or classify data, order a series of events, plot data, generate equations, highlight text, or fill in a blank. One example of a TEI is an interaction in which students are asked to drag response options onto a Venn diagram to show the relationship among ideas.

Pre- and Post-Equating

Spring 2017 PARCC assessments (except ELA/L online) were pre-equated, meaning that the scoring tables were based on item parameters estimated using data from earlier administrations. The ELA/L online tests continued to be post-equated due to the structure of these tests. ELA/L paper tests used only established operational items to make pre-equating possible. Mathematics tests have embedded field-test items that provide reliable statistics for new items. This design, coupled with the evolution of the item bank given the longevity of the program, made pre-equating possible. Pre-equating allows for faster reporting of results. In this technical report, item statistics are from the administration the item parameters were estimated that were used to score the students.

Classical and IRT Item Analysis

Classical item analyses and differential item functioning analyses were performed on the data to evaluate the psychometric characteristics of the operational test items after items were administered and before scores were reported. The two-parameter logistic/generalized partial credit (2PL/GPC) IRT model was used for calibrations and scaling. Multiple operational core forms were administered for each grade in ELA/L and mathematics. The forms included sets of embedded common items to provide data to support horizontal linking across test forms within a grade and content area and for post-equated tests across years. Forms that were pre-equated, meaning that the conversion tables are based on item parameters estimated using data from earlier operational administrations, had parameter estimates already on the base IRT scale, so there was no need to create year-to-year common item (linking) sets for these grades/subjects.

For post-equated ELA/L online tests, IRT calibration and scaling placed all operational items for a single grade onto a common scale. After calibration and model fit evaluation was completed, a master list of all items flagged as problematic was compiled and brought to the PARCC Priority Alert Task Force.[1] The task force reviewed each item, its content, and the statistical properties, and made decisions about whether to include the item in the operational scores. Sometimes, an item was rejected because it appeared to have content issues, and sometimes an item was excluded because it could not be calibrated or showed extremely poor IRT model fit. Ultimately the decision about whether to keep or exclude each flagged item was made by the task force. The goals of the task force were to: a) minimize the number of items excluded from the operational test forms, and b) avoid advantaging or disadvantaging any test takers.

In response to several practical constraints based on the number of forms constructed for each mode and to meet the blueprints (e.g., inclusion of TEI on computer-based testing [CBT] forms), there was no single CBT form that was administered intact in the paper delivery mode at any grade level. For example, TEI from online forms were replaced in the paper forms with items having similar content, but appropriate for paper-based testing (PBT). Mode comparability studies were conducted in 2015 and 2017. The results are presented in full in separate reports and are summarized in Section 9 of this report. The studies evaluated the extent to which scores from CBT and PBT forms could be considered as comparable with regard to psychometric characteristics. The studies indicated that score

[1] The Priority Alert Task Force comprised New Meridian staff, state leads and content experts, and Pearson staff.

comparability was inconsistent across the content domains and grade levels investigated. However, when differences were found the differences tended to be within measurement error.

Overall Scale Scores, Claim Scores, and Subclaim Scores

The PARCC ELA/L and mathematics scores are expressed as various types of scale scores (both total scores and claim scores, related to the claims structures described below), as well as by performance levels used to describe how well students meet the academic standards for their grade level. On the basis of a student's total score, an inference is drawn about how much knowledge and skill in the content area the student has acquired. The total score is also used to classify students in terms of the level of knowledge and skill in the content area as students progress in their K–12 education. These levels are called performance levels and are reported as:

- Level 5: Exceeded Expectations
- Level 4: Met Expectations
- Level 3: Approached Expectations
- Level 2: Partially Met Expectations
- Level 1: Did Not Yet Meet Expectations

Students classified as either Level 4 or Level 5 are meeting or exceeding the grade level expectations. Additionally, information on more specific skills is provided and is reported as *Below Expectations*, *Nearly Meets Expectations*, and *Meets or Exceeds Expectations*.

PARCC developed performance level descriptors (PLDs) to assist with the understanding and interpretations of the ELA/L and mathematics scores (https://parcc-assessment.org/performance-levels/). Additionally, resource information is available online to educators, parents, and students (http://understandthescore.org/score-report-guide/), which includes information on understanding and interpreting the ELA/L and mathematics score reports.

The claim structures for ELA/L and mathematics, grounded in the Common Core State Standards, inform the design and development of the summative assessments.

Claim Structure for ELA/L

Master Claim. The master claim is the overall performance goal for the PARCC ELA/L Assessment System—students must demonstrate that they are college- and career-ready or on track to readiness as demonstrated through reading and comprehending grade-level texts of appropriate complexity and writing effectively when using and/or analyzing sources.

Major Claims: 1) reading and comprehending a range of sufficiently complex texts independently, and 2) writing effectively when using and/or analyzing sources.

Subclaims: The subclaims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence tables for reading and writing (https://parcc-assessment.org/ela-literacy). The claims and evidences are grouped into the following categories:

1. Vocabulary, Interpretation, and Use
2. Reading Literature
3. Reading Informational Text
4. Written Expression
5. Knowledge of Language and Conventions

Claim Structure for Mathematics

Master Claim. The degree to which a student is college- or career-ready or on track to being ready in mathematics. The student solves grade-level/course-level problems aligned to the Standards for Mathematical Content with connections to the Standards for Mathematical Practice.

Subclaims: The subclaims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence statement tables for mathematics (https://parcc-assessment.org/mathematics/). The claims and evidence are grouped into the following categories:

Subclaim A: Major Content with Connections to Practices

Subclaim B: Additional and Supporting Content with Connections to Practices

Subclaim C: Highlighted Practices with Connections to Content: Expressing mathematical reasoning by constructing viable arguments, critiquing the reasoning of others, and/or attending to precision when making mathematical statements

Subclaim D: Highlighted Practice with Connections to Content: Modeling/Application by solving real-world problems by applying knowledge and skills articulated in the standards

Scale Scores

Scale scores were defined for each test as a linear transformation of the IRT theta (θ) scale. The test characteristic curves associated with the performance level setting forms were used to identify the theta values associated with the Level 2 and Level 4 point scores. By defining Level 2 and Level 4 scale scores to be 700 and 750, respectively, the linear relationship between theta and scale scores was established.

The result was 201 defined full summative scale score points for each ELA/L and mathematics assessment, ranging from 650 to 850. A scale score of 700 is always the minimum for Level 2 performance, and a scale score of 750 is always the minimum for Level 4 performance.

The thresholds for summative performance levels on the scale score metric recommended by the scale score task force are described in Table 1.1.

Table 1.1 Defined Summative Scale Scores and Cut Scores

	Lowest Obtainable Scale Score	Cut Score Level 2	Cut Score Level 4	Highest Obtainable Scale Score
Full Summative	650	700	750	850

As with the full summative scores, scale scores for reading and writing were defined for each test as a linear transformation of the IRT theta (θ) scale. The same IRT theta scale was used for reading and writing as was used for the ELA/L full summative scores. The theta values associated with the Level 2 and Level 4 performance levels were identified using the test characteristic curves associated with the performance level setting forms. Parallel to the full summative scores, the relationship between theta and scale scores was established with Level 2 and Level 4 theta scores and the corresponding predefined scale scores.

The result was 81 defined scale score points for reading, ranging from 10 to 90. A scale score of 30 is the cut score for minimum Level 2 performance, and a scale score of 50 is the cut score for minimum Level 4 performance. There are 51 defined scale score points for writing, ranging from 10 to 60. A scale score of 25 is the cut score for minimum Level 2 performance, and a scale score of 35 is the cut score for minimum Level 4 performance. The threshold reading and writing performance levels on the scale score metric recommended by the scale score task force are described in Table 1.2.

Table 1.2 Defined Scaled Scores and Cut Scores for Reading and Writing Claim Scores

	Lowest Obtainable Scale Score	Cut Score Level 2	Cut Score Level 4	Highest Obtainable Scale Score
Reading	10	30	50	90
Writing	10	25	35	60

Regarding the subclaim scores, the Level 4 cut is defined as *Meets or Exceeds Expectations* because grade 3–8 and high school students at Level 4 or above are likely to have the skills and knowledge to meet the definition of career and college readiness. Subclaim outcomes center on that performance level and are reported as *Below Expectations*, *Nearly Meets Expectations*, and *Meets or Exceeds Expectations*.

Quality Control

To ensure IRT calibrations, scaling and conversion tables were produced accurately, Human Resources Research Organization (HumRRO) replicated the data processing, IRT calibrations, and scale score transformations carried out by Pearson, and the generation of the score conversion tables. Pearson and HumRRO independently generated incomplete data matrices and conducted the calibrations using IRTPRO (Cai, Thissen & du Toit, 2011) calibration software. Pearson's scaling constants were compared to those generated by HumRRO and found to be consistent. Measured Progress (MP) performed

independent quality control comparisons between the Pearson and HumRRO item parameter estimates to identify any differences. In addition, MP independently made certain that the same items were excluded from the linking sets and compared transformed parameter estimates computed by Pearson and HumRRO. If items had large differences across years or modes, the items were discussed and any remaining issues resolved. MP prepared reports documenting their findings. Exact matches were found between all Pearson and HumRRO conversion tables before scores were reported.

Section 1: Introduction

1.1 Background

States associated with the Partnership for Assessment of Readiness for College and Careers (PARCC) came together in early 2010 with a shared vision of ensuring that all students—regardless of income, family background, or geography—have equal access to a world-class education that will prepare them for success after high school in college and/or careers. The PARCC goal was to develop new assessments that tie into more rigorous academic expectations and help prepare students for success in college and the workforce, as well as to provide information back to teachers and parents about where students are on their path to success. Calling on the expertise of thousands of teachers, higher education faculty, and other educators in multiple states, the PARCC assessment system is a high-quality set of summative assessments, diagnostic assessments, formative tasks, and other support materials for teachers including professional development and communications tools.

The PARCC consortium develops and administers next-generation assessments that, compared to traditional K–12 assessments, more accurately measure student progress toward college and career readiness. The assessments are aligned to the Common Core State Standards (CCSS) and include both English language arts/literacy (ELA/L) assessments (grades 3 through 11) and mathematics assessments (grades 3 through 8 and high school). Compared to traditional standardized tests, these assessments are intended to measure more complex skills like critical-thinking, persuasive writing, and problem-solving.

In 2013, the PARCC Governing Board launched Parcc Inc., a non-profit organization designed to support the successful delivery of the tests in 2014–2017, and the long-term success of the multi-state partnership. States continue to govern decisions about the assessment system; the non-profit organization was their "agent" for overseeing the many vendors involved in the PARCC assessment system, coordinating the multiple work groups and committees (including Governing Board meetings), managing the PARCC intellectual property, overseeing the research agenda and the Technical Advisory Committee, and developing and launching the multiple non-summative tools.

Following the Parcc, Inc. contract ending in June 2017, PARCC states released the intellectual property (IP) of the PARCC contract to the Council of Chief State School Officers (CCSSO), and also contracted with New Meridian Corporation to manage the IP and provide item development, forms construction, and governance. Starting in August 2017, New Meridian oversaw the phase 7 item development, the data review activities from the items field tested in the spring 2018 administration, and test construction activities for the spring 2019 administration.

Summative assessments for the first operational administration were constructed in 2014. Eleven states including the District of Columbia participated in the first administration of the PARCC assessments during the 2014–2015 school year. A small subset of students tested in the fall 2014 window. ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered in the fall; these assessments were administered on paper only. The majority of students tested during the spring 2015 window when all grades and content areas were administered online and on paper.

Six states, the Bureau of Indian Education, and District of Columbia participated in the second administration in school year 2015–2016. Not all participating states had students testing in all grades. In the fall 2015 window, ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2016 window when all grades and content areas were administered online and on paper.

Five states, the Bureau of Indian Education, the Department of Defense Education Activity, and District of Columbia participated in the third administration in school year 2016–2017. Not all participating states had students testing in all grades. In the fall 2016 window, ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2017 window when all grades and content areas were administered online and on paper.

Four states, the Bureau of Indian Education, the Department of Defense Education Activity, and District of Columbia participated in the fourth administration in school year 2017–2018. Not all participating states had students testing in all grades. In the fall 2017 window, ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2016 window when all grades and content areas were administered online and on paper.

The purpose of this technical report is to describe the fourth operational administration of the PARCC summative assessments in the 2017–2018 academic year, including test form construction, test administration, item scoring, test taker characteristics, classical item analysis results, reliability results, evidence of validity, item response theory (IRT) calibrations and scaling, performance level setting procedure, growth measures, and quality control procedures.

1.2 Purpose of the Operational Tests

The PARCC assessments are designed to achieve several purposes. First, the tests are intended to provide evidence to determine whether students are on track for college- and career-readiness. Second, the tests are structured to access the full range of CCSS and measure the total breadth of student performance. Finally, the tests are designed to provide data to help inform classroom instruction, student interventions, and professional development.

1.3 Composition of Operational Tests

Each operational test form was constructed to reflect the full test blueprint in terms of content, standards measured, and item types. Sets of common items, included to provide data to support horizontal linking across test forms within a grade and content area, were proportionally representative of the operational test blueprint.

The current PARCC assessments were administered in either computer-based (CBT) or paper-based (PBT) format. ELA/L assessments focused on writing effectively when analyzing text. Mathematics assessments focused on applying skills and concepts, and understanding multi-step problems that

require abstract reasoning and modeling real-world problems, precision, perseverance, and strategic use of tools. In both content areas, students also demonstrated their acquired skills and knowledge by answering selected-response items and fill-in-the-blank questions.

PARCC is a mixed-format test. The ELA/L assessment has a total of six to nine literary and informational texts; each passage set has four to eight brief comprehension and vocabulary questions. ELA/L constructed-response items include three types of tasks: literary analysis, narrative writing, and research simulation. For each task, students are instructed to read one or more texts, answer several brief questions, and then write an essay based on the material they read. The mathematics assessment contains tasks that measure a combination of conceptual understanding, applications, skills, and procedures. Mathematics constructed-response items consist of tasks designed to assess a student's ability to use mathematics to solve real-life problems. Some of the tasks require students to describe how they solved a problem, while other tasks measure conceptual understanding and ability to apply concepts by means of selected-response or technology-enhanced items. In addition, students are required to demonstrate their skills and knowledge by answering innovative selected-response and short-answer questions that measure concepts and skills.

Each assessment consisted of multiple units, and additionally, one of the mathematics units was split into two sections: a non-calculator section and a calculator section.

1.4 Intended Population

The PARCC tests are intended for students taking ELA/L and/or mathematics in grades 3 through 11, as well as students taking high school mathematics (i.e., Algebra I, Geometry, Algebra II, and Integrated Mathematics I–III). For these students, the PARCC tests measured whether students were meeting state academic standards and mastering the knowledge and skills needed to progress in their K–12 education and beyond.

1.5 Groups and Organizations Involved with PARCC

New Meridian is a nonprofit organization that assumes the responsibility for management of the PARCC consortium, as well as item development and forms construction of PARCC assessments.

Committees of educators, state education agency staff, and national experts lead the work of the PARCC consortium. These committees include:

- the PARCC consortium Governing Board that makes major policy and operational decisions;
- the Technical Advisory Committee that helps ensure all assessments will provide reliable results to inform valid instructional and accountability decisions;
- the State Lead Council that coordinates all aspects of development of the PARCC assessment system and serves as the conduit to the Technical Advisory Committee and the Governing Board; and
- ELA/L, Mathematics, and Accessibility and Accommodation Features operational working groups.

Test and item development activities were conducted by Pearson under the guidance and oversight of New Meridian.

Pearson served as the primary contractor for the PARCC operational administration and was responsible for producing all testing materials, packaging and distribution, receiving and scanning of materials, and scoring, as well as program management and customer service.

Pearson Psychometrics was responsible for all psychometric analyses of the PARCC operational test data. This included classical item analyses, differential item functioning (DIF) analyses, item calibrations based on item response theory (IRT), scaling, and development of all conversion tables.

Human Resources Research Organization (HumRRO) served as a subcontractor and was responsible for replicating item calibrations based on item response theory (IRT), scaling, and development of all conversion tables.

Measured Progress (MP) served as a subcontractor to conduct external evaluations; they were responsible for reviewing and comparing the psychometric IRT calibrations performed by Pearson, which were replicated by HumRRO. MP also provided comparisons of results obtained independently from Pearson and from HumRRO for conversion tables, summative and claim scale scores, performance level classifications, and subclaim performance level classifications.

1.6 Overview of the Technical Report

This report begins by providing explanations of the test form construction process, test administration, and scoring of the test items. Subsequent sections of the report present descriptions of test taker characteristics, results of classical item analyses, results of reliability analyses, evidence of validity, item response theory (IRT) calibrations and scaling, performance level setting procedure, measures of student growth, and quality control procedures.

The technical report contains the following sections:

- Section 2 – Test Development

  This section describes the PARCC test design and the procedures followed during the development of operational test forms.

- Section 3 – Test Administration

  This section presents the operational administration schedule, information regarding test security and confidentiality, accessibility features and accommodations, and testing irregularities and security breaches.

- Section 4 – Scoring of the Items

  The key-based and rule-based processes for machine-scored items, as well as the training and monitoring processes for human-scored items, are provided in this section.

- Section 5 – Test Taker Characteristics

  This section describes the composition of test forms, rules for inclusion of students in analyses, distributions of test takers by grade, mode, and gender, and distributions of demographic variables of interest.

- Section 6 – Classical Item Analysis

  The classical item-level statistics calculated for the operational test data, the flagging criteria used to identify items that performed differently than expected, and the results of these analyses are presented in this section.

- Section 7 – Differential Item Functioning

  In this section, the methods for conducting differential item functioning analyses as well as corresponding flagging criteria are described. This is followed by definitions of the comparison groups and subsequent results for the comparison groups.

- Section 8 – Reliability

  The results of internal consistency reliability analyses and corresponding standard errors of measurement, for each grade, content area, and mode (CBT or PBT) for all test takers, and for subgroups of interest, is provided in this section. This is followed by reliability results for subscores and reliability of classification (i.e., decision accuracy and decision consistency). Finally, expectations and results for inter-rater agreement for handscored items are summarized.

- Section 9 – Validity

  Validity evidence based on analyses of the internal structure of the tests is provided in this section. Correlations between subscores are reported by grade, content area, and mode (CBT or PBT) for all test takers.

- Section 10 – IRT Calibration and Scaling

  This section presents the information related to the calibration and scaling of item response data including: data preparation, the calibration process, model fit evaluation, and items excluded from score reporting. In addition, the scaling process is described and evaluated.

- Section 11 – Performance Level Setting (PLS) Procedure and Results

  Performance levels and policy definitions, as well as the processes followed to establish performance level thresholds, are described in this section.

- Section 12 – Scale Scores

  This section provides an overview of the claims and subclaims, describes the development of the reporting scales and conversion tables, and presents scale score distributions. Finally, information regarding the interpretation of claim scores and subclaim scores is presented.

- Section 13 – Student Growth Measures

  This section provides details on student growth percentiles (SGP). Information about the model, model fit, and SGP averages at the consortium level for all test takers, and for subgroups of interest, are provided in this section.

- Section 14 – Quality Control Procedures

  All aspects of quality control are presented in this section. These activities range from quality assurance of item banking, test form construction, and all testing materials to quality control of scanning, image editing, and scoring. This is followed by a detailed description of the steps taken to ensure that all psychometric analyses were of the highest quality.

- References

- Appendices

  To facilitate utility, tables in the appendices are numbered sequentially according to the section represented by the tables. For example, the first appendix table for Section 5 is numbered A.5.1, the second appendix table for Section 5 is numbered A.5.2, and so on.

- Addendum

  The addendum presents the results of analyses for the fall operational administration. These results are reported separately from the spring results because fall testing involved a nonrepresentative subset of students testing only ELA/L grades 9, 10, and 11, as well as Algebra I, Geometry, and Algebra II.

  To organize the addendum, tables are numbered sequentially according to the section represented by the tables. For example, the first addendum table for Section 5 is numbered ADD.5.1, the second addendum table for Section 5 is numbered ADD.5.2, and so on.

1.7 Glossary of Abbreviations

Table 1.3 Glossary of PARCC Abbreviations and Acronyms

Abbreviation/Acronym	Definition
1PL/PC	One-parameter/Partial Credit Model
2PL/GPC	Two-parameter Logistic/Generalized Partial Credit Model
3PL/GPC	Three-parameter Logistic/Generalized Partial Credit Model
AAF	Accessibility, Accommodations, and Fairness
ABBI	Assessment Banking for Building and Interoperability
AERA	American Educational Research Association
AIS	Average Item Score
AIQ	Assessment and Information Quality
APA	American Psychological Association
ASC	Additional and Supporting Content (Mathematics)
ASL	American Sign Language
ATA	Automatic Test Assembler
CBT	Computer-Based Test
CCSS	Common Core State Standards
CDQ	Customer Data Quality
CSEM	Conditional Standard Error of Measurement
DIF	Differential Item Functioning
DPL	Digital Production Line
DPP	Digital Pre-press
EBSS	Evidence-based Standard Setting
ELA/L	English Language Arts/Literacy
EL	English Learners
EOC	End-of-Course
EOY	End-of-Year
ePEN2	Electronic Performance Evaluation Network second generation
ESEA	Elementary and Secondary Education Act
FRL	Free or Reduced-price Lunch
FS	Full Summative
FT	Field Test
IA	Item Analysis
ICC	Item Characteristic Curve
IDEA	Individuals with Disabilities Education Act
IEP	Individualized Education Program
INF	Information Curve
IRA	Inter-rater Agreement
IRF	Item Response File
IRT	Item Response Theory
ISR	Individual Student Report
K–12	Kindergarten to Grade 12
LEA	Local Education Agency
LID	Local Item Dependence
MAD	Mean Absolute Difference
MC	Major Content (Mathematics)
MH	Mantel-Haenszel
MP	Measured Progress

Abbreviation/Acronym	Definition
MP	Modeling Practice (Mathematics)
MR	Mathematical Reasoning
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind
NCME	National Council on Measurement in Education
NSLP	National School Lunch Program
OE responses	Open-ended responses
OMR	Optical Mark Reading
OWG	Operational Working Group
PARCC	Partnership for Assessment of Readiness for College and Careers
PBA	Performance-Based Assessment
PBT	Paper-Based Test
PCR	Prose Constructed Response (ELA/L)
PEJ	Postsecondary Educators' Judgment
PLD	Performance Level Descriptor
PLS	Performance Level Setting
PV	Product Validation
QA	Quality Assurance
RD	Reading (ELA/L)
RI	Reading Information (ELA/L)
RL	Reading Literature (ELA/L)
RMSD	Root Mean Square Difference
RV	Reading Vocabulary (ELA/L)
RST	Raw-score-to-theta
SD	Standard Deviation
SDF	Student Data File
SE	Standard Error
SEJ	Standard Error of Judgment
SEM	Standard Error of Measurement
SIRB	Scored Item Response Block
SMD	Standardized Mean Difference
SSMC	Single Select Multiple Choice
SWD	Students with Disabilities
TCC	Test Characteristic Curve
TTS	Text to Speech
UIN	Unique Item Number
WE	Writing Written Expression (ELA/L)
WKL	Writing Knowledge Language and Conventions (ELA/L)
WLS	Weighted Least Squares
WR	Writing (ELA/L)
WRMSD	Weighted Root Mean Square Difference

Section 2: Test Development

2.1 Overview of the PARCC Assessment, Claims, and Design

Aligned to the Common Core State Standards (CCSS) as articulated in the PARCC Model Content Frameworks, the PARCC assessments are designed to determine whether students are college- and career-ready or on track, assess the full range of the CCSS, measure the full range of student performance, and provide data to help inform instruction, interventions, and professional development. Test development is an ongoing process involving educators, researchers, psychometricians, subject matter professionals, and assessment experts who participate in the development of the PARCC test design and its underlying foundational documents; develop and review passages and items used to build the PARCC assessments; monitor the program for quality, accessibility, and fairness for all students; and construct, review, and score the assessments

The PARCC summative assessments include both English language arts/literacy (ELA/L) and mathematics assessments in grades 3 through 8 and high school. The high school mathematics tests include traditional mathematics and integrated mathematics course pathways. Tests contain selected response, brief and extended constructed response, technology-enabled and technology-enhanced items (TEI), as well as performance tasks. Technology-enabled items are single-response or constructed-response items that involve some type of digital stimulus or open-ended response box with which the students engage in answering questions. Technology-enhanced items involve specialized student interactions for collecting performance data. In other words, the act of performing the task is the way in which data is collected. Students may be asked, among other interactions, to categorize information, organize or classify data, order a series of events, plot data, generate equations, highlight text, or fill in a blank. One example of a TEI is an interaction in which students are asked to drag response options onto a Venn diagram to show the relationship among ideas.

The PARCC assessments offer a wide range of accessibility features for all students and accommodations for students with disabilities (e.g., screen reader, assistive technology, braille, large print [LP], text-to-speech [TTS], and American Sign Language [ASL] video versions of the test, as well as response accommodations that allow students to respond to test items using different formats). For English learners who are native Spanish speakers, PARCC offers the mathematics assessments in Spanish, and both LP and TTS versions of the test in Spanish (refer to the PARCC Accessibility Features and Accommodations Manual for in-depth information).

2.1.1 English Language Arts/Literacy (ELA/L) Assessments—Claims and Subclaims

The ELA/L summative assessment at each grade level consists of three task types: literary analysis, research simulation, and narrative writing. For each performance-based task, students are asked to read or view one or more texts, answer comprehension and vocabulary questions, and write an extended response that requires them to draw evidence from the text(s). The summative assessment also contains literary and informational reading passages with comprehension and vocabulary questions.

The claim structure, grounded in the CCSS, undergirds the design and development of the ELA/L summative assessments.

Master Claim. The master claim is the overall performance goal for the PARCC ELA/L Assessment System—students must demonstrate that they are college- and career-ready or on track to readiness as demonstrated through reading and comprehending of grade-level texts of appropriate complexity and writing effectively when using and/or analyzing sources.

Major Claims: 1) reading and comprehending a range of sufficiently complex texts independently, and 2) writing effectively when using and/or analyzing sources.

Subclaims: The subclaims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence tables for reading and writing (refer to PARCC test specifications documents). The claims and evidences are grouped into the following categories:

1. Vocabulary Interpretation and Use
2. Reading Literature
3. Reading Informational Text
4. Written Expression
5. Knowledge of Language and Conventions

2.1.2 Mathematics Assessments—Claims and Subclaims

The summative mathematics assessment at each grade level includes both short- and extended-response questions focused on applying skills and concepts to solve problems that require demonstration of the mathematical practices from the CCSS with a focus on modeling and reasoning with precision. The assessments also include performance-based short-answer questions focused on conceptual understanding, procedural skills, and application.

The claim structure, grounded in the CCSS, undergirds the design and development of the summative assessments.

Master Claim. The degree to which a student is college- or career-ready or on track to being ready in mathematics. The student solves grade-level/course-level problems aligned to the Standards for Mathematical Content with connections to the Standards for Mathematical Practice.

Subclaims: The subclaims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence statement tables for mathematics (refer to PARCC test specifications documents). The claims and evidence are grouped into the following categories.

Subclaim A: Major Content with Connections to Practices

Subclaim B: Additional and Supporting Content with Connections to Practices

Subclaim C: Highlighted Practices with Connections to Content: Expressing mathematical reasoning by constructing viable arguments, critiquing the reasoning of others, and/or attending to precision when making mathematical statements

Subclaim D: Highlighted Practice with Connections to Content: Modeling/Application by solving real-world problems by applying knowledge and skills articulated in the standards

2.2 Test Development Activities

Test development activities began with the standards and model content frameworks. From these, PARCC, in collaboration with more than 2,000 educators, researchers, and psychometricians, has developed the PARCC test specifications documents that guide the development of test items and the composition of the tests. These documents include the College- and Career-Ready Determinations and Performance-Level Descriptions, Claim Structure, Evidence Statement Tables, Blueprints, Informational Guides, Passage Selection Guidelines, Mathematics Sequencing Guidelines, Task Generation Models, Fairness and Sensitivity Guidelines, Text Selection Guidelines, and the Style Guide. Refer to the PARCC website for further information about these documents.

2.2.1 Item Development Process

PARCC test and item development activities were conducted by Pearson under the guidance and oversight of PARCC leadership, including the PARCC Governing Board, the K–12 state leads, the Higher Education Leadership Team, the Technical Advisory Committee, the Operational Working Group members from each of the member states, the PARCC Text and Content Item Review Committees, and staff members from New Meridian Corporation, the project management partner for the PARCC Consortium.

Developing high quality assessment content with authentic stimuli for computer-based tests (CBT) and paper-based tests (PBT) measuring rigorous standards is a complex process involving the services of many experts including assessment designers, psychometricians, managers, trainers, content providers, content experts, editors, artists, programmers, technicians, human scorers, advisors, and members of the PARCC Operational Working Groups (OWGs).

Bank Analysis and Item Development Plan

The PARCC summative item bank houses passages and items at each assessed grade level and subject. The bank supports the administration of the assessments, along with item release and practice tests. Items are developed and field tested annually. Prior to the annual item development cycle, the item development teams, in conjunction with members of the OWGs for ELA/L and mathematics, evaluated the strengths of the bank and considered the needs for future tests to establish an item development plan.

Text Selection for ELA/L

Using the PARCC Passage Selection Guidelines, English language arts subject matter experts were trained to search for appropriate passages to support an annual pool of passages for consideration. Guided by the PARCC test specifications documents, Pearson recruited, trained, and managed the contracted subject matter experts to deliver the number of texts specified in the annual asset development plan. The Passage Selection Guidelines provided a text complexity framework and guidance on selecting a variety of text types and passages that allow for a range of standards/evidences to be demonstrated to meet the PARCC claims. PARCC ELA/L tests are based on authentic texts, including multi-media stimulus. Authentic texts are grade-appropriate texts that are not developed for the purposes of the assessment or to achieve a particular readability metric, but reflect the original language of the authors. Pearson content experts reviewed the passages for adherence to the PARCC Passage Selection Guidelines to meet the annual asset development plan described above in the number and distribution of genres and topics prior to review and consideration by the Text Review Committee. ELA/L item development was not conducted until after texts were approved by the Text Review Committee.

Item Development

Guided by the PARCC foundational documents, Pearson recruited and trained the item writers and managed the item writing to develop the number of items specified in the annual asset development plan. Prior to further committee reviews, the assessment teams at Pearson reviewed the items for content accuracy, alignment to the standards, range of difficulty, adherence to universal design principles (which maximize the participation of the widest possible range of students), bias and sensitivity, and copy edit to enable the accurate measurement of the PARCC standards.

2.2.2 Item and Text Review Committees

Members of the PARCC OWGs for ELA/L and mathematics, state-level experts, local educators, post-secondary faculty, and community members from the PARCC states conducted rigorous reviews of every item and passage being developed for the PARCC assessment system to ensure all test items are of the highest quality, aligned to the standards, and fair for all student populations. All PARCC reviewers were nominated by their state education agency. The purpose of the educator reviews was to provide feedback to Pearson and PARCC on the quality, accuracy, alignment, and appropriateness of the test passages and items developed annually for the summative PARCC assessments. The meetings were conducted either in person or virtually and included large group training on the expectations and processes of each meeting, followed by breakout meetings of grade/subject working committees where additional training was provided.

Text Review

The Text Review is a review and approval by the Text Review Committee of the texts eligible for item development. Participants reviewed and provided feedback to Pearson and PARCC about the grade-level appropriateness, content, and potential bias concerns, and reached consensus about which texts would move forward for development. The Text Review Committee was made up of members of both Content Item Review and Bias and Sensitivity Review Committees.

Content Item Review
During Content Item Review, committees reviewed and edited test items for adherence to the PARCC foundational documents, basic universal design principles, PARCC Accessibility Guidelines, associated item metadata, and the PARCC Style Guide. Committees accessed the item content within the Pearson Assessment Banking for Building and Interoperability (ABBI) system that previews how the passages and items will be displayed in an operational online environment. Committees also verified that the appropriate scoring rule had been applied to each item. The Content Item Review Committees were made up of OWG members and educators nominated by PARCC member states.

Bias and Sensitivity Review
Educators and community members make up the committee that reviews items and tasks to confirm that there are no bias or sensitivity issues that would interfere with a student's ability to achieve his or her best performance. The committee reviewed items and tasks to evaluate adherence to the Fairness and Sensitivity Guidelines, and to ensure that items and tasks do not unfairly advantage or disadvantage one student or group of students over another. Bias and Sensitivity Committee members made edits and modifications to items and passages to eliminate sources of bias and improve accessibility for all students.

Editorial Review
The PARCC Editorial Review Committee consists of editors who reviewed up to 10 percent of the items and tasks. The committee reviewed the items for grammar, punctuation, clarity, and adherence to the PARCC Style Guide.

Data Review
Following the field test, educator and bias committee members met to evaluate test items and associated performance data with regard to appropriateness, level of difficulty, and potential gender, ethnic, or other bias, then recommended acceptance or rejection of each field-test item for inclusion on an operational assessment. The Data Review Committee also made recommendations that items be revised and re-field tested. Items that were approved by the committee are eligible for use on operational summative assessments.

2.2.3 Operational Test Construction

Under the guidance in the operational test form creation specifications, Pearson constructed the operational forms to adhere to the test blueprints and the assessment goals outlined in the form creation specifications. These goals were:

- test forms designed to measure well across the full range of student ability;
- scores that are comparable among forms and across test administrations;
- scales that support classification of students into performance levels;
- maximization of the number of parallel forms;
- minimization of overexposure of items; and

- adherence to standards for validity, reliability, and fairness (*Standards for Educational and Psychological Testing, 2014)*.

Each content-area and grade-level assessment was based on a specific test blueprint that guided how each test was built. Test blueprints determined the range and distribution of content, and the distribution of points across the PARCC subclaims and task types.

Multiple operational forms were constructed for each grade/subject. These forms were designed to facilitate psychometric equating through a common item linking strategy and to be constructed as "parallel" as possible from a content and test-taking experience. Evaluation criteria for parallelism included adherence to blueprint; sequencing of content across the forms; statistical averages and distributions for difficulty (e.g., p-value) and discrimination (e.g., polyserial correlation); item type and cognitive complexity; and passage characteristics for ELA/L including genre, topics, word count, and text complexity.

Core forms are the operational test forms consisting of only those items that will count toward a student's score. Core forms are constructed to meet the blueprint and psychometric properties outlined in the test construction specifications. PARCC creates multiple core forms for a given assessment to enhance test security and to support opportunity for item release. The number of core operational forms per grade/subject is provided in Table 2.1. Additionally, appropriate forms were identified as accessibility and accommodated forms, and the core forms for all mathematics assessments included embedded field-test items. A sample of students were administered ELA/L core forms that included a unit of field-test items. Accessibility and accommodated forms and embedded field testing are described later in this section.

Table 2.1 Number of Core Operational Forms per Grade/Subject and Mode for ELA/L and Mathematics*

Grade/Subject	ELA/L		Mathematics	
	CBT	PBT	CBT	PBT
Grade 3	2	1	2	1
Grade 4	2	1	2	1
Grade 5	2	1	2	1
Grade 6	2	1	2	1
Grade 7	2	1	2	1
Grade 8	2	1	2	1
Grade 9	2	1		
Grade 10	2	1		
Grade 11	2	1		
Algebra I			2	1
Geometry			2	1
Algebra II			2	1
Integrated Mathematics I			1	1
Integrated Mathematics II			1	1
Integrated Mathematics III			1	1

*Grades 3–11 ELA/L and Integrated Mathematics I have two operational accommodated forms and mathematics grades 3–8 and the high school traditional assessments have three accommodated forms. The forms are accommodated to support Braille, large print, human reader/human signers, assistive technology, text-to-speech, closed captioning, and Spanish. Human reader/human signers and Spanish are provided for mathematics assessments only. Closed captioning is provided for ELA/L assessments only.

Test Construction Activities

After the data review meetings and prior to the test construction meetings, Pearson assessment specialists constructed initial versions of all the core forms, as depicted in Table 2.1. The construction model varied slightly between the two subject areas.

For ELA/L, content specialists constructed the initial core forms shown in Table 2.1 based on the support documents and specific processes to achieve fair parallel forms. The following steps were used to construct the operational core ELA/L form inputs taken to the Test Construction Committee for review.

1. constructed the online forms to match blueprint and test construction specifications
2. constructed the paper forms to match the blueprint and test construction specifications
3. constructed accommodated and accessibility forms to match the blueprint, test construction specifications, and AAF constraints

The ELA/L construction process included iterative steps between content specialists and psychometricians. Custom PARCC test construction reports generated by the Pearson psychometric team provided information on adherence to blueprint and statistical averages/distributions of item difficulty and discrimination describing the forms and allowing comparison of the forms. These reports facilitated content changes to better achieve the test construction goals. Equating across post-equated

operational forms within an administration was accomplished by repeating core items across forms. Linking across administrations for post-equated operational forms was accomplished by including prior operational items on the current operational test forms.

For mathematics, Pearson employed the use of an automatic test assembler (ATA) to select the items for the initial forms. Based on the blueprints and other test construction goals and specifications, the ATA was able to create sets of items best satisfying the statistical parameters outlined in the test construction specifications; however, the ATA was unable to sequence the items as required by the PARCC Mathematics Sequencing Guidelines. Sequencing was conducted by assessment specialists who ordered the items according to the sequencing guidelines.

Similar to the ELA/L construction process, mathematics included iterative steps between assessment specialists and psychometricians. Custom PARCC test construction reports generated by the Pearson psychometric team provided information on adherence to blueprint and statistical averages/distributions of item difficulty and discrimination allowing a comparison of the forms and facilitating content changes to better achieve the test construction goals. Since the mathematics forms were generated by the ATA, psychometricians could also generate the reports prior to content experts reviewing the forms.

Pearson assessment specialists identified forms for each grade/subject suitable for use as the accommodated forms. Pearson psychometrics reviewed the psychometric properties of each of the accommodated forms with respect to the required criteria. The content of these forms was also reviewed by Pearson accessibility specialists allowing for content changes prior to the Test Construction Committee meetings.

These test construction activities provided significant inputs to commence the meetings including:

- the proposed items for the initial operational core forms and the accommodated forms described above
- reports describing each form and comparing parallel forms
- recommended accommodated forms

Test Construction Meeting to Review Test Construction Inputs

Members of the Content Item Review Committees and the Accessibility, Accommodations, and Fairness (AAF) OWG participated in the building of operational core forms that met PARCC summative assessments requirements. In that process, they met in an in-person meeting to review and make recommendations for changes so that test forms conformed to both the content and psychometric requirements of the assessment.

Accommodated Form Review Process

In addition to participating in many of the development activities including the Text Review and the Bias and Sensitivity Review meetings, the AAF OWG reviewed the proposed accommodated forms at the Test Construction Committee meeting for accessibility to make sure that the content can be accommodated for students with disabilities and English learners without changing the underlying measured construct.

Forms were identified to support the following accommodations:

Accommodated Base 1

- Spanish paper (also serves Spanish LP, Spanish human reader paper)
- Spanish human reader/human signer online
- base accommodated paper (serves Braille, LP, human reader paper)
- human reader/human signer online
- assistive technology screen reader
- assistive technology non-screen reader
- ASL

Accommodated Base 2

- closed captioning
- text-to-speech first form
- Spanish online
- Spanish text-to-speech

Accommodated Base 3 (mathematics only)

- text-to-speech second form

Spanish is mathematics only. Closed captioning is ELA/L only.

At the conclusion of the meetings, all test forms were constructed to meet test blueprints and PARCC requirements, and if necessary, reflect the operational linking design. Each test form reflected the test blueprint in terms of content, item types, and test length, as well as *expected* difficulty and performance along the ability continuum. ELA/L linking sets were proportionally representative of the operational test blueprint. The operational core forms, linking set forms, and field-test forms were reviewed by PARCC Forms Review Committees and approved prior to the test administration.

Spanish-Language Assessments for Mathematics

For English learners, PARCC offers the mathematics assessments in Spanish, as well as large print and text-to-speech (TTS) versions of the test in Spanish. Once the operational form was approved, the form was sent to Pearson's subcontractor, Teneo, for transadaption of the items. Transadaption differs from translation in that it takes into consideration the grade-level appropriateness of the words, as well as the linguistic and cultural differences that exist between speakers of two different languages. Accounting for these differences allows the item to measure the achievement of Spanish language speakers in the same way that the original version of the item does for native speakers of English. The PARCC Spanish Glossary provided guidance to the translator conducting the transadaption in grade-level and culturally appropriate ways of transadapting the items. For the Spanish language TTS form, the alternate text (used for description and/or text in art and graphics) was transadapted from the alternate text for the English language version of the TTS form. Phonetic mark-up, which guides how the TTS reader pronounces content-specific words and phrases, was also applied in this process.

In addition to the expert review of potential content for all accommodated forms conducted by the AAF OWG with assistance from content experts at the test construction meetings, the transadapted forms

underwent additional quality checks: a Pearson Spanish copy edit services review and approval, and an AAF OWG review and approval.

2.2.4 Linking Design of the Operational Test

To support the goal of score comparability within and across administrations and years, PARCC implemented a hybrid approach that incorporated the strengths of common item linking and randomly equivalent groups. For ELA/L, the use of repeated operational core items was leveraged for common item linking. In addition, all forms were available throughout the operational administration, with spiraling at the student level, leveraged to support linking through randomly equivalent groups.

The ELA/L operational test forms involved various types of linking: horizontal linking, testing mode linking, and across-administration linking. Horizontal linking consisted of linking items, or common items, included in both forms in a single administration. Testing mode linking consisted of common items placed in computer-based forms and paper-based forms within an administration to support the development of scores on the same reporting scale. Across-administration linking, or year-to-year linking, consisted of common items included in two different administrations. The placement of linking items across forms or administrations supports the development of comparable scores.

Linking item sets can be internal or external linking sets. Internal linking sets consist of common items in operational positions such that the items contribute to the students' scores. External linking sets consist of common items in positions resulting in the items not contributing to students' scores. The current linking designs included internal linking sets.

2.2.5 Field Test Data Collection Overview

Field-test items were embedded in the spring operational forms to collect data for psychometric analysis necessary to support the assessment system for future administrations. Field-test administration entailed paper and computer administration modes, with computer administration as the dominant mode. The ELA/L unit of field-test items were administered to a sample of students.

Field-test sets were constructed to balance the expected cognitive load and difficulty across forms, reflected in the number of points, distribution of task types, and balance of passages for ELA/L. Forms for each content area were spiraled at the student level. The data collection design entailed two conditions. Condition 1, which comprised the mathematics assessment, was an embedded census field-test model in which all students taking the summative assessment participated in the field test.

Under Condition 2, which comprised the ELA/L assessment, PARCC sampled approximately one-third of the schools across the consortium states. Students in the sampled schools took forms containing ELA/L embedded field-test tasks. Schools were selected so that the sample for each ELA/L assessment was representative of the general PARCC testing populations in terms of achievement (i.e., average scale score and percentage of students at Level 4 and Level 5 in the previous year) and demographics (i.e., ethnicity composition, percentage of economically disadvantaged, English learners, and students with disabilities). A three-year sampling plan was created such that if a given school was part of the ELA/L

field test one year (e.g., spring 2016), it would not be required to participate in the field test for the subsequent two years (e.g., spring 2017 and spring 2018).

Section 3: Test Administration

3.1 Testing Windows

The 2017–2018 operational administration of PARCC assessments included a fall/winter block administration beginning in the fall of 2017 as well as a spring administration in the spring of 2018. Each PARCC assessment consisted of multiple units, and additionally, one of the mathematics units for grade 7 and high school course assessments was split into two sections: a non-calculator section and a calculator section.

Table 3.1 PARCC Fall/Winter Block 2017 and Spring 2018 Testing Windows

Fall/Winter Block 2017	November 13, 2017– January 26, 2018
Spring 2018	March 12, 2018– June 8, 2018

3.2 Test Security and Administration Policies

The administration of any PARCC assessment is a secure testing event. Maintaining the security of test materials before, during, and after the test administration is crucial to obtaining valid and reliable results. School Test Coordinators are responsible for ensuring that all personnel with authorized access to secure materials are trained in and subsequently act in accordance with all security requirements.

School Test Coordinators must implement chain-of-custody requirements for specified materials. School Test Coordinators are responsible for distributing materials to Test Administrators, collecting materials from Test Administrators, returning secure test materials, and securely destroying certain specified materials after testing.

The administration of the PARCC assessment includes both secure and non-secure materials, and these materials are further delineated by whether they are “scorable” or “nonscorable,” depending on whether the assessments were administered via paper/pencil (i.e., paper-based assessments) or online (i.e., computer-based assessments). For the 2017–2018 paper-based administration, students used paper-based answer documents (except in grade 3 where students responded directly into test booklets). About 96 percent of the PARCC assessments administered during the 2017–2018 administration were online assessments, and about 4 percent were paper-based assessments.

Secure vs. Non-Secure Materials

PARCC defines secure materials as those that must be closely monitored and tracked to prevent unauthorized access to or prohibited use or distribution of secure content such as test items, reading

passages, student work, etc. For paper-based tests, secure materials include both used and unused test booklets and used scratch paper, while for computer-based tests, secure materials include student testing tickets, secure administration scripts (e.g., mathematics read-aloud), and used scratch paper. PARCC defines non-secure materials as any authorized testing materials that do not include secure content (e.g., test items or student work). These include test administration manuals, unused scratch paper, and mathematics reference sheets that have not been written upon, etc.

Scorable vs. Nonscorable Materials

Paper-based assessments have both scorable and nonscorable materials while computer-based assessments have only nonscorable materials. Scorable materials for paper-based assessments consist of used (includes student work) test booklets (grade 3) and answer documents (grades 4 and above) only. Scorable materials must be returned to the vendor to be scored. All other materials for paper-based testing, such as blank (i.e., unused) test booklets, test administration manuals, scratch paper, mathematics reference sheets, etc., are deemed nonscorable. For computer-based tests, there are no scorable materials as student work is submitted electronically for scoring; thus there are limited physical materials to return (e.g., secure administration scripts for certain accommodations).

Students taking the computer-based test may not have access to secure test materials before testing, including printed student testing tickets. Printed mathematics reference sheets (if applicable) and scratch paper must be new and unmarked.

Students taking the paper-based test may not have access to scorable or nonscorable secure test content before or after testing. Scorable secure materials that are to be provided by Test Administrators to students include test booklets (grade 3) or answer documents (grades 4 through high school). Nonscorable secure materials that are distributed by Test Administrators to paper-based testing students include large print test booklets, braille test booklets, scratch paper (paper used by students to take notes and work through items), and printed mathematics reference sheets (grades 5 through 8 and high school).

School Test Coordinators are required to maintain a tracking log to account for collection and destruction of test materials, including mathematics reference sheets and scratch paper written on by students. As part of the test administration policy, schools are required to maintain the Chain-of-Custody Form or tracking log of secure materials for at least three years unless otherwise directed by state policy. Copies of the Chain-of-Custody Form for paper-based testing are included in each Local Education Agency (LEA) or school's test materials shipment.

Test Administrators are not to have extended access to test materials before or after administration (except for certain accessibility or accommodations purposes). Test Administrators must document the receipt and return of all secure test materials (used and unused) to the School Test Coordinator immediately after testing.

All PARCC test security and administration policies are found in the *PARCC Test Coordinator Manual* and the *PARCC Test Administrator Manuals*. State security and administration policies may exceed that of the PARCC policies. State-specific policies are included in Appendix C of the *Test Coordinator Manual*.

3.3 Accessibility Features and Accommodations

3.3.1 Participation Guidelines for PARCC Assessments

All students, including students with disabilities and English learners, are required to participate in statewide assessments and have their assessment results be part of the state's accountability systems, with narrow exceptions for English learners in their first year in a U.S. school, and certain students with disabilities who have been identified by the Individualized Education Program (IEP) team to take their state's alternate assessment. All eligible students will participate in the PARCC ELA/L and mathematics assessments. Federal laws governing student participation in statewide assessments include the No Child Left Behind Act of 2001 (NCLB), the Individuals with Disabilities Education Act of 2004 (IDEA), Section 504 of the Rehabilitation Act of 1973 (reauthorized in 2008), and the Elementary and Secondary Education Act (ESEA) of 1965, as amended. All students can receive accessibility features on PARCC assessments.

Four distinct groups of students may receive accommodations on PARCC assessments:

1. **students with disabilities who have an Individualized Education Program (IEP)**;
2. **students with a Section 504 plan** who have a physical or mental impairment that substantially limits one or more major life activities, have a record of such an impairment, or are regarded as having such an impairment, but who do not qualify for special education services;
3. **students who are English learners**; and
4. **students who are English learners with disabilities who have an IEP or 504 plan.** These students are eligible for accommodations intended for both students with disabilities and English learners.

Testing accommodations for students with disabilities or students who are English learners (EL) must be documented according to the guidelines and requirements outlined in the *PARCC Accessibility Features and Accommodations Manual*.

3.3.2 PARCC Accessibility System

Through a combination of universal design principles and accessibility features, PARCC has designed an inclusive assessment system by considering accessibility from initial design through item development, field testing, and implementation of the assessments for all students, including students with disabilities, English learners, and English learners with disabilities. Accommodations may still be needed for some students with disabilities and English learners to assist in demonstrating what they know and can do. However, the accessibility features available to students should minimize the need for accommodations during testing and ensure the inclusive, accessible, and fair testing of the diverse students being assessed.

3.3.3 What are Accessibility Features?

On the PARCC computer-based assessments, accessibility features are tools or preferences that are either built into the assessment system or provided externally by Test Administrators, and may be used

by any student taking the PARCC assessments (i.e., students with and without disabilities, gifted students, English learners, and English learners with disabilities). Since accessibility features are intended for all students, they are not classified as accommodations. Students should have the opportunity to select and practice using them prior to testing to determine which are appropriate for use on the PARCC assessment. Consideration should be given to the supports a student finds helpful and consistently uses during instruction. Practice tests that include accessibility features are available for teacher and student use throughout the year. Practice tests are available at parcc.pearson.com.

3.3.4 Accommodations for Students with Disabilities and English Learners

It is important to ensure that performance in the classroom and on assessments is influenced minimally, if at all, by a student's disability or linguistic/cultural characteristics that may be unrelated to the content being assessed. For PARCC assessments, accommodations are considered to be adjustments to the testing conditions, test format, or test administration that provide equitable access during assessments for students with disabilities and students who are English learners. In general, the administration of the assessment should not be the first occasion on which an accommodation is introduced to the student. To the extent possible, accommodations should:

- provide equitable access during instruction and assessments;
- mitigate the effects of a student's disability;
- not reduce learning or performance expectations;
- not change the construct being assessed; and
- not compromise the integrity or validity of the assessment.

Accommodations are intended to reduce and/or eliminate the effects of a student's disability and/or English language proficiency level; however, **accommodations should never reduce learning expectations by reducing the scope, complexity, or rigor of an assessment.** Moreover, accommodations provided to a student on the PARCC assessments must be generally consistent with those provided for classroom instruction and classroom assessments. There are some accommodations that may be used for instruction and for formative assessments that are not allowed for the summative assessment because they impact the validity of the assessment results—for example, allowing a student to use a thesaurus or access the Internet during a PARCC assessment. There may be consequences (e.g., excluding a student's test score) for the use of non-allowable accommodations during PARCC assessments. It is important for educators to become familiar with PARCC policies regarding accommodations used for assessments.

To the extent possible, accommodations should adhere to the following principles.

- Accommodations enable students to participate more fully and fairly in instruction and assessments and to demonstrate their knowledge and skills.
- Accommodations should be based upon an individual student's needs rather than on the category of a student's disability, level of English language proficiency alone, level of or access to

grade-level instruction, amount of time spent in a general classroom, current program setting, or availability of staff.

- Accommodations should be based on a documented need in the instruction/assessment setting and should not be provided for the purpose of giving the student an enhancement that could be viewed as an unfair advantage.
- Accommodations for students with disabilities must be described and documented in the student's appropriate plan (i.e., either a 504 plan or an approved IEP), and must be provided if they are listed.
- Accommodations for English learners should be described and documented.
- Students who are English learners with disabilities are eligible to receive accommodations for both students with disabilities and English learners.
- Accommodations should become part of the student's program of daily instruction as soon as possible after completion and approval of the appropriate plan.
- Accommodations should not be introduced for the first time during the testing of a student.
- Accommodations should be monitored for effectiveness.
- Accommodations used for instruction should also be used, if allowable, on local district assessments and state assessments.

In the following scenarios, the school must follow each state's policies and procedures for notifying the state assessment office:

- a student **was provided a test accommodation that was not listed** in his or her IEP/504 plan/documentation for an English learner, or
- a student **was not provided a test accommodation that was listed** in his or her IEP/504 plan/documentation for an English learner.

3.3.5 Unique Accommodations

PARCC provides a comprehensive list of accessibility features and accommodations in the *PARCC Accessibility Features and Accommodations Manual* that are designed to increase access to PARCC assessments and that will result in valid, comparable assessment scores. However, students with disabilities or English learners may require additional accommodations that are not already listed. PARCC states individually review requests for unique accommodations in their respective states and provide a determination as to whether the accommodation would result in a valid score for the student, and if so, would approve the request.

3.3.6 Emergency Accommodations

An emergency accommodation may be appropriate for a student who incurs a temporary disabling condition that interferes with test performance shortly before or during the PARCC assessment window. A student, whether or not they already have an IEP or 504 plan, may require an accommodation as a result of a recently occurring accident or illness. Cases include a student who has a recently fractured limb (e.g., arm, wrist, or shoulder); a student whose only pair of eyeglasses has broken; or a student

returning to school after a serious or prolonged illness or injury. An emergency accommodation should be given only if the accommodation will result in a valid score for the student (i.e., does not change the construct being measured by the test[s]). If the principal (or designee) determines that a student requires an emergency accommodation on the PARCC assessment, an Emergency Accommodation Form must be completed and maintained in the student's assessment file. If required by a PARCC state, the school may need to consult with the state or district assessment office for approval**. The parent must be notified that an emergency accommodation was provided.** If appropriate, the Emergency Accommodation Form may also be submitted to the District Assessment Coordinator to be retained in the student's central office file. Requests for emergency accommodations will be approved after it is determined that use of the accommodation would result in a valid score for the student.

3.3.7 Student Refusal Form

If a student refuses an accommodation listed in his or her IEP, 504 plan, or (if required by the PARCC member state) an English learner plan, the school should document in writing that the student refused the accommodation, and the accommodation must be offered and remain available to the student during testing. This form must be completed and placed in the student's file and a copy must be sent to the parent on the day of refusal. Principals (or designee) should work with Test Administrators to determine who, if any others, should be informed when a student refuses an accommodation documented in an IEP, 504 plan, or (if required by the PARCC member state) English learner plan.

3.4 Testing Irregularities and Security Breaches

Any action that compromises test security or score validity is prohibited. These may be classified as testing irregularities or security breaches. Below are examples of activities that compromise test security or score validity (note that these lists are not exhaustive). It is highly recommended that School Test Coordinators discuss other possible testing irregularities and security breaches with Test Administrators during training.

Examples of test security breaches and irregularities include but are not limited to:

- **Electronic Devices**
    - Using a cell phone or other prohibited handheld electronic device (e.g., smartphone, iPod, smart watch, personal scanner) while secure test materials are still distributed, while students are testing, after a student turns in his or her test materials, or during a break
        - Exception: Test Coordinators, Technology Coordinators, Test Administrators, and Proctors are permitted to use cell phones in the testing environment only in cases of emergencies or when timely administration assistance is needed. LEAs may set additional restrictions on allowable devices as needed.

- **Test Supervision**
  - Coaching students during testing, including giving students verbal or nonverbal cues, hints, suggestions, or paraphrasing or defining any part of the test
  - Engaging in activities (e.g., grading papers, reading a book, newspaper, or magazine) that prevent proper student supervision at all times while secure test materials are still distributed or while students are testing
  - Leaving students unattended for any period of time while secure test materials are still distributed or while students are testing
  - Deviating from testing time procedures
  - Allowing cheating of any kind
  - Providing unauthorized persons with access to secure materials
  - Unlocking a test in PearsonAccess^{next} during non-testing times
  - Failing to provide a student with a documented accommodation or providing a student with an accommodation that is not documented and therefore is not appropriate
  - Allowing students to test before or after the state's test administration window

- **Test Materials**
  - Losing a student test booklet or answer document
  - Losing a student testing ticket
  - Leaving test materials unattended or failing to keep test materials secure at all times
  - Reading or viewing the passages or test items before, during, or after testing
    - Exception: Administration of a human reader/signer accessibility feature for mathematics or accommodation for English language arts/literacy, which requires a Test Administrator to access passages or test items
  - Copying or reproducing (e.g., taking a picture of) any part of the passages or test items or any secure test materials or online test forms
  - Revealing or discussing passages or test items with anyone, including students and school staff, through verbal exchange, email, social media, or any other form of communication
  - Removing secure test materials from the school's campus or removing them from locked storage for any purpose other than administering the test

- **Testing Environment**
  - Allowing unauthorized visitors in the testing environment
  - Failing to follow administration directions exactly as specified in the *Test Administrator Manual*
  - Displaying testing aids in the testing environment (e.g., a bulletin board containing relevant instructional materials) during testing

All instances of security breaches and testing irregularities must be reported to the School Test Coordinator immediately. The Form to Report a Testing Irregularity or Security Breach must be completed within two school days of the incident.

If any situation occurred that could cause any part of the test administration to be compromised, schools should refer to the *PARCC Test Coordinator Manual* for each state's policy and immediately follow those steps. Instructions for the School Test Coordinator or LEA Test Coordinator to report a testing irregularity or security breach is available in the *PARCC Test Coordinator Manual*.

3.5 Data Forensics Analyses

Maintaining the validity of test scores is essential in any high-stakes assessment program, and misconduct represents a serious threat to test score validity. When used appropriately, data forensic analyses can serve as an integral component of a wider test security protocol. The results of these data forensic analyses may be instrumental in identifying potential cases of misconduct for further follow-up and investigation.

In 2016–2017, PARCC conducted the following data forensics analyses on its operational assessments:

- Response Change Analysis
- Aberrant Response Analysis
- Plagiarism Analysis
- Longitudinal Performance Modeling
- Internet and Social Media Monitoring
- Off-Hours Testing Monitoring

An overview of each data forensics analysis method is provided next.

3.5.1 Response Change Analysis

Response change analysis looks at how often student answers are changed, focusing specifically on an excessive number of wrong answers changed to right answers. In traditional paper-based, multiple-choice testing programs, this is sometimes referred to as "erasure analysis"[2]. The rationale for erasure analysis is that a teacher or administrator who is intent on improving classroom performance might be motivated to change student responses after the answer sheets are collected. A clustered number of student answer documents from the same school or classroom with unusually high numbers of answers changed from wrong to right might provide evidence to support follow-up investigation. PARCC's response change analysis extended the traditional erasure method to account for issues specific to computer-based testing as well as the variety of item types on the PARCC assessments, such as partial-credit, multi-part, and multiple-select items.

[2] The term "erasure analysis" is sometimes objected to because it is inferential rather than descriptive. A more descriptive term is "mark discrimination analysis," which recognizes that the scanning approach makes discriminations among the darkness of selected answer choices when multiple responses to a multiple-choice item are detected during answer sheet processing.

3.5.2 Aberrant Response Analysis

Aberrant response pattern detection analysis looks at the unusualness of student responses compared with what would be expected. Most simply, this can be thought of as quantifying the extent to which higher-scoring students miss easy questions and lower-scoring students answer difficult questions correctly. While it would be difficult to draw a definitive inference about a single student flagged as having an aberrant response pattern, a cluster of students with aberrant response patterns within a classroom or school might warrant further investigation.

3.5.3 Plagiarism Analysis

Plagiarism analysis compares the responses given for a group of written composition items, looking for high degrees of similarity. For the PARCC assessments, the primary item type of interest was the prose constructed-response (PCR) tasks in the English language arts/literacy (ELA/L) content area. This analysis was conducted for PCR tasks administered online using some of the same artificial intelligence (AI) techniques that are applied in automated essay scoring. Specifically, this method was based on Latent Semantic Analysis (LSA) technology to detect possible plagiarism. Using LSA, the content of each constructed response was compared against the content of every other constructed response and a measure that indicated the degrees of similarity was generated for each pair of response comparison. Because LSA provided a semantic representation of language, rather than a syntactic or word-based representation, it allowed the detection of potential copying behaviors, even when test takers or administrators substituted synonymous words or phrases.

3.5.4 Longitudinal Performance Monitoring

Longitudinal performance modeling evaluates the performance on PARCC assessments across test administrations and identifies unusual performance gains in the unit of interest (e.g., school or district). The original proposal was to use the cumulative logit regression (CLR) model approach (Clark, Skorupski, Jirka, McBride, Wang & Murphy, 2014) to identify unusual changes in test performance across two consecutive administrations of the PARCC assessment. A research study comparing the cumulative logit regression model and the weighted least squares (WLS) regression approach found that the cumulative logit regression model worked well when there were more units exhibiting unexpected performance. WLS resulted in lower false positives when few units exhibited unexpected performance. In addition, the units flagged by WLS were almost always flagged by CLR.

At the November 2016 Technical Advisory Committee (TAC) meeting, the results from this study were discussed and the recommendation to use the Weighted Least Squares methodology was endorsed by the TAC. In January of 2017, the State Leads approved the TAC recommendation.

In the weighted least squares regression approach, mean current year scale scores are regressed on mean prior year scale scores, weighting by unit sample size. The model is fit using SAS PROC REG or the lm function in R. Standardized residuals are calculated by dividing raw residuals by their respective standard deviations. Units with a standardized residual exceeding 3.0 are flagged for unexpected performance.

3.5.5 Internet and Social Media Monitoring

Internet and social media monitoring was conducted by Caveon, LLC. Caveon's team monitored English-language websites and searchable forums that were publicly available for suspected proxy testing solicitations and website postings that contain, or appear to contain, infringements of PARCC's protected operational test content. The Internet and social media outlets monitored included popular websites (such as Facebook and Twitter), blogs, discussion forums, video archives, document archives, brain dumps, auction sites, media outlets, peer-to-peer servers, etc. Caveon's process generated regular updates that categorize identified threats by level of actual or potential risk based upon the representations made on the websites, or actual analysis of the proffered content. For example, categorizations typically ranged from "cleared" (lowest risk but bookmarked for continued monitoring) to "severe" (highest risk). Note that this process only considered potential breaches of secure item content, not violations of testing administration policies. Potential breaches were reported directly to the state(s) implicated for further action. Summary reports describing the threats were provided to PARCC through notification emails.

3.5.6 Off-Hours Testing Monitoring

Off-hours testing monitoring checks for suspicious testing activities at test administration locations occurring outside of the set windows for computer-based testing sessions. PARCC states established set start and end times for administering computer-based assessments. Based on these hours, authorized users (that is, users with the State Role) were allowed to override the start and end times for a test session. The off-hours testing monitoring process tracked such occurrences and logged them in an operational report, which listed the sessions within an organization that selected to test outside the set window. PARCC states could use this report to follow-up with the organizations identified in the report.

Section 4: Item Scoring

4.1 Machine-Scored Items

4.1.1 Key-Based Items

Pearson performed a key review prior to the test administration to verify that the scoring (answer) keys were correct for each item. Once the forms were constructed and approved by PARCC for publication, an independent key review was performed by an experienced third-party vendor. The vendor reviewed each item and confirmed that the key was correct. If discrepancies were identified, a Pearson senior content specialist or content manager reviewed the flagged item(s) and worked with the item developers to resolve the issue.

4.1.2 Rule-Based Items

Rule-based scoring refers to item types that use various scoring models. PARCC uses Question and Test Interoperability (QTI) item type implementation based on scoring model rules. Examples of these item types include "choice interaction," which presents a set of choices where one or more choices can be selected; text entry, where the response is entered in a text box; hot spot or text interaction, where an area in a graph or text in a paragraph (for example) can be highlighted; or match interaction, where an association can be made between pairs of choices in a set. These items include the scoring rules and correct responses as part of their item XML (markup language) coding.

During the initial stages of item development for PARCC, Pearson staff worked closely with PARCC to first delineate the rules for the scoring rubrics and then to adjust those rules based on student responses. During the item tryout[3] planning phase, Pearson content staff received input from PARCC staff to develop a thorough rule-based scoring process that met PARCC needs.

Pearson worked with the item developers to review initial scoring rules created during the item development. Once the rule-based scoring process was approved by PARCC, and prior to test construction, Pearson content staff worked closely with the item developers to finalize scoring rubrics for items to be scored via the rule-based scoring method. The proposed scoring rubrics were sent to PARCC for review, and if any additional changes were needed or new rules added, Pearson documented and applied the requested edits.

During test construction, Pearson monitored and evaluated the scoring and updated the scoring keys/ scoring rules in the item bank. After the tryout items were scored, Pearson prepared a frequency distribution of student responses for each item or task scored using a rule-based approach and compared this to the expected response based on correct answers to ensure that scoring keys and rules were appropriately applied. The content team does this by analyzing the student response data to determine if scoring is acceptable using the item metadata and the student response file in conjunction with any potential item issues as flagged by psychometrics. These frequency distributions included an

[3] The item tryout was a set of item studies conducted in spring 2015.

indication of right/wrong and other identifying information defined by PARCC and those items that showed a statistical anomaly, whereby the frequency distribution was outside of the expected range, were sent to content experts to verify that the items were coded with the correct key.

Following the Rule-Based Scoring Educator Committee's review, which occurred prior to year one test construction, Pearson analyzed the feedback from the committees and made recommendations about adjustments to the scoring rubrics based on the results of the reviews. Upon submission of the results, Pearson worked with PARCC staff to discuss these findings and determine next steps prior to the completion of scoring. In subsequent years as scoring inquiries arise throughout the process of test construction, forms creation, testing, scoring, and psychometric analysis, items with scoring discrepancies are brought before the PARCC Priority Alert Task Force for resolution. This committee consists of representatives from each state as well as PARCC and Pearson content specialists.

Following the initial development of the PARCC rule-based scoring rubrics, Pearson has continued to monitor and evaluate new item development to ensure the scoring rules established are maintained within all item types as approved.

Pearson continues to use several avenues to monitor scoring each year. Prior to testing, a third-party key review checks operational and field test items for correct keys. Any disputed items go to a second review with Pearson content experts and anything still in question is taken before the PARCC task force for review and possible key change. During testing, Pearson creates early testing files for frequency distribution analysis whereby items for which an incorrect key receives a high distribution of responses are further evaluated for accuracy. After testing, all responses are again evaluated for the distribution of responses and potential scoring abnormalities during psychometric analysis. Any change in scoring that may be requested as a result of the psychometric analysis is also taken before the PARCC task force for decisions. These processes are the same for both paper and online modes of testing.

4.2 Human or Handscored Items

PARCC constructed-response items were scored by human scorers in a process referred to as handscoring. Online training units were used to train all scorers. The online training units included prompts (items), passages, rubrics, training sets, and qualification sets. Scorers who successfully completed the training and qualified, demonstrating they could correctly score student responses based on the guidelines in the online training units, were permitted to score student responses using the ePEN2 (Electronic Performance Evaluation Network, second generation) scoring platform. All online and paper responses were scored within the ePEN2 system. Pearson monitored quality throughout scoring.

Pearson staff roles and responsibilities were as follows:

- Scorers applied scores to student responses.
- Scoring Supervisors monitored the work of a team of scorers through review of scorer statistics and backreading, which is a review of responses scored by each scorer. When backreading, a supervisor sees the scores applied by scorers, which helps the supervisor provide additional coaching or instruction to the scorer being backread.

- Scoring directors managed the scoring quality of a subset of items and monitored the work of supervisors and scorers for their assigned items. Directors backread responses scored by supervisors and scorers as part of their quality-monitoring duties.
- English language arts/literacy (ELA/L) and mathematics content specialists managed the scoring quality and monitored the work of the scoring directors.
- Project managers documented the procedures, identified risks, and managed day-to-day administrative matters.
- A program manager provided oversight for the entire scoring process.

All Pearson employees involved in the scoring or the supervision of scoring possessed at least a four-year college degree.

4.2.1 Scorer Training

Key steps in the development of scorer training materials were rangefinding and rangefinder review meetings where educators and administrators from PARCC states met to interpret the scoring rubrics and determine consensus scores for student responses. Rangefinding meetings were held prior to scoring field-test items, and rangefinder review meetings were held prior to scoring operational items.

At rangefinding meetings, educators and administrators from PARCC states reviewed student responses and used scoring rubrics to determine consensus scores. Those responses scored in rangefinding were used to create field-test scorer training sets. After items were selected for operational testing, PARCC educators and administrators attended rangefinder review meetings to review and approve proposed operational scorer training sets.

When developing scorer training materials, Pearson scoring directors carefully reviewed detailed notes and records from PARCC rangefinding and rangefinder review committee meetings. Training sets were developed using the responses scored by the committees and additional suitable student response samples (as needed). PARCC reviewed and approved all scorer training sets prior to scorer training.

During training, scorers reviewed training sets of scored student responses with annotations that explained the rationale for the score assigned. The anchor set was the primary reference for scorers as they internalized the rubric during training. Each anchor set consisted of responses that were clear examples of student performance at each score point. The responses selected were representative of typical approaches to the task and arranged to reflect a continuum of performance. All scorers had access to the anchor set when they were training and scoring and were directed to refer to it regularly during scoring.

Practice sets were used in training to help trainees practice applying the scoring guidelines. Scorers reviewed the anchor sets, scored the practice sets, and then were able to compare their assigned scores for the practice sets to the actual PARCC-assigned scores to help them learn.

Qualification sets were used to confirm that scorers understood how to score student responses accurately. Qualification sets were composed of responses that were clear examples of score points.

Scorers were required to meet specified agreement percentages on qualification sets in order to score student responses.

Pearson has developed two types of training sets to train scorers: prototype and abbreviated sets. Prototype training sets were complete training sets consisting of anchor, practice, and qualification sets (refer to 4.2.2 for information on the qualification process). In ELA/L, there was one prototype training set per task type (Research Simulation Task, Literary Analysis Task, and Narrative Writing Task) at each of the nine grade levels (grades 3 through 11). In mathematics, a prototype training set was built for a grouping of similar items for a total of approximately three to five prototype sets per grade level or course.

The prototype training approach promoted consistency in scoring, as each subsequent abbreviated training set for the ELA/L task type or mathematics item grouping was based on the prototype. Once a prototype was chosen, full training materials were developed for that item, and at each grade level, scorers were trained to score a particular task type using the prototype training materials for that type.

Abbreviated training sets were prepared for all items not selected for prototype training sets. The abbreviated training sets included an anchor set and two practice sets so scorers could internalize the scoring standards for these new items, which were similar to prototype items they had previously scored.

Anchor and practice sets for both prototype and abbreviated items included annotations for each response. Annotations are formal written explanations of the score for each student response.

Table 4.1 details the composition of the anchor sets, practice sets, and qualification sets.

Table 4.1 Training Materials Used During Scoring

<table>
<tr><th colspan="2">Training Set Development</th></tr>
<tr><th>Description</th><th>Specification</th></tr>
<tr><td>Anchor Set</td><td></td></tr>
<tr><td rowspan="2">The anchor set is the primary reference for scorers as they internalize the rubric during training. All scorers have access to the anchor set when they are training and scoring, and are directed to refer to it regularly.

The anchor set comprises clear examples of student performance at each score point. The responses selected may be representative of typical approaches to the task or arranged to reflect a continuum of performance.</td><td>The anchor set for mathematics prototype items comprises three annotated responses per score point.

The anchor set for subsequent abbreviated items for mathematics comprise one to three annotated responses per score point.</td></tr>
<tr><td>The anchor sets for ELA/L prototype items comprise three annotated responses per score point. Anchor sets for prototype items include separate complete anchor sets for each applicable scoring trait (Reading Comprehension and Written Expression and Conventions [RCWE] for Research Simulation and Literary Analysis Tasks, Written Expression [WE] for Narrative Writing Tasks, and Knowledge of Language and Conventions for all task types).</td></tr>
<tr><td>Practice Sets</td><td></td></tr>
<tr><td rowspan="2">Practice sets are used to help trainees develop experience in independently applying the scoring guide (the rubric) to student responses. Some of these responses clearly reinforce the scoring guidelines presented in the anchor set. Other responses are selected because they are more difficult to evaluate, fall near the boundary between two score categories, or represent unusual approaches to the task.

The practice sets provide guidance and practice for trainees in defining the line between score categories, as well as applying the scoring criteria to a wider range of types of responses.</td><td>The practice sets for mathematics prototype and abbreviated items include two to three sets of ten annotated responses.</td></tr>
<tr><td>ELA/L practice sets for prototype items include two sets of five annotated responses and two sets of ten annotated responses.

The subsequent ELA/L practice sets for abbreviated items include two sets of ten annotated responses.</td></tr>
</table>

Table 4.1 Training Materials Used During Scoring

Training Set Development	
Description	**Specification**
Qualification Sets	
Qualification sets are used to confirm that scorer trainees understand the scoring criteria and are able to assign scores to student responses accurately. The responses in these sets are selected to reinforce the application of the scoring criteria illustrated in the anchor set. Scorer trainees must demonstrate acceptable performance on these sets by meeting a pre-determined standard for accuracy in order to qualify to score. Pearson scoring staff define and document qualifying standards in conjunction with PARCC prior to scoring.	The qualification sets for mathematics prototype items include three sets of ten responses each (not annotated). The subsequent mathematics abbreviated items for mathematics do not include qualification sets. The qualification sets for ELA/L prototype items include three sets of ten responses each (not annotated). The subsequent ELA/L abbreviated items do not include qualification sets.

4.2.2 Scorer Qualification

In order to score items, scorers were required to show that they were able to apply PARCC scoring methodology accurately through a qualification process. Scorers were asked to apply scores to three qualification sets consisting of ten responses each. ELA/L scorers applied a score for each trait on each response in the qualification sets. Literary Analysis and Research Simulation Tasks each had two traits: the Reading Comprehension and Written Expression trait and the Conventions trait. The Narrative Writing Task had two traits: Written Expression and Conventions. Mathematics scorers applied a score for each part of an item that was a constructed response. The number of constructed-response parts for each mathematics item ranged from one to four. Scorers were required to match the PARCC-approved score at a percentage agreed to by PARCC in order to qualify.

For ELA/L qualification, scorers were required to meet the following three conditions:

1. On at least one of the three qualifying sets, at least 70 percent of the ratings on each of the two scoring traits (considered separately) must agree exactly with the PARCC-approved scores.
2. On at least two of the three qualifying sets, at least 70 percent of the ratings (combined across the three scoring traits) must agree exactly with the PARCC-approved scores.
3. Combining over the three qualifying sets and across the two scoring traits, at least 96 percent of the ratings must be within one point of the PARCC-approved scores.

For mathematics qualification, the requirements were based on the item types and score point ranges. Because mathematics items can have one or more scoring traits, a scorer needed to achieve the following requirements separately for each scoring trait (when applicable to the item):

Table 4.2 Mathematics Qualification Requirements

Category	Score Point Range	Perfect Agreement	Within One Point
2	0–1	90%	100%
3	0–2	80%	96%
4	0–3	70%	96%
5	0–4	70%	95%
6	0–5	70%	95%
7	0–6	70%	95%

On at least two of the three qualifying sets, a scorer was required to meet the “perfect agreement” percentage indicated in the table above for each category. “Perfect agreement” was achieved when the scores applied exactly matched the PARCC-approved scores. Over the three qualifying sets, a scorer was required to meet the “within one point” percentage indicated in the table above for each category. The average is exclusive to each trait, so an item with multiple scoring traits would have multiple trait rating averages within one point of the PARCC-approved score.

4.2.3 Managing Scoring

Pearson created a handscoring specifications document that detailed the handscoring schedule, customer requirements, rangefinding plans, quality management plans, item information, and staffing plans for each scoring administration.

4.2.4 Monitoring Scoring

Second Scoring

During scoring, Pearson's ePEN2 scoring system automatically and randomly distributed a minimum of 10 percent of student responses for second scoring; scorers had no indication whether a response had been scored previously. Humans applied the second score for all mathematics items. Second scoring for ELA/L was performed either by human scorers or by the Intelligent Essay Assessor. If the first and second scores applied were non-adjacent, a third and occasionally a fourth score was assigned to resolve scorer disagreements. When a resolution score (i.e., third score) was nonadjacent to one or both of the first and second scores, the content specialist or scoring director would apply an adjudication score (fourth score).

Table 4.3 Scoring Hierarchy Rules

If a response was scored more than once, the following rules were applied to determine the final score:		
Score Type	**Rank**	**Final Score Calculation**
Adjudication	1	If an adjudication score is assigned, this is the final score.
Resolution	2	If no adjudication score is assigned, this is the final score.
Backread	3	If no adjudication or resolution score is assigned, the latest backreading score is the final score.
Human First Score	4	If no adjudication, resolution, or backreading score is assigned, this is the final score.
Human Second Score	5	If no adjudication, resolution, backreading, or human first score is assigned, this is the final score.
Intelligent Essay Assessor Score	6	If no human score is assigned, this is the final score.

Backreading

Backreading was one of the major responsibilities of Pearson Scoring Supervisors and a primary tool for proactively guarding against scorer drift, where scorers score responses in comparison to one another instead of in comparison to the training responses. Scoring supervisory staff used the ePEN2 backreading tool to review scores assigned to individual student responses by any given scorer in order to confirm that the scores were correctly assigned and to give feedback and remediation to individual scorers. Pearson backread approximately 5 percent of the handscored responses. Backreading scores did not override the original score but were used to monitor scorer performance.

Validity

Validity responses are pre-scored responses strategically interspersed in the pool of live responses. These responses were not distinguishable from any other responses so that scorers were not aware they

were scoring validity responses rather than live responses. The use of validity responses provided an objective measure that helped ensure that scorers were applying the same standards throughout the project. In addition, validity was at times shared with scorers in a process known as "validity as review." Validity as review provided scorers automated, immediate feedback: a chance to review responses they mis-scored, with reference to the correct score and a brief explanation of that score. One validity response was sent to scorers for every 25 "live" responses scored.

PARCC validity agreement requirements for scorers are listed in Table 4.4. Scorers had to meet the required validity agreement percentages to continue working on the PARCC project. Scorers who did not maintain expected agreement statistics were given a series of interventions culminating in a targeted calibration set: a test of scorer knowledge. Scorers who did not pass targeted calibration were removed from scoring the item, and all the scores they assigned were deleted.

Table 4.4 Scoring Validity Agreement Requirements

Subject	Score Point Range	Perfect Agreement	Within One Point
Mathematics	0–1	90%	96%*
Mathematics	0–2	80%	96%
Mathematics	0–3	70%	96%
Mathematics	0–4	65%	95%
Mathematics	0–5	65%	95%
Mathematics	0–6	65%	95%
ELA/L	Multi-trait	65%	96%

*A zero or 1 score compared to a blank score will have a disagreement greater than 1 point.

Calibration Sets

Calibration sets are special sets created during scoring to help train scorers on particular areas of concern or focus. Scoring directors used calibration sets to reinforce rangefinding standards, introduce scoring decisions, or address scoring issues and trends. Calibration was used either to correct a scoring issue or trend, or to continue scorer training by introducing a scoring decision. Calibration was administered regularly throughout scoring.

Inter-rater Agreement

Inter-rater agreement is the agreement between the first and second scores assigned to student responses and is the measure of how often scorers agree with each other. Pearson scoring staff used inter-rater agreement statistics as one factor in determining the needs for continuing training and intervention on both individual and group levels. PARCC inter-rater agreement expectations are shown in Table 4.5.

Table 4.5 Inter-rater Agreement Expectations and Results

Subject	Score Point Range	Perfect Agreement Expectation	Perfect Agreement Result	Within One Point Expectation	Within One Point Result
Mathematics	0–1	90%	97%	96%*	100%
Mathematics	0–2	80%	95%	96%	100%
Mathematics	0–3	70%	94%	96%	99%
Mathematics	0–4	65%	95%	95%	99%
Mathematics	0–5	65%	89%	95%	98%
Mathematics	0–6	65%	95%	95%	98%
ELA/L	Multi-trait	65%	77%	96%	99%

*A zero or 1 score compared to a blank score will have a disagreement greater than 1 point.

Pearson's ePEN2 scoring system included comprehensive inter-rater agreement reports that allowed supervisory personnel to monitor both individual and group performance. Based on reviews of these reports, scoring experts targeted individuals for increased backreading and feedback, and if necessary, retraining.

The perfect agreement rate for all mathematics responses scored by two scorers was 94 percent and the within one point rate was 99.5 percent. For all ELA/L responses scored by two scorers, the perfect agreement rate was 77 percent and the within one point rate was 99.3 percent.

The results by grade level for ELA/L are provided in Section 4.3.7: Inter-rater Agreement for Prose Constructed Response.

4.3 Automated Scoring for PARCC PCRs

Automated scoring performed by Pearson's Intelligent Essay Assessor (IEA) was the default option for scoring the PARCC assessment's online prose constructed-response (PCR) tasks. Under the default option, it was assumed that operational scores for approximately 90 percent of the online PCR responses would be assigned by IEA for the spring administration. The operational scores for the remaining online responses were assigned by human scorers. Human scoring was applied to responses that were scored while IEA was being trained as well as to additional responses routed to human scoring when there was uncertainty about the automated scores.

For 10 percent of responses, a second "reliability" score was assigned. The purpose of the reliability score was to provide data for evaluating the consistency of scoring, which is done by evaluating scoring agreement. When IEA provided the first score of record, the second reliability score was a human score.

4.3.1 Concepts Related to PARCC Automated Scoring

The text below describes concepts related to PARCC automated scoring.

Continuous Flow
Continuous flow scoring results in an integrated connection between human scoring and automated scoring. It refers to a system of scoring where either an automated score, a human score, or both can be assigned based on a predetermined asynchronous operational flow.

Training of IEA using Operational Data
Continuous flow scoring facilitates the training of IEA using human scores assigned to operational online data collected early in the administration. Once IEA obtains sufficient data to train, it can be "turned on" and becomes the primary source of scoring (although human scoring continues for the 10 percent reliability sample and other responses that may be routed accordingly).

Smart Routing
Smart routing refers to the practice of using automated scoring results to detect responses that are likely to be challenging to score, and applying automated routing rules to obtain one or more additional human scores. Smart routing can be applied prompt by prompt to the extent needed to meet scoring quality criteria for automated scoring.

Quality Criteria for Evaluating Automated Scoring
The PARCC state leads approved specific quality criteria for evaluating automated scoring at the time IEA was trained. The primary evaluation criteria for IEA was based on responses to validity papers with "known" scores assigned by experts. For each prompt scored, a set of validity papers is used to monitor the human-scoring process over time. Validity papers are seeded into human scoring throughout the administration. The expectation is that IEA can score validity papers at least as accurately as humans can.

Additional measures of inter-rater agreement for evaluating automated scoring were proposed based on the research literature (Williamson, Xi, & Breyer 2012). These measures were previously utilized in Pearson's automated scoring research and include Pearson correlation, kappa, quadratic-weighted kappa, exact agreement, and standardized mean difference. These measures are computed between pairs of human scores, as well as between IEA and humans, to evaluate how performance was the same or different. Criteria for evaluating the training of IEA given these measures include the following:

- Pearson correlation between IEA-human should be within 0.1 of human-human.
- Kappa between IEA-human should be within 0.1 of human-human.
- Quadratic-weighted kappa between IEA-human should be within 0.1 of human-human.
- Exact agreement between IEA-human should be within 5.25 percent of human-human.
- Standardized mean difference between IEA-human should be less than 0.15.

The specific criteria for evaluating IEA included both primary and secondary criteria and are noted below.

- Primary Criteria—Based on responses to validity papers: *With smart routing applied as needed, IEA agreement is as good as or better than human agreement for each trait score.*
- Contingent Primary Criteria—Based on the training responses if validity responses are not available: *With smart routing applied as needed, IEA-human exact agreement is within 5.25 percent of human-human exact agreement for each trait score.*
- Secondary Criteria—Based on the training responses: *With smart routing applied as needed, IEA-human differences on statistical measures for each trait score are within the Williamson et al. tolerances for subgroups with at least 50 responses.*

Hierarchy of Assigned Scores for Reporting

When multiple scores are assigned for a given response, the following hierarchy determines which score was reported operationally:

- The IEA score is reported if it is the only score assigned.
- If an IEA score and a human score are assigned, the human score is reported.
- If two human scores are assigned, the first human score is reported.
- If a backread score and human and/or IEA scores are assigned, the backread score is reported.
- If a resolution score is assigned and an adjudicated score is not assigned, the resolution score is reported (note that if nonadjacent scores are encountered, responses are automatically routed to resolution).
- If an adjudicated score is assigned, it is reported (note that if a resolution score is nonadjacent to the other scores assigned, responses are automatically routed to adjudication).

4.3.2 Sampling Responses Used for Training IEA

For prompts trained using 2018 operational data, the early performance of human scoring was closely monitored to verify that an appropriate set of data would be available for training IEA. In particular, several characteristics of the human scoring data were monitored, including:

- exact agreement between human scorers (the goal was for this to be at least 65 percent for each trait);
- exact agreement between human scores conditioned on score point (the goal was for this to be at least 50 percent for each trait);
- the number of responses at each score point (the goal was to have at least 40 responses at the highest score points in the training samples used by IEA); and
- the number of responses with two human scores assigned (note that IEA "ordered" additional scoring of responses during the sampling period as needed).

Although the desired characteristics of the training data were easily achieved for some prompts, they were more challenging to achieve for others. For some prompts, a subset of scores were reset and

clarifying directions were provided to scorers to improve human-human agreement. For other prompts, special sampling approaches were used to increase the numbers of responses that received top scores. In addition, a healthy percentage of responses were backread during the sampling period and these scores as well as double human scores were all part of the data used to train IEA.

4.3.3 Primary Criteria for Evaluating IEA Performance

The primary criteria for evaluating IEA performance is based on evaluating validity papers and is stated as follows: *With smart routing applied as needed, IEA agreement is as good as or better than human agreement for each trait score.*

To operationalize the primary criteria for a given prompt, the following general steps are undertaken:

1. Determine agreement of the human scores with the validity papers for each trait.
2. Calculate agreement of the IEA scores with the validity papers for each trait.
3. Compare the IEA validity agreement with the human agreement.
4. If the IEA validity agreement is greater than or equal to the human agreement for each trait, IEA can be deployed operationally.

In addition to looking at overall validity agreement, conditional agreement was also examined. In general, it was desirable for IEA to exceed 65 percent agreement at every score point as well as be close to or exceed the human validity agreement at each score point.

4.3.4 Contingent Primary Criteria for Evaluating IEA Performance

For many of the prompts trained in 2018, it was not possible to utilize human-scored validity responses in evaluating IEA performance. In these cases, IEA was evaluated based on IEA-human exact agreement for each trait score and compared to agreement based on responses that were double-scored by humans. A portion of the data was held out for evaluating IEA-human exact agreement according to the following steps:

1. Determine exact agreement of the two human scores with each other for each trait.
2. Calculate agreement of the IEA scores with the human scores for each trait.
3. Compare the IEA-human agreement with the human-human agreement.
4. If the IEA-human agreement is within 5.25 percent of the human-human agreement, IEA can be deployed operationally.

In addition to the overall comparison, we targeted the following performance thresholds in the test data set: 1) at least 65 percent overall IEA-human agreement; and 2) 50 percent IEA-human agreement by score point (i.e., conditioned on the human score). These targets went beyond the contingent primary criteria approved by the state leads.

4.3.5 Applying Smart Routing

With smart routing, the quality of automated scoring can be increased by routing responses that are more likely to disagree with a human score to receive an additional human score.

When human scorers read a paper, they typically apply integer scores based on a scoring rubric. When there is strong agreement between two independent human readers, they might both assign a score of 3 such that the average score over both raters is also a 3 (i.e., (3+3)/2 = 3). IEA simulates this behavior, but because its scores come from an artificial intelligence algorithm, it generates continuous (i.e., decimalized) scores. In this case, the IEA score might be a 2.9 or 3.1. When human readers disagree on the score for a paper, say one reader gives the paper a score of 3 and another reader gives the paper a score of 4, the average of the two scores would be 3.5 (i.e., 3+4=7/2=3.5). For this paper, IEA would likely provide a score between 3 and 4, say 3.4 or 3.6. Because this continuous score needs to be rounded to an integer score for reporting, it might be reported as a 3 or a 4, depending on the rounding rules. Smart routing involves routing those responses with "in between" IEA scores to additional human scoring because the nature of the responses suggests there may be less confidence in the IEA score. Since these "in between" IEA scores are based on modeling human scores, it follows that human scores may be less certain as well, and thus such responses tend to be the ones that it makes sense to have double-scored and possibly to resolve if the IEA and human scores are non-adjacent.

Smart routing was utilized as needed to help IEA achieve targeted quality metrics (e.g., validity agreement or agreement with human scorers). Smart routing involved the application of the following four steps:

1. The continuous IEA score for each of the two trait scores was rounded to the nearest score interval of 0.2, starting from zero. For example, IEA scores between 0 and 0.1 were rounded to an interval score of 0, scores between 0.1 and 0.3 were rounded to an interval score of 0.2, scores between 0.3 and 0.5 were rounded to an interval score of 0.4, and so on.
2. Within each of these intervals, the percentage of exact agreement between IEA integer scores and the human scores was calculated for each trait.
3. For each prompt, agreement rates were evaluated by rounding interval. Those intervals for which the agreement rates were below a designated threshold for either trait were identified.
4. Once IEA scoring was implemented, responses within intervals for which IEA-human agreement was below the designated threshold were routed for additional human scoring.

In training IEA, we first evaluated the scoring models without smart routing by applying either the primary validity criteria or the contingent criteria as described in Section 4.3. For those prompts that did not meet these criteria, we applied increasing smart routing thresholds in an iterative fashion to filter scores and evaluate the remaining scores against the criteria. That is, in any one iteration a particular smart routing threshold was applied such that only scores falling in intervals for which exact agreement exceeded the threshold were included in evaluating the criteria. If the primary or contingent criteria were not met with this level of smart routing, we repeated the analysis applying an increased smart routing threshold. If the primary or contingent criteria were still not met, we repeated the analysis

applying a still higher threshold. If the criteria were still not met after a maximum threshold was applied, we investigated different models and/or utilized additional human scoring data until an IEA scoring model was found that met the criteria.

4.3.6 Evaluation of Secondary Criteria for Evaluating IEA Performance

The secondary criteria for evaluating IEA performance involved comparing agreement indices for IEA-human scoring for various demographic subgroups. Because of the importance of protecting personally identifiable information (PII), student demographic data is stored and managed separately from the performance scoring data. For this reason, it was not possible to evaluate subgroup performance in real time as IEA was being trained.

For those prompts trained on early operational data, attempts were made to prioritize the data being returned from the field to include data from states or districts where more diverse populations of students were anticipated. In addition, requests for additional human scores were made to increase the likelihood that there would be sufficient numbers of responses with two human scores for most of the demographic subgroups of interest.

Once IEA was trained and deployed, scoring sets used in training were matched to demographic information so that agreement between IEA and human scorers could be evaluated across subgroups. The analysis was conducted for the following ten comparison groups:

Table 4.6 Comparison Groups

Group Type	**Comparison Groups**
Sex	Female
	Male
Ethnicity	American Indian/Alaska Native
	Asian
	Black/African American
	Hispanic/Latino
	Native Hawaiian or Other Pacific Islander
	White
Special Instructional Needs	English Language Learners (ELL)
	Students with Disabilities (SWD)

IEA-human agreement indices were calculated for all cases with an IEA score and at least one human score. Human-human agreement was calculated for all cases with two human scores.

To evaluate the training of IEA for subgroups, we applied the following criteria approved by the state leads for subgroups with at least 50 IEA-human scores and at least 50 human-human scores:

- Pearson correlation between IEA-human should be within 0.1 of human-human.
- Kappa between IEA-human should be within 0.1 of human-human.

- Quadratic-weighted kappa between IEA-human should be within 0.1 of human-human.
- Exact agreement between IEA-human should be within 5.25 percent of human-human.
- Standardized mean difference between IEA-human should be less than ±0.15 (this criterion was applied to subgroups with at least 50 IEA-human scores).

Although it was not expected that these criteria would be met for all subgroups for all prompts, if results of the evaluation between IEA and human scoring for subgroups for any prompt indicated that IEA performance persistently failed on the criteria listed above, consideration would be given to resetting the responses scored by IEA and reverting to human scoring until such time that an alternate IEA model could be established with improved subgroup performance.

In addition to the secondary criteria approved by the State Leads, we also compared the performance of IEA to the following targets on the various measures for subgroups with at least 50 responses:

- Pearson correlation between IEA-human should be 0.70 or above.
- Kappa between IEA-human should be 0.40 or above.
- Quadratic-weighted kappa between IEA-human should be 0.70 or above.
- Exact agreement between IEA-human should be 65 percent or above.

These targets were not intended to be directly applied in decisions about whether to deploy IEA operationally or not. Such targets may or may not be met by human scoring for any particular prompt and/or subgroup, and if they are not met by human scoring, they are unlikely to be met by IEA scoring. Nevertheless, comparisons to these targets provided additional information about IEA performance (and human scoring) in an absolute sense.

4.3.7 Inter-rater Agreement for Prose Constructed Response

This section presents the inter-rater agreement for operational results for the online prose constructed-response (PCR) tasks by trait and grade level in spring 2018. PCR task items are scored on two traits: (1) Reading Comprehension and Written Expression and (2) Knowledge of Language and Conventions.

For 10 percent of responses, a second “reliability” score was assigned. The purpose of the reliability score is to provide data for evaluating the consistency of scoring, which is done by evaluating scoring agreement. Inter-rater agreement is the agreement between the first and second scores assigned to student responses and is the measure of how often scorers agree with each other. Pearson scoring staff used inter-rater agreement indices as one factor in determining the needs for continuing training and intervention on both individual and group levels. PARCC inter-rater agreement expectations are provided in Table 4.5 in Section 4.2.4. For ELA/L PCR traits, the expectation for agreement is an inter-rater agreement of 65 percent or higher between two scorers. When IEA provided the first score of record, the second reliability score was a human score. For those states choosing the human-scoring option, the second reliability score was assigned by IEA. For a subset of responses, the first and second score were both human scores.

Table 4.7 presents the average agreement across the PCRs for each grade level by trait. The number of prompts included in the analyses is listed for each grade level. The agreement indices (exact agreement, kappa, quadratic-weighted kappa, and Pearson correlation) were calculated separately by PCR for each trait (Written Expression and Conventions). For each grade level, the agreement indices were averaged across the PCRs. The table presents the average count and the average for the agreement indices.

The exact agreement for the PCR traits is above the 65 percent agreement rate criteria for all PCRs. The strength of agreement between raters is moderate to substantial agreement as defined by Landis and Koch (1977) for all PCRs. The quadratic-weighted kappa (QW Kappa) distinguishes between differences in ratings that are close to each other versus larger differences. The weighted kappa is substantial to almost perfect agreement for all grades. The Pearson correlations (*r*) ranged from .71 to .89.

During operational scoring, the PCR agreement rates are monitored for quality and items not meeting the criteria are shared with the PARCC handscoring group. After the operational administration, the performance of all the PCRs is provided to the content team as feedback for re-using PCRs and in order to inform development of future PCRs. This provides evidence for continuous improvement of the testing program.

Table 4.7 PARCC PCR Average Agreement Indices by Test

			Written Expression				Conventions			
Test	Number of PCRs	Count	Exact	Kappa	QW Kappa	*r*	Exact	Kappa	QW Kappa	*r*
ELA03	5	25,547	70.72	0.53	0.72	0.73	71.20	0.55	0.76	0.76
ELA04	5	24,748	67.62	0.54	0.81	0.82	70.82	0.58	0.82	0.82
ELA05	5	24,196	69.84	0.56	0.82	0.82	70.20	0.58	0.82	0.82
ELA06	5	29,958	73.72	0.63	0.87	0.87	73.28	0.62	0.85	0.85
ELA07	5	23,718	73.88	0.64	0.89	0.89	73.88	0.64	0.87	0.87
ELA08	5	24,163	74.00	0.66	0.91	0.91	76.00	0.67	0.89	0.89
ELA09	5	9,273	72.40	0.64	0.89	0.90	74.94	0.66	0.88	0.88
ELA10	5	12,763	74.62	0.66	0.89	0.89	75.48	0.67	0.88	0.88
ELA11	5	7,529	75.66	0.66	0.89	0.89	76.42	0.67	0.87	0.87

Section 5: Test Taker Characteristics

5.1 Overview of Test Taking Population

Approximately two and half million students participated in the operational administration of the PARCC assessments during the 2017–2018 school year in the Bureau of Indian Education (BIE), the Department of Defense Education Activity (DD), the District of Columbia (DC), Illinois (IL), Maryland (MD), New Jersey (NJ), and New Mexico (NM). Not all participating states had students testing in all grades. Assessments were administered for English language arts/literacy (ELA/L) in grades 3 through 11; mathematics assessments were administered in grades 3 through 8, as well as for traditional high school mathematics (Algebra I, Geometry, and Algebra II) and integrated high school mathematics (Integrated Mathematics I, II, and III). A small subset of students tested in ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II during fall of 2017. Test taker characteristics for this group are presented in an addendum.[4] The majority of students tested during the spring 2018 window when all grades and content areas were administered online and on paper.

5.2 Composition of Operational Forms

The fall 2017 and spring 2018 PARCC assessments were administered in either a computer-based test (CBT) or a paper-based test (PBT) format. ELA/L assessments focused on writing effectively when analyzing text. Mathematics assessments focused on applying skills and concepts, and featured multi-step problems that require abstract reasoning and modeling of real-world problems. Solving such problems requires the strategic use of tools, as well as precision and perseverance. In both content areas, students also demonstrated their acquired skills and knowledge by answering selected response items and fill-in-the-blank questions. Each assessment was comprised of multiple units; one of the mathematics units was split into calculator and non-calculator sections.

5.3 Rules for Inclusion of Students in Analyses

Criteria for inclusion of students were implemented prior to all operational analyses. These rules were established by Pearson psychometricians in consultation with PARCC to determine which, if any, student records should be removed from analyses. This data screening process resulted in higher quality, albeit slightly smaller, data sets.

Student response data were included in analyses if:

1. Valid form numbers were observed for each unit for online assessments or for the full form for paper assessments,
2. Student records were not flagged as “void” (i.e., do not score), and
3. The student attempted at least 25 percent of the items in each unit or form.

[4] Addendum 5 presents a summary of the test taker characteristics for the fall 2017 administration.

Additionally, in cases where students had more than one valid record, the record with the higher raw score was chosen. Records for students with administration issues or anomalies were excluded from analyses.

5.4 Test Takers by Grade, Mode, and Gender

Table 5.1 presents, for each grade of ELA/L, the number and percentage of students who took the test in each mode (CBT or PBT). This information is provided for all participating states combined. Table 5.2 presents the same type of information for all students who took the mathematics assessments, and Table 5.3 provides this information for students who took the mathematics assessments in Spanish.

Markedly more students tested online than on paper across all grades for both content areas. For ELA/L, the percentages of online test takers by grade level, for all states combined, ranged from 87.5 percent to 99.5 percent, while the percentages of paper test takers ranged from .5 percent to 12.5 percent. For all mathematics test takers, the percentages of students testing online ranged from 87.7 percent to 100 percent, whereas the percentages of students testing on paper ranged from 0 percent to 12.3 percent. The percentages of students taking Spanish-language mathematics online forms ranged from 86.1 percent to 100 percent and the percentages of students taking Spanish-language mathematics paper forms ranged from 0 percent to 13.9 percent. Generally, the percentage of students who tested online increased steadily from the lower grades to the higher grades. For example, about 88 percent of the ELA/L grade 3 students tested online, while about 99 percent of the grade 11 students tested online. Overall, fewer students tested at the higher grades for both content areas.

Table 5.1 ELA/L Test Takers by Grade and Mode: All States Combined

	No. of Valid	CBT		PBT	
Grade	Cases	N	%	N	%
3	338,927	296,516	87.5	42,411	12.5
4	345,483	304,903	88.3	40,580	11.7
5	348,524	333,909	95.8	14,615	4.2
6	344,520	333,518	96.8	11,002	3.2
7	338,731	331,387	97.8	7,344	2.2
8	339,283	331,657	97.8	7,626	2.2
9	128,229	127,586	99.5	643	0.5
10	188,597	187,426	99.4	1,171	0.6
11	102,755	101,851	99.1	904	0.9
Grand Total	2,475,049	2,348,753	94.9	126,296	5.1

Note: Includes students taking accommodated forms of ELA/L.

Table 5.2 Mathematics Test Takers by Grade and Mode: All States Combined

Grade	No. of Valid Cases	CBT		PBT	
		N	%	N	%
3	348,117	305,323	87.7	42,794	12.3
4	354,080	313,273	88.5	40,807	11.5
5	355,854	341,143	95.9	14,711	4.1
6	345,712	334,675	96.8	11,037	3.2
7	323,440	316,116	97.7	7,324	2.3
8	263,809	256,332	97.2	7,477	2.8
A1	221,242	220,012	99.4	1,230	0.6
GO	130,412	129,723	99.5	689	0.5
A2	123,787	122,954	99.3	833	0.7
M1	750	750	100	n/a	n/a
M2	873	854	97.8	19	2.2
M3	191	186	97.4	5	2.6
Grand Total	2,468,267	2,341,341	94.9	126,926	5.1

Note: Includes students taking mathematics in English, students taking Spanish-language forms for mathematics, and students taking accommodated forms. A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. n/a = not applicable.

Table 5.3 Spanish-Language Mathematics Test Takers, by Grade and Mode: All States Combined

Grade	No. of Valid Cases	CBT		PBT	
		N	%	N	%
3	4,906	4,223	86.1	683	13.9
4	3,647	3,178	87.1	469	12.9
5	2,960	2,916	98.5	44	1.5
6	2,536	2,516	99.2	20	0.8
7	2,586	2,566	99.2	20	0.8
8	2,108	2,103	99.8	5	0.2
A1	3,579	3,203	89.5	376	10.5
GO	1,961	1,958	99.8	3	0.2
A2	1,300	1,300	100	n/a	n/a
M1	6	6	100	n/a	n/a
M2	11	11	100	n/a	n/a
M3	1	n/a	n/a	1	100
Grand Total	25,601	23,980	93.7	1,621	6.3

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. n/a = not applicable.

Tables A.5.1, A.5.2, and A.5.3 in Appendix 5 show the number and percentage of students with valid test scores in each content area (including Spanish-language mathematics), grade, and mode of assessment for all states combined and for each state separately. Tables A.5.4, A.5.5, and A.5.6 present the distribution by content area, grade, mode, and gender, for all states combined.

5.5 Demographics

Also presented in Appendix 5 is student demographic information for the following characteristics: economically disadvantaged, students with disabilities, English learners (EL), gender, and race/ethnicity (American Indian/Alaska Native; Asian; Black/African American; Hispanic/Latino; White/Caucasian; Native Hawaiian or Other Pacific Islander; two or more races reported; race not reported). Student demographic information was provided by the states and districts and captured in PearsonAccess^{next} by means of a student data upload. The demographic data was verified by the states and districts prior to score reporting.

Tables A.5.7 through A.5.15 provide demographic information for students with valid ELA/L scores, and Tables A.5.16 through A.5.27 present demographics for students with valid mathematics scores. All tables of demographic information are organized by grade; the results are first aggregated across all PARCC states and then presented for each state. Percentages are not reported for any states in which fewer than 20 students tested in a grade/content area.

Section 6: Classical Item Analysis

6.1 Overview

This section describes the results of the classical item analysis conducted for data obtained from the operational test items. For the spring 2018 administration, mathematics online and paper and ELA/L paper tests were pre-equated; ELA/L online tests were post-equated. For pre-equated tests, the item statistics provided in this section are from prior operational administrations and reflect the statistics that were used at test construction. For the post-equated tests, the statistics provided in this section are from the spring 2018 administration. Item analysis serves two purposes: to inform item exclusion decisions for IRT analysis and to provide item statistics for the item bank.

PARCC item analysis included data from the following types of items: key-based selected-response items, rule-based machine-scored items, and hand-scored constructed-response items. For each item, the analysis produced item difficulty, item discrimination, and item response frequencies.

6.2 Data Screening Criteria

Item analyses were conducted by test form based on administration mode. In preparation for item analysis, student response files were processed to verify that the data were free of errors. Pearson Customer Data Quality (CDQ) staff ran predefined checks on all data files and verified that all fields and data needed to perform the statistical analyses were present and within expected ranges.

Before beginning item analysis, Pearson performed the following data screening operations:

1. All records with an invalid form number were excluded.
2. All records that were flagged as “void” were excluded.
3. All records where the student attempted fewer than 25 percent of items were excluded.
4. For students with more than one valid record, the record with the higher raw score was chosen.
5. Records for students with administration issues or anomalies were excluded.

6.3 Description of Classical Item Analysis Statistics

A set of classical item statistics were computed for each operational item by form and by administration mode. Each statistic was designed to evaluate the performance of each item.

The following statistics and associated flagging rules were used to identify items that were not performing as expected:

Classical item difficulty indices (p-value and average item score)
When constructing PARCC tests, a wide range of item difficulties is desired (i.e., from easy to hard items) so that students of all ability levels can be assessed with precision. At the operational stage, item difficulty statistics are used by test developers to build forms that meet desired test difficulty targets.

For dichotomously scored items, item difficulty is indicated by its p-value, which is the proportion of test takers who answered that item correctly. The range for p-values is from .00 to 1.00. Items with high p-values are easy items and those with low p-values are difficult items. Dichotomously scored items were flagged for review if the p-value was above .95 (i.e., too easy) or below .20 (i.e., too difficult).

For polytomously scored items, difficulty is indicated by the average item score (AIS). The AIS can range from .00 to the maximum total possible points for an item. To facilitate interpretation, the AIS values for polytomously scored items are often expressed as percentages of the maximum possible score, which are equivalent to the p-values of dichotomously scored items. The desired p-value range for polytomously scored items is .30 to .80; items with values outside this range were flagged for review.

The percentage of students choosing each response option

Selected-response items on PARCC assessments refer primarily to single-select multiple choice scored items. These items require that the test taker select a response from a number of answer options. These statistics for single-select multiple choice items indicate the percentage of students who select each of the answer options and the percentage that omit the item. The percentages are also computed for the high-performing subgroup of students who scored at the top 20 percent on the assessment. Items were flagged for review if more high-performing test takers chose the incorrect option than the correct response. Such a result could indicate that the item has multiple correct answers or is miskeyed.

Item-total correlation

This statistic describes the relationship between test takers' performance on a specific item and their performance on the total test. The item-total correlation is usually referred to as the item discrimination index. For PARCC operational item analysis, the total score on the assessment was used as the total test score. The polyserial correlation was calculated for both selected-response items and constructed-response items as an estimate of the correlation between an observed continuous variable and an unobserved continuous variable hypothesized to underlie the variable with ordered categories (Olsson, Drasgow, and Dorans, 1982). Item-total correlations can range from -1.00 to 1.00. Desired values are positive and larger than .15. Negative item-total correlations indicate that low-ability test takers perform better on an item than high-ability test takers, an indication that the item may be potentially flawed. Item-total correlations below .15 were flagged for review. Items with extremely low or negative values were considered for exclusion from IRT calibrations or linking (refer to Section 10 for details on item inclusion and exclusion criteria for IRT analyses).

Distractor-total correlation

For selected-response items, this estimate describes the relationship between selecting an incorrect response (i.e., a distractor) for a specific item and performance on the total test. The item-total correlation is calculated (refer to #3) for the distractors. Items with distractor-total correlations above .00 were flagged for review as these items may have multiple correct answers, be miskeyed, or have other content issues.

Percentage of students omitting or not reaching each item

For both selected-response and constructed-response items, this statistic is useful for identifying problems with test features such as testing time and item/test layout. Typically, if students have an adequate amount of testing time, approximately 95 percent of students should attempt to answer each question on the test. A distinction is made between "omit" and "not reached" for items without responses.

- An item is considered "omit" if the student responded to subsequent items.
- An item is considered "not reached" if the student did not respond to any subsequent items.

Patterns of high omit or not-reached rates for items located near the end of a test section may indicate that test takers did not have adequate time. Items with high omit rates were flagged. Omit rates for constructed-response items tend to be higher than for selected-response items. Therefore, the omit rate for flagging individual items was 5 percent for selected-response items and 15 percent for constructed-response items. If a test taker omitted an item, then the test taker received a score of 0 for that item and was included in the n-count for that item. However, if an item was near the end of the test and classified as not reached, the test taker did not receive a score and was not included in the n-count for that item.

Distribution of item scores

For constructed-response items, examination of the distribution of scores is helpful to identify how well the item is functioning. If no students' responses are assigned the highest possible score point, this may indicate that the item is not functioning as expected (e.g., the item could be confusing, poorly worded, or just unexpectedly difficult), the scoring rubric is flawed, and/or test takers did not have an opportunity to learn the content. In addition, if all or most test takers score at the extreme ends of the distribution (e.g., 0 and 2 for a 3-category item), this may indicate that there are problems with the item or the rubric so that test takers can receive either full credit or no credit at all, but not partial credit.

The raw score frequency distributions for constructed-response items were computed to identify items with few or no observations at any score points. Items with no observations or a low percentage (i.e., less than 3 percent) of test takers obtaining any score point were flagged. In addition, constructed-response items were flagged if they had U-shaped distributions, with high frequencies for extreme scores and very low frequencies for middle score categories. Items with such response patterns may pose problems during the IRT calibrations and therefore may need to be excluded (refer to Section 10 for more information).

6.4 Summary of Classical Item Analysis Flagging Criteria

In summary, items are flagged for review if the item analysis yielded any of the following results:

1. p-value above .95 for dichotomous items and above .80 for polytomous items
2. p-value below .25 for dichotomous items and below .30 for polytomous items
3. item-total correlation below .15

4. any distractor-total correlation above .00
5. greater number of high-performing students (top 20 percent) choosing a distractor rather than the keyed response
6. high percentage of omits: above 5 percent for selected-response items and above 15 percent for constructed-response items
7. high percentage that did not reach the item: above 5 percent for selected-response items and above 15 percent for constructed-response items
8. constructed-response items with a score value obtained by less than 3 percent of responses

Pearson's psychometric staff carefully reviewed the flagged items and brought items to the PARCC Priority Alert Task Force to decide if the items were problematic and should be excluded from scoring.

6.5 Classical Item Analysis Results

This section presents tables summarizing the analyses for items on the spring 2018 operational forms. Spring 2018 PARCC assessments (except ELA/L online) were pre-equated, meaning that the scoring was based on item parameters estimated using data from earlier administrations. For the pre-equated grades/subjects, item analysis results in this section are the item statistics from prior administrations that were used to make decisions during test construction and for scoring. For ELA/L online, the item analysis is from the spring 2018 operational administration.

- Tables 6.1 and 6.2 present p-value information by grade and mode for the ELA/L and mathematics operational items.
- Tables 6.3 and 6.4 present item-total correlations by grade and mode for the ELA/L and mathematics operational items.

An operational item could appear on multiple test forms. The tables list only unique items in each test mode, and the reported item statistics are based on student responses across multiple occurrences of an item.

Spoiled or "do not score" items were excluded from the total test score in item analysis. These items were removed from scoring because of item performance, technical scoring issues, content concerns, or multiple/no correct answers. Additionally, some items were dropped during item calibrations due to:

- a low weighted polyserial,
- a low p-value (e.g., extremely difficult item), or
- extremely poor IRT model fit or item not able to calibrate.

Table 10.2 in section 10 presents the count and percentage of items excluded from IRT calibration along with the reasons the items were excluded for ELA/L. The tables in this section and in the Appendix include only those items that were used for operational scoring.

The fall 2017 forms were based on the spring 2017 operational forms; therefore, the item analyses for these forms were reported in the 2016–2017 Technical Report.

Table 6.1 Summary of p-Values for ELA/L Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean p-Value	SD p-Value	Min p-Value	Max p-Value	Median p-Value
3	CBT	46	0.44	0.17	0.14	0.83	0.44
3	PBT	27	0.40	0.17	0.17	0.77	0.41
4	CBT	59	0.45	0.14	0.22	0.78	0.44
4	PBT	36	0.42	0.14	0.23	0.72	0.40
5	CBT	56	0.44	0.16	0.17	0.80	0.39
5	PBT	36	0.39	0.14	0.17	0.76	0.35
6	CBT	60	0.44	0.15	0.20	0.86	0.44
6	PBT	36	0.46	0.17	0.22	0.76	0.39
7	CBT	61	0.46	0.13	0.22	0.82	0.44
7	PBT	36	0.43	0.14	0.24	0.82	0.40
8	CBT	57	0.43	0.13	0.16	0.80	0.40
8	PBT	36	0.49	0.17	0.27	0.86	0.42
9	CBT	63	0.47	0.12	0.26	0.74	0.45
9	PBT	36	0.44	0.14	0.19	0.74	0.43
10	CBT	60	0.41	0.12	0.14	0.73	0.41
10	PBT	36	0.37	0.10	0.20	0.64	0.36
11	CBT	58	0.37	0.11	0.12	0.65	0.37
11	PBT	36	0.33	0.11	0.13	0.65	0.31

Note: CBT = computer-based testing (online); PBT = paper-based testing (paper).

Table 6.2 Summary of p-Values for Mathematics Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean p-Value	SD p-Value	Min p-Value	Max p-Value	Median p-Value
3	CBT	86	0.54	0.22	0.14	0.94	0.52
3	PBT	42	0.54	0.23	0.11	0.94	0.51
4	CBT	80	0.52	0.19	0.15	0.94	0.51
4	PBT	40	0.49	0.24	0.09	0.94	0.49
5	CBT	80	0.47	0.19	0.10	0.86	0.47
5	PBT	40	0.46	0.18	0.15	0.79	0.47
6	CBT	76	0.40	0.19	0.09	0.93	0.38
6	PBT	38	0.41	0.20	0.09	0.93	0.39
7	CBT	76	0.37	0.20	0.08	0.83	0.33
7	PBT	38	0.37	0.21	0.07	0.83	0.30
8	CBT	72	0.34	0.20	0.06	0.84	0.30
8	PBT	36	0.36	0.20	0.08	0.84	0.34
A1	CBT	82	0.27	0.17	0.05	0.69	0.22
A1	PBT	42	0.29	0.18	0.05	0.73	0.27
GO	CBT	81	0.27	0.15	0.05	0.83	0.23
GO	PBT	41	0.26	0.13	0.06	0.59	0.27
A2	CBT	84	0.28	0.18	0.05	0.80	0.25
A2	PBT	43	0.32	0.21	0.06	0.90	0.30
M1	CBT	42	0.31	0.15	0.05	0.65	0.30
M1	PBT	42	0.34	0.16	0.05	0.70	0.34
M2	CBT	41	0.25	0.17	0.01	0.62	0.22
M2	PBT	42	0.70	0.53	0.01	1.57	0.52
M3	CBT	40	0.28	0.17	0.04	0.82	0.26
M3	PBT	39	0.54	0.43	0.03	1.83	0.41

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based testing (online); PBT = paper-based testing (paper).

Table 6.3 Summary of Item-Total Correlations for ELA/L Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean Polyserial	SD Polyserial	Min Polyserial	Max Polyserial	Median Polyserial
3	CBT	46	0.53	0.14	0.23	0.78	0.52
3	PBT	27	0.56	0.15	0.18	0.74	0.57
4	CBT	59	0.49	0.15	0.22	0.81	0.47
4	PBT	36	0.51	0.16	0.22	0.81	0.46
5	CBT	56	0.48	0.17	0.20	0.83	0.43
5	PBT	36	0.47	0.17	0.20	0.79	0.41
6	CBT	60	0.51	0.16	0.27	0.84	0.48
6	PBT	36	0.56	0.16	0.32	0.84	0.51
7	CBT	61	0.48	0.18	0.19	0.86	0.44
7	PBT	36	0.53	0.19	0.29	0.86	0.46
8	CBT	57	0.49	0.18	0.21	0.86	0.45
8	PBT	36	0.54	0.18	0.25	0.85	0.48
9	CBT	63	0.52	0.18	0.26	0.88	0.49
9	PBT	36	0.54	0.16	0.27	0.85	0.52
10	CBT	60	0.47	0.19	0.18	0.86	0.41
10	PBT	36	0.49	0.21	0.18	0.86	0.41
11	CBT	58	0.49	0.17	0.21	0.85	0.44
11	PBT	36	0.49	0.19	0.24	0.85	0.43

Note: CBT = computer-based testing (online); PBT = paper-based testing (paper).

Table 6.4 Summary of Item-Total Correlations for Mathematics Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean Polyserial	SD Polyserial	Min Polyserial	Max Polyserial	Median Polyserial
3	CBT	86	0.53	0.13	0.22	0.79	0.54
3	PBT	42	0.53	0.14	0.24	0.78	0.55
4	CBT	80	0.53	0.12	0.22	0.79	0.55
4	PBT	40	0.54	0.13	0.27	0.80	0.55
5	CBT	80	0.50	0.14	0.19	0.74	0.51
5	PBT	40	0.53	0.13	0.25	0.81	0.54
6	CBT	76	0.53	0.14	0.22	0.82	0.54
6	PBT	38	0.55	0.13	0.23	0.82	0.59
7	CBT	76	0.53	0.14	0.23	0.86	0.54
7	PBT	38	0.55	0.16	0.25	0.85	0.52
8	CBT	72	0.48	0.13	0.22	0.82	0.49
8	PBT	36	0.52	0.15	0.22	0.79	0.54
A1	CBT	82	0.50	0.15	0.13	0.81	0.49
A1	PBT	42	0.50	0.15	0.24	0.78	0.49
GO	CBT	81	0.48	0.14	0.18	0.76	0.48
GO	PBT	41	0.47	0.15	0.19	0.75	0.47
A2	CBT	84	0.51	0.13	0.20	0.84	0.51
A2	PBT	43	0.50	0.18	0.17	0.97	0.52
M1	CBT	42	0.49	0.13	0.24	0.78	0.48
M1	PBT	42	0.45	0.15	0.18	0.71	0.42
M2	CBT	41	0.40	0.13	0.16	0.67	0.38
M2	PBT	42	0.43	0.17	0.16	0.76	0.38
M3	CBT	40	0.46	0.16	0.20	0.75	0.47
M3	PBT	39	0.44	0.16	0.12	0.74	0.44

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based testing (online); PBT = paper-based testing (paper)

Section 7: Differential Item Functioning

7.1 Overview

Differential item functioning (DIF) analyses were conducted using the data obtained from the operational items. If an item performs differentially across identifiable subgroups (e.g., gender or ethnicity) when students are matched on ability, the item may be measuring something other than the intended construct (i.e., possible evidence of DIF). It is important, however, to recognize that item performance differences flagged for DIF might be related to actual differences in relevant knowledge or skills (item impact) or statistical Type I error. As a result, DIF statistics are used to identify *potential* item bias. Subsequent reviews by content experts and bias/sensitivity committees are required to determine the source and meaning of performance differences.

In this section, DIF statistics are presented for the spring 2018 ELA/L online post-equated tests. In addition, the DIF statistics used at test construction to make decisions about items are provided for pre-equated grades/subjects: mathematics online and paper and ELA/L paper.

7.2 DIF Procedures

Dichotomous Items

The Mantel-Haenszel (MH) DIF statistic was calculated for selected-response items and for dichotomously scored constructed-response items. In this method, test takers are classified to relevant subgroups of interest (e.g., gender or ethnicity). Using the PARCC raw score total as the criteria, test takers in a certain total score category in the focal group (e.g., females) are compared with examinees in the same total score category in the reference group (e.g., males). For each item, test takers in the focal group are also compared to test takers in the reference group who performed equally well on the test as a whole. The common odds ratio is estimated across all categories of matched test taker ability using the following formula (Dorans & Holland, 1993), and the resulting estimate is interpreted as the relative likelihood of success on a particular item for members of two groups when matched on ability.

$$\hat{\alpha}_{MH} = \frac{\sum_{s=1}^{S} R_{rs} W_{fs} / N_{ts}}{\sum_{s=1}^{S} R_{fs} W_{rs} / N_{ts}}, \qquad (7\text{-}1)$$

in which:

S = the number of score categories,
R_{rs} = the number of test takers in the reference group who answer the item correctly,
W_{fs} = the number of test takers in the focal group who answer the item incorrectly,
R_{fs} = the number of test takers in the focal group who answer the item correctly,
W_{rs} = the number of test takers in the reference group who answer the item incorrectly, and
N_{ts} = the total number of test takers.

To facilitate the interpretation of MH results, the common odds ratio is frequently transformed to the delta scale using the following formula (Holland & Thayer, 1988):

$$MH\ D\text{-}DIF = -2.35 \ln (\hat{\alpha}_{MH}) \quad (7\text{-}2)$$

Positive values indicate DIF in favor of the focal group (i.e., positive DIF items are differentially easier for the focal group), whereas negative values indicate DIF in favor of the reference group (i.e., negative DIF items are differentially easier for the reference group).

Polytomous Items

For polytomously scored constructed-response items, the MH D-DIF statistic is not calculated; instead the standardization DIF (Dorans & Schmitt, 1991; Zwick, Thayer & Mazzeo, 1997; Dorans, 2013), in conjunction with the Mantel chi-square statistic (Mantel, 1963; Mantel & Haenszel, 1959), is used to identify items with DIF.

The standardization DIF compares the item means of the two groups after adjusting for differences in the distribution of test takers across the values of the matching variable (i.e., total test score) and is calculated using the following formula:

$$STD - EISDIF = \frac{\sum_{s=1}^{S} N_{fs} * E_f(Y|X=s)}{\sum_{s=1}^{S} N_{fs}} - \frac{\sum_{s=1}^{S} N_{fs} * E_r(Y|X=s)}{\sum_{s=1}^{S} N_{fs}} = \frac{\sum_{s=1}^{S} D_s}{\sum_{s=1}^{S} N_{fs}}, \quad (7\text{-}3)$$

in which:

X= the total score,
Y= the item score,
S = the number of score categories on X,
N_{rs} = the number of test takers in the reference group in score category s,
N_{fs} = the number of test takers in the focal group in score category s,
E_r = the expected item score for reference group, and
E_f = the expected item score for focal group.

A positive *STD-EISDIF* value means that, conditional on the total test score, the focal group has a higher mean item score than the reference group. In contrast, a negative STD-EISDIF value means that, conditional on the total test score, the focal group has a lower mean item score than the reference group.

Classification

Based on the DIF statistics and significance tests, items are classified into three categories and assigned values of A, B, or C (Zieky, 1993). Category A items contain negligible DIF, Category B items exhibit slight to moderate DIF, and Category C items possess moderate to large DIF values. Positive values indicate that, conditional on the total score, the focal group has a higher mean item score than the reference group. In contrast, negative DIF values indicate that, conditional on the total test score, the focal group has a lower mean item score than the reference group. The flagging criteria for dichotomously scored

items are presented in Table 7.1; the flagging criteria for polytomously scored constructed-response items are provided in Table 7.2.

Table 7.1 DIF Categories for Dichotomous Selected-Response and Constructed-Response Items

DIF Category	Criteria
A (negligible)	Absolute value of the MH D-DIF is not significantly different from zero, or is less than one.
B (slight to moderate)	1. Absolute value of the MH D-DIF is significantly different from zero but not from one, and is at least one; or 2. Absolute value of the MH D-DIF is significantly different from one, but is 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 one, and is at least 1.5. Positive values are classified as "C+" and negative values as "C-".

Table 7.2 DIF Categories for Polytomous Constructed-Response Items

DIF Category	Criteria
A (negligible)	Mantel Chi-square p-value > 0.05 or $\lvert STD\text{-}EISDIF/SD \rvert \leq 0.17$
B (slight to moderate)	Mantel Chi-square p-value < 0.05 and $\lvert STD\text{-}EISDIF/SD \rvert > 0.17$
C (moderate to large)	Mantel Chi-square p-value < 0.05 and $\lvert STD\text{-}EISDIF/SD \rvert > 0.25$

Note: *STD-EISDIF* = standardized DIF; SD = total group standard deviation of item score.

7.3 Operational Analysis DIF Comparison Groups

Traditional Comparisons

DIF analyses were conducted on each test form for designated comparison groups defined on the basis of demographic variables including: gender, race/ethnicity, economic disadvantage, and special instructional needs such as students with disabilities (SWD) or English learners (EL). Student demographic information was provided by the states and district and captured in PearsonAccess by means of a student data upload. The demographic data was verified by the states and district prior to score reporting. These comparison groups are specified in Table 7.3.

Table 7.3 Traditional DIF Comparison Groups

Grouping Variable	Focal Group	Reference Group
Gender	Female	Male
Ethnicity	American Indian/Alaska Native (AmerIndian)	White
	Asian	White
	Black or African American	White
	Hispanic/Latino	White
	Native Hawaiian or Pacific Islander	White
	Multiple Race Selected	White
Economic Status*	Economically Disadvantaged (EcnDis)	Not Economically Disadvantaged (NoEcnDis)
Special Instructional Needs	English Learner (ELY)	Non English Learner (ELN)
	Students with Disabilities (SWDY)	Students without Disabilities (SWDN)

Note: * Economic status was based on participation in National School Lunch Program (receipt of free or reduced-price lunch).

Comparison across Languages

DIF analyses were also conducted for Spanish-language items vs. English-language items in mathematics for items that previously were not evaluated for Spanish-language DIF. The purpose of the Spanish vs. English DIF analysis was to evaluate how similarly the items functioned between the two languages because the data from the Spanish-language forms were not separately calibrated using IRT. The item parameter estimates based on the English-speaking test takers were used to generate conversion tables for the Spanish-language forms. Spanish-language mathematics items flagged for C-DIF were reviewed by content specialists and the PARCC Priority Alert Task Force to decide if the items were problematic and should be excluded from scoring. An item could be dropped from a Spanish-language form but remain in the English-language form if no other issues were detected; in those cases, separate conversion tables were generated for the two versions of the form that had different numbers of items.

The Spanish-language forms did not have a non-accommodated English-language form counterpart with the same set of items (refer to Section 2 for more information on the development of Spanish-language forms). Most of the Spanish-language items were previously evaluated for Spanish-language DIF. For items that had not been evaluated for Spanish-language DIF, the analyses were conducted for items that had an English-language item on a non-accommodated test form.

Sample Size Requirement

DIF analyses were conducted when the following sample size requirements were met:

- the smaller group, reference or focal, had at least 100 students, and
- the combined group, reference and focal, had at least 400 students.

Spanish-language vs. English-language DIF analyses were not conducted for Integrated Mathematics I, II, and III because of insufficient sample sizes.

7.4 Operational Differential Item Functioning Results

Appendix 7 presents tables summarizing the DIF results for the spring 2018 ELA/L post-equated online operational items as well as the pre-equated item DIF results that were used to inform decisions at test construction. There is one table prepared for each content and grade level (e.g., ELA/L Grade 3). The fall 2017 forms were based on spring 2017 operational forms. The DIF analyses for these forms are reported in the 2016–2017 Technical Report.

Spoiled or "do not score" items were excluded from the total test score for each form in DIF analysis. These items were removed from scoring because of item performance, technical scoring issues, content concerns, multiple correct answers, or no correct answers. However, the tables in this section may include items for certain grade levels that were excluded from scoring based on later analyses (refer to Section 10.5 Items Excluded from Score Reporting for more information).

In the DIF results tables, the column "DIF Comparisons" identifies the focal and reference groups for the analysis performed; the column "Mode" identifies the test delivery mode. "Total N of Unique Items" reports the number of unique items included in the analysis, whereas "Total N of Item Occurrences" reports the number of times items were used on test forms. An item could be used in multiple test forms; therefore, items were counted according to the occurrences. For example, if the same item appeared in five test forms, it was counted as five occurrences; if this item was classified as B+ on one form and C+ on another form, both occurrences were reported in the corresponding columns. For the Spanish-language DIF, "Total N of Item Occurrences" reports the number of items previously not analyzed for Spanish-language DIF and a non-accommodated English-language item was available in spring 2017. "Total N of Item Occurrences Included in DIF Analysis" reports the number of occurrences with sufficient sample sizes to be included in DIF analyses. In addition, "0" indicates that the DIF analysis did not classify any items in the particular DIF category, while "n/a" indicates that the DIF analysis was not performed due to insufficient sample sizes.

Table 7.4 Differential Item Functioning for ELA/L Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs AmerIndian	**CBT**	42	46	46					46	100				
	PBT	27	27	26			5	19	17	65	4	15		
White vs Asian	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs Black	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs Hispanic	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs Pacific Islander	**CBT**	42	46	46			1	2	44	96	1	2		
	PBT	27	27	0										
White vs Multiracial	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				
NoEcnDis vs EcnDis	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				
ELN vs ELY	**CBT**	42	46	46			1	2	45	98				
	PBT	27	27	26			1	4	25	96				
SWDN vs SWDY	**CBT**	42	46	46					46	100				
	PBT	27	27	26					26	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table 7.5 Differential Item Functioning for Mathematics Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	**CBT**	86	86	86	1	1	5	6	79	92	1	1		
	PBT	42	42	42	1	2			41	98				
White vs AmerIndian	**CBT**	86	86	86			3	3	83	97				
	PBT	42	42	42	1	2	3	7	38	90				
White vs Asian	**CBT**	86	86	86					83	97	3	3		
	PBT	42	42	42					39	93	3	7		
White vs Black	**CBT**	86	86	86			3	3	78	91	4	5	1	1
	PBT	42	42	42					41	98	1	2		
White vs Hispanic	**CBT**	86	86	86			1	1	85	99				
	PBT	42	42	42			1	2	41	98				
White vs Pacific Islander	**CBT**	86	86	86			2	2	82	95	2	2		
	PBT	42	42	0										
White vs Multiracial	**CBT**	86	86	86					86	100				
	PBT	42	42	42					42	100				
NoEcnDis vs EcnDis	**CBT**	86	86	86					86	100				
	PBT	42	42	42			1	2	41	98				
ELN vs ELY	**CBT**	86	86	86			1	1	85	99				
	PBT	42	42	42					42	100				
SWDN vs SWDY	**CBT**	86	86	86			1	1	85	99				
	PBT	42	42	42					42	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability. Small sample sizes may result in fewer items in the column "Total N of Item Occurrences Included in DIF Analysis."

Section 8: Reliability

8.1 Overview

Reliability focuses on the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested rather than fluctuations due to chance. Thus, reliability measures the consistency of the scores across conditions that can be assumed to differ at random, especially which form of the test the test taker is administered and which persons are assigned to score responses to constructed-response questions. In statistical terms, the variance in the distributions of test scores, essentially the differences among individuals, is partly due to real differences in the knowledge, skill, or ability being tested (true variance) and partly due to random errors in the measurement process (error variance). Reliability is an estimate of the proportion of the total variance that is true variance.

There are several different ways of estimating reliability. The type of reliability estimate reported here is an internal-consistency measure, which is derived from analysis of the consistency of the performance of individuals across items within a test. It is used because it serves as a good estimate of alternate forms reliability, but it does not take into account form-to-form variation due to lack of test form parallelism, nor is it responsive to day-to-day variation due to, for example, the student's state of health or the testing environment.

Reliability coefficients range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely students would be to obtain very similar scores upon repeated testing occasions, if the students do not change in their level of the knowledge or skills measured by the test. The reliability estimates in the tables to follow attempt to answer the question, "How consistent would the scores of these students be over replications of the entire testing process?"

Reliability of classification estimates the proportion of students who are accurately classified into proficiency levels. There are two kinds of classification reliability statistics: decision accuracy and decision consistency. Decision accuracy is the agreement between the classifications actually made and the classifications that would be made if the test scores were perfectly reliable. Decision consistency is the agreement between the classifications that would be made on two independent forms of the test.

Another index is inter-rater reliability for the human-scored constructed-response items, which measures the agreement between individual raters (scorers). The inter-rater reliability coefficient answers the question, "How consistent is the scoring such that a set of similarly trained raters would produce similar scores to those obtained?"

Standard error of measurement (SEM) quantifies the amount of error in the test scores. SEM is the extent by which students' scores tend to differ from the scores they would receive if the test were perfectly reliable. As the SEM increases, the variability of students' observed scores is likely to increase across repeated testing. Observed scores with large SEMs pose a challenge to the valid interpretation of a single test score.

Reliability and SEM estimates were calculated at the full assessment level—both paper-based tests (PBTs) and computer-based tests (CBTs)—and at the claim and subclaim levels. In addition, conditional SEMs were calculated and reported in Section 12 and Appendix 12.4.

8.2 Reliability and SEM Estimation

8.2.1 Raw Score Reliability Estimation

Coefficient alpha (Cronbach, 1951), which measures internal consistency reliability, is the most commonly used measure of reliability. Coefficient alpha is estimated by substituting sample estimates for the parameters in the formula below:

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} \sigma_i^{\,2}}{\sigma_x^2}\right], \tag{8-1}$$

where n is the number of items, σ_i^2 is the variance of scores on the *i*-th item, and σ_x^2 is the variance of the total score (sum of scores on the individual items). Other things being equal, the more items a test includes, the higher the internal consistency reliability.

Since PARCC test forms have mixed item types (dichotomous and polytomous items), it is more appropriate to report stratified alpha (Feldt & Brennan, 1989). Stratified alpha is a weighted average of coefficient alphas for item sets with different maximum score points or "strata." Stratified alpha is a reliability estimate computed by dividing the test into parts (strata), computing alpha separately for each part, and using the results to estimate a reliability coefficient for the total score. Stratified alpha is used here because different parts of the test consist of different item types and may measure different skills. The formula for the stratified alpha is:

$$\rho_{strata} = 1 - \frac{\sum_{j=1}^{J} \sigma_{x_j}^2 (1-\alpha_j)}{\sigma_X^2} \tag{8-2}$$

where $\sigma_{X_j}^2$ is the variance for part *j* of the test, σ_X^2 is the variance of the total scores, and α_j is coefficient alpha for part *j* of the test. Estimates of stratified alpha are computed by substituting sample estimates for the parameters in the formula. The average stratified alpha is a weighted average of the stratified alphas across the test forms.

The formula for the standard error of measurement is:

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}} \tag{8-3}$$

where σ_X is the standard deviation of the test raw score and $\rho_{XX'}$ is the reliability estimated by substitution of appropriate statistics for the parameters in equation 8-1 or 8-2.

Cronbach's alpha and stratified alpha coefficients are influenced by test length, test characteristics and sample characteristics (Lord & Novick, 1968; Tavakol & Dennick, 2011; Cortina, 1993). As test length decreases and samples become smaller and more homogeneous, lower estimates of alpha are obtained (Tavakol & Dennick, 2011; Pike & Hudson, 1998). In this section, reliability estimates are reported for overall summative scores, claim scores, and subclaim scores. Estimates are also reported for subgroups for summative scores. A smaller more homogenous sample will likely result in lower stratified alpha estimates. Moderate to acceptable ranges of reliability tend to exceed .5 (Cortina, 1993; Schmitt, 1996). Estimates of reliability lower than 0.50 prompt additional analyses to investigate whether the observed low estimates of alpha are due to restriction in range of the sample. In these cases, the alpha estimates are not a reliable measure of internal consistency. Scale-score reliability estimates which are sample free are provided (Kolen, Zeng, & Hanson, 1996).

A decrease in the number of items may result in a decrease in stratified alpha estimates. The decrease in sample size and the homogeneity of the samples is likely to result in lower stratified alpha estimates. Estimates lower than 0.50 may indicate a lack of internal consistency. Additional analyses investigate whether lower estimates of alpha are due to restriction in range of the sample. In these cases, the alpha estimates are not appropriate measures of internal consistency. As a result, sample-free reliability estimates are also provided such as scale score reliability (Kolen, Zeng, & Hanson, 1996).

8.2.2 Scale Score Reliability Estimation

Like the stratified alpha coefficients, scale score reliability coefficients range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain similar scores upon repeated testing occasions, if the students do not change in their level of the knowledge or skills measured by the test. Because PARCC scale scores are computed from a total score and do not have an item-level component, a stratified alpha coefficient cannot be computed for scale scores. Instead, Kolen, Zeng, and Hanson's (1996) method for scale score reliability was used.

The general formula for a reliability coefficient,

$$\rho = 1 - \frac{\sigma^2(E)}{\sigma^2(X)}, \tag{8-4}$$

involves the error variance, $\sigma^2(E)$, and the total score variance, $\sigma^2(X)$. Using Kolen et al.'s (1996) method, conditional raw score distributions are estimated using Lord and Wingersky's (1984) recursion formula. The conditional raw score distributions are transformed into conditional scale score distributions. Denote *x* as the raw sum score ranging from 0 to *X*, and *s* as a resulting scale score after transformation. The conditional distribution of scale scores is written as $P(X = x|\theta)$. The mean and variance, $\sigma^2[s(X)]$, of this distribution can be computed using these scores and their associated probabilities.

The average error variance of the scale scores is computed as

$$\sigma^2(Error_{scale}) = \int_\theta \sigma^2(s(X)|\theta)\, g(\theta) d\theta, \tag{8-5}$$

where $g(\theta)$ is the ability distribution. The square root of the error variance is the conditional standard error of measurement of the scale scores.

Just as the reliability of raw scores is one minus the ratio of error variance to total variance, the reliability of scale scores is one minus the ratio of the average variance of measurement error for scale scores to the total variance of scale scores,

$$\rho_{scale} = 1 - \frac{\sigma^2(Error_{scale})}{\sigma^2[s(X)]}. \quad (8\text{-}6)$$

The Windows program POLYCSEM (Kolen, 2004) was used to estimate scale score error variance and reliability.

8.3 Reliability Results for Total Group

8.3.1 Raw Score Reliability Results

Tables 8.1 and 8.2 summarize test reliability estimates for the total testing group for English language arts/literacy (ELA/L) and mathematics, respectively. The section includes only spring 2018 results. The fall 2017 results are located in the Addendum.[5] The tables provide the average reliability, which is estimated by averaging the internal consistency estimates computed for all the individual forms of the test and the raw score SEMs, separately for the CBTs and PBTs within each grade level. In addition, the number of forms, the total sample size across all forms, and the average maximum possible score for each set of tests are provided. Estimates were calculated only for groups of 100 or more students administered a specific test form.

English Language Arts/Literacy

The average reliability estimates for the CBT tests for grades 3 through 11 ELA/L range from a low of .90 to a high of .94. The average reliability estimates for the PBT tests for ELA/L grades 3 through 11 range from a low of .88 to a high of .93. The tests for grades 3 through 5 have fewer maximum possible points than for the grades 6 through 11 tests. The average reliability estimates are at least .90 except for grades 3, 4 and 5 PBT tests, which are .88, .89, and .89, respectively. The average raw score SEM is consistently between 5 percent and 6 percent of the maximum possible score.

[5] Addendum 8 provides a summary of reliability information for the fall 2017 administration.

Table 8.1 Summary of ELA/L Test Reliability Estimates for Total Group

Grade Level	Testing Mode	Number of Forms	Avg. Max. Possible Score	Avg. Raw Score SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
3	CBT	4	81	4.59	0.90	8,217	0.84	140,480	0.91
	PBT	1	80	4.90	0.88	40,778	0.88	40,778	0.88
4	CBT	4	106	5.45	0.92	1,035	0.87	144,981	0.92
	PBT	1	106	6.15	0.89	39,738	0.89	39,738	0.89
5	CBT	4	106	5.44	0.92	1,120	0.86	158,843	0.92
	PBT	2	106	5.80	0.89	942	0.85	13,405	0.89
6	CBT	4	109	5.58	0.93	9,681	0.88	158,346	0.93
	PBT	1	109	6.07	0.92	10,859	0.92	10,859	0.92
7	CBT	4	109	5.69	0.93	9,033	0.89	158,668	0.93
	PBT	1	109	6.44	0.91	7,249	0.91	7,249	0.91
8	CBT	4	108	5.79	0.93	8,272	0.90	157,847	0.94
	PBT	1	109	6.21	0.92	7,532	0.92	7,532	0.92
9	CBT	3	109	5.70	0.94	3,409	0.90	59,987	0.94
	PBT	1	109	5.74	0.92	609	0.92	609	0.92
10	CBT	5	109	5.95	0.92	130	0.87	85,684	0.93
	PBT	1	109	6.06	0.93	1,119	0.93	1,119	0.93
11	CBT	3	109	6.03	0.92	1,731	0.86	45,337	0.92
	PBT	1	109	6.17	0.90	865	0.90	865	0.90

Mathematics

The average reliability estimates for the grades 3 through 8 mathematics assessments range from .91 to .94 for the CBT tests and from .91 to .93 for the PBT tests. Most of the average reliability estimates are above .90 except for three PBT tests (Algebra I, Geometry, and Algebra II) and three CBT tests (Integrated Mathematics I, II, and III). The raw score SEM consistently ranges from 4 percent to 6 percent of the maximum score.

In Table 8.2, the Integrated Mathematics I form with a stratified coefficient below .5 has a sample size of 104. The stratified coefficient alpha value less than .5 prompts additional investigations. The sample for this form has a mean scale score of 700.1, standard deviation of 18.4, minimum of 650 and maximum of 739. The form with the maximum reliability has a sample of 572, a mean scale score of 726.8, standard deviation of 30.4, minimum of 650 and maximum of 850. The restriction in range impacts the stratified alpha coefficient. Therefore, other methods for evaluating reliability are considered for this form. Table 8.4 summarizes the scale score reliability for Integrated Mathematics I forms, and for this particular form the scale score reliability is .88. Scale score reliability is sample-free and provides an estimate of measurement precision based on the IRT parameters and an ability distribution.

Table 8.2 Summary of Mathematics Test Reliability Estimates for Total Group

Grade Level	Testing Mode	Number of Forms	Avg. Max. Possible Score	Avg. Raw Score SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
3	CBT	5	66	3.37	0.94	2,070	0.92	33,176	0.94
	PBT	2	65	3.39	0.93	4,405	0.90	37,527	0.93
4	CBT	5	66	3.50	0.93	2,002	0.92	34,021	0.94
	PBT	2	66	3.40	0.93	4,383	0.87	34,426	0.93
5	CBT	5	66	3.58	0.93	1,892	0.89	119,440	0.94
	PBT	2	66	3.63	0.92	2,067	0.86	12,319	0.93
6	CBT	5	66	3.46	0.93	874	0.90	125,724	0.94
	PBT	2	66	3.54	0.92	1,440	0.85	9,154	0.93
7	CBT	5	66	3.37	0.94	722	0.90	30,373	0.94
	PBT	2	66	3.38	0.92	979	0.86	6,059	0.93
8	CBT	5	66	3.17	0.91	674	0.84	55,740	0.92
	PBT	2	66	3.27	0.91	899	0.82	6,055	0.92
A1	CBT	6	81	3.58	0.94	118	0.66	81,445	0.94
	PBT	2	81	3.20	0.88	408	0.84	683	0.91
GO	CBT	4	81	3.46	0.94	7,898	0.92	58,740	0.94
	PBT	1	81	3.11	0.85	544	0.85	544	0.85
A2	CBT	4	81	3.68	0.93	4,424	0.93	45,816	0.94
	PBT	1	81	3.47	0.89	731	0.89	731	0.89
M1	CBT	2	81	3.22	0.83	104	*	572	0.91
	PBT	-	-	-	-	-	-	-	-
M2	CBT	1	80	3.16	0.89	778	0.89	778	0.89
	PBT	-	-	-	-	-	-	-	-
M3	CBT	1	81	2.85	0.77	177	0.77	177	0.77
	PBT	-	-	-	-	-	-	-	-

***** Stratified alpha below 0.50, further investigation summarized at the end of section 8.3.1
Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

8.3.2 Scale Score Reliability Results

Tables 8.3 and 8.4 summarize scale score reliability estimates for the total testing group for English language arts/literacy (ELA/L) and mathematics, respectively, for spring 2018. The tables provide average reliabilities by grade level and mode, which are estimated by averaging the reliability estimates computed for all forms of the test within mode and grade level. In addition, the number of forms, the total sample size across all forms, and the average maximum possible score for each set of tests are provided. Since estimates of scale score reliability are sample independent, all of the form-level results are included.

English Language Arts/Literacy

The average reliability estimates for CBTs for grades 3–11 English language arts/literacy (ELA/L) range from .90 to .92. The average reliability estimates for the PBTs for ELA/L grades 3–11 range from .89 to .92. The average reliability estimates are at least .89. The average scale score SEM ranges from 8.93 to 12.54 for CBTs and between 8.59 and 13.04 for PBTs.

Mathematics

The average scale score reliability estimates for the grades 3–8 mathematics assessments range from .88 to .93 for CBTs and PBTs. For the high school assessments, these quantities range from .84 to .89 and from .85 and .90 for CBTs and PBTs, respectively. For grades 3–8, the scale score SEM ranges from 8.54 to 12.28 for CBTs and from 8.41 to 12.28 for PBTs. For high school tests, the scale score SEM ranges from 9.11 to 13.82 for CBTs and from 9.08 to 14.03 for PBTs.

Table 8.3 Summary of ELA/L Test Scale Score Reliability Estimates for Total Group

Grade Level	Testing Mode	Number of Forms	Avg. Scale Score SEM	Avg. Scale Score Reliability	Min. Scale Score Reliability	Max. Scale Score Reliability
3	CBT	5	12.54	0.90	0.89	0.90
	PBT	1	12.62	0.89	0.89	0.89
4	CBT	5	10.33	0.91	0.90	0.91
	PBT	1	10.39	0.90	0.90	0.90
5	CBT	4	9.77	0.90	0.89	0.91
	PBT	2	10.55	0.89	0.89	0.89
6	CBT	4	8.93	0.91	0.91	0.92
	PBT	1	8.59	0.92	0.92	0.92
7	CBT	4	10.35	0.92	0.91	0.92
	PBT	1	10.03	0.92	0.92	0.92
8	CBT	4	9.98	0.92	0.91	0.93
	PBT	1	9.82	0.92	0.92	0.92
9	CBT	5	9.86	0.92	0.92	0.93
	PBT	1	10.22	0.92	0.92	0.92
10	CBT	5	12.30	0.92	0.91	0.94
	PBT	1	13.04	0.91	0.91	0.91
11	CBT	5	11.67	0.90	0.89	0.91
	PBT	1	12.49	0.89	0.89	0.89

Table 8.4 Summary of Mathematics Test Scale Score Reliability Estimates for Total Group

Grade Level	Testing Mode	Number of Forms	Avg. Scale Score SEM	Avg. Scale Score Reliability	Min. Scale Score Reliability	Max. Scale Score Reliability
3	CBT	5	8.92	0.93	0.93	0.93
	PBT	2	9.04	0.93	0.93	0.93
4	CBT	5	8.54	0.93	0.92	0.93
	PBT	2	8.41	0.93	0.93	0.93
5	CBT	5	8.90	0.92	0.91	0.92
	PBT	2	8.82	0.92	0.92	0.92
6	CBT	5	8.72	0.92	0.91	0.92
	PBT	2	8.86	0.91	0.91	0.92
7	CBT	6	8.65	0.90	0.88	0.92
	PBT	2	8.77	0.90	0.89	0.90
8	CBT	6	12.28	0.88	0.85	0.89
	PBT	2	12.38	0.88	0.86	0.89
A1	CBT	6	11.54	0.89	0.87	0.90
	PBT	2	11.27	0.89	0.88	0.90
GO	CBT	6	9.11	0.89	0.88	0.91
	PBT	2	9.08	0.90	0.89	0.90
A2	CBT	6	12.90	0.89	0.88	0.90
	PBT	2	12.82	0.89	0.89	0.90
M1	CBT	4	11.87	0.88	0.86	0.90
	PBT	2	11.93	0.88	0.88	0.88
M2	CBT	3	12.86	0.84	0.83	0.85
	PBT	2	12.48	0.85	0.85	0.85
M3	CBT	3	13.82	0.88	0.87	0.88
	PBT	2	14.03	0.87	0.87	0.88

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

8.4 Reliability Results for Subgroups of Interest

When the sample size was sufficiently large, raw score reliability and SEM were estimated for the groups identified for DIF analysis. Estimates were calculated only for groups of 100 or more students administered a specific test form.

Tables 8.5 through 8.13 summarize test reliability for groups of interest for ELA/L grades 3 through 11, and Tables 8.14 through 8.25 summarize test reliability for groups of interest for mathematics grades/subjects. Note that reliability estimates are dependent on score variance, and subgroups with smaller variance are likely to have lower reliability estimates than the total group.

Gender

English Language Arts/Literacy

The average reliability estimates and the average SEMs for males and females are similar to the corresponding values for the total group. For most tests, the reliabilities between males and females are equal or within .01 for the majority of tests and modes. For some PBT tests, the differences between the male and female average reliability range from .02 to .04. The SEMs for females are slightly higher than for males for the majority of tests and modes.

Mathematics

As with the ELA/L test components, the average reliability estimates and SEMs for males and females reflect the corresponding reliabilities for the total group. For most tests, the reliabilities between males and females are equal or within .01 for the majority of tests and modes. For some PBT tests, the differences between the male and female average reliability range from .02 to .04. The SEMs for females are slightly higher than for males for the majority of tests and modes.

Ethnicity

English Language Arts/Literacy

The majority of the average reliabilities for the ethnicity groups are .01 to .03 lower than for the total group. There is not a consistent difference among the average reliabilities for white, African American, Asian/Pacific Islander, Hispanic, and multiple-ethnicity students, with the majority of the reliabilities between .87 and .93. However, the average reliabilities for American Indian/Alaskan native students range from .80 to .91. In general, the SEMs are similar to the total group SEMs.

Mathematics

As with the ELA/L reliabilities, the reliabilities for ethnicity groups are marginally lower than for the total group of students. The average SEMs reflect the total group SEMs. While there is variation across tests, the average reliabilities are generally highest for Asian/Pacific Islander students. The American Indian/Alaskan native and the African American groups have the lowest reliabilities.

Special Education Needs

English Language Arts/Literacy

The average reliabilities for five groups of students (economically disadvantaged, not economically disadvantaged, non-English learner, students with disabilities, and students without disabilities) are generally .01 to .02 less than the average reliability for the total group of students. The majority of the average reliabilities range from .87 to .93. The average reliabilities for English learner students are lower, ranging from .77 to .89. The SEMs are generally similar to the total group SEMs.

Mathematics

The average reliabilities for the larger student groups (not economically disadvantaged, non-English learner, and students without disabilities) are generally .01 to .03 less than the average reliability for the total group of students. For economically disadvantaged, English learner, and students with disabilities, the average reliabilities are lower than those for the total group. In general, the SEMs are similar to the total group SEMs.

In Table 8.24 for Integrated Mathematics II, the English Language Learners subgroup has a stratified coefficient alpha value less than .5 which prompts additional investigations. The sample for this form has 114 students, a mean scale score of 700.3, standard deviation of 15.6, minimum of 650 and maximum of 746. The non-English Language Learners has a sample of 664, a mean scale score of 719.7, standard deviation of 32.3, minimum of 650 and maximum of 850. The low performance of the subgroup results in a restriction in range which impacts the stratified alpha coefficient. Therefore, other methods for evaluating reliability are considered for this subgroup. In Table 8.4, the scale score reliability for Integrated Mathematics II forms ranges from .83 to .85. Scale score reliability is sample-free and provides an estimate of measurement precision based on the IRT parameters and an ability distribution.

Students Taking Accommodated Forms

English Language Arts/Literacy

Two of the four accommodation form types (closed caption and text-to-speech) had sufficient sample sizes to allow for estimation of reliability and SEM for grades 3 through 8. Grades 9 and 11 had one of the four forms that had a sufficient sample (text-to-speech). Grade 10 had sufficient sample sizes for three of the four accommodation form types (ASL, closed caption, and text-to-speech). In the lower grades, the ASL and screen reader groups did not have at least 100 students take any specific form. Within grades, the reliabilities of the closed caption forms are similar to the average reliabilities for the total group. For the text-to-speech forms, the reliabilities are somewhat lower than for the total group.

Mathematics

The text-to-speech forms had sufficient sample sizes for reliability and SEM estimation across grades, except for the Integrated Mathematics I, II, and III courses where the sample was not sufficient. ASL forms had more than 100 students for Algebra I. For almost all tests, text-to-speech reliabilities and SEMs are very close to the total group reliabilities. The ASL reliability was lower. The SEMs for accommodated forms were similar to the total group.

Students Taking Translated Forms

Mathematics

With the exception of Integrated Mathematics I, II, and III, there were sufficient numbers of students taking the Spanish-language form for reliability and SEM estimation. The average reliability ranged from .81 to .92 for grades 3 through 8. The average reliability ranged from .59 to .69 for the high school courses. The SEMs are generally lower for the students administered the Spanish-language forms. The moderate to high correlations suggest the translated forms are sufficient for individual student reporting.

Table 8.5 Summary of Test Reliability Estimates for Subgroups: Grade 3 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	81	4.59	0.90	8,217	0.84	140,480	0.91	80	4.90	0.88	40,778	0.88	40,778	0.88
Gender														
Male	81	4.49	0.90	5,386	0.84	70,672	0.90	80	4.78	0.88	20,523	0.88	20,523	0.88
Female	81	4.68	0.90	2,831	0.85	69,808	0.91	80	5.03	0.88	20,255	0.88	20,255	0.88
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	81	4.68	0.89	3,058	0.86	282	0.91	80	5.04	0.87	10,908	0.87	10,908	0.87
African American	81	4.41	0.89	2,031	0.78	25,516	0.90	80	4.73	0.87	11,819	0.87	11,819	0.87
Asian/Pacific Islander	81	4.90	0.88	211	0.86	10,112	0.89	80	5.19	0.87	1,790	0.87	1,790	0.87
Am. Indian/Alaska Native	81	4.32	0.87	1,832	0.87	1,813	0.88	80	4.82	0.82	1,087	0.82	1,087	0.82
Hispanic	81	4.45	0.89	2,629	0.82	37,354	0.90	80	4.89	0.87	14,167	0.87	14,167	0.87
Multiple	81	4.62	0.90	224	0.84	4,999	0.91	80	4.92	0.89	907	0.89	907	0.89
Special Instructional Needs														
Economically Disadvantaged	81	4.41	0.89	5,000	0.80	64,540	0.89	80	4.82	0.87	28,690	0.87	28,690	0.87
Not Economically Disadvantaged	81	4.73	0.89	3,215	0.86	382	0.91	80	5.08	0.86	12,053	0.86	12,053	0.86
English Learner	81	4.22	0.86	1,083	0.77	19,331	0.86	80	4.80	0.84	9,463	0.84	9,463	0.84
Non-English Learner	81	4.64	0.90	7,126	0.85	121,104	0.90	80	4.93	0.89	31,279	0.89	31,279	0.89
Students with Disabilities	81	4.05	0.90	8,217	0.84	18,071	0.91	80	4.29	0.88	5,565	0.88	5,565	0.88
Students w/o Disabilities	81	4.68	0.90	118	0.86	121,967	0.90	80	4.99	0.87	35,050	0.87	35,050	0.87
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	80	4.60	0.90	155	0.90	155	0.90	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	80	3.72	0.84	8,062	0.84	8,062	0.84	-	-	-	-	-	-	-

Table 8.6 Summary of Test Reliability Estimates for Subgroups: Grade 4 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Relia-bility	Min. Sample Size	Min. Relia-bility	Max. Sample Size	Max. Relia-bility	Max. Raw Score	Avg. SEM	Avg. Relia-bility	Min. Sample Size	Min. Relia-bility	Max. Sample Size	Max. Relia-bility
Total Group	106	5.45	0.92	1,035	0.87	144,981	0.92	106	6.15	0.89	39,738	0.89	39,738	0.89
Gender														
Male	106	5.31	0.92	694	0.86	73,149	0.92	106	5.92	0.89	19,916	0.89	19,916	0.89
Female	106	5.57	0.92	341	0.88	71,832	0.92	106	6.37	0.88	19,822	0.88	19,822	0.88
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	106	5.57	0.91	292	0.87	62,915	0.91	106	6.36	0.88	10,233	0.88	10,233	0.88
African American	106	5.21	0.90	288	0.79	26,056	0.91	106	5.90	0.87	11,402	0.87	11,402	0.87
Asian/Pacific Islander	106	5.67	0.91	245	0.91	10,413	0.91	106	6.52	0.89	1,699	0.89	1,699	0.89
Am. Indian/Alaska Native	106	5.14	0.89	1,980	0.89	1,951	0.89	106	6.09	0.82	1,159	0.82	1,159	0.82
Hispanic	106	5.31	0.90	3,076	0.86	38,347	0.91	106	6.11	0.87	14,323	0.87	14,323	0.87
Multiple	106	5.49	0.92	251	0.85	5,014	0.92	106	6.32	0.90	837	0.90	837	0.90
Special Instructional Needs														
Economically Disadvantaged	106	5.24	0.90	5,783	0.83	66,567	0.90	106	6.02	0.86	28,255	0.86	28,255	0.86
Not Economically Disadvantaged	106	5.60	0.91	364	0.89	78,398	0.91	106	6.43	0.88	11,455	0.88	11,455	0.88
English Learner	106	4.97	0.86	1,178	0.77	17,794	0.87	106	5.93	0.83	9,372	0.83	9,372	0.83
Non-English Learner	106	5.50	0.92	784	0.88	127,115	0.92	106	6.22	0.89	30,337	0.89	30,337	0.89
Students with Disabilities	106	4.83	0.91	955	0.85	19,831	0.93	106	5.14	0.87	6,135	0.87	6,135	0.87
Students w/o Disabilities	106	5.56	0.91	124,028	0.91	124,695	0.92	106	6.31	0.87	33,350	0.87	33,350	0.87
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	106	5.36	0.94	232	0.94	232	0.94	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	106	4.51	0.87	9,505	0.87	9,505	0.87	-	-	-	-	-	-	-

Table 8.7 Summary of Test Reliability Estimates for Subgroups: Grade 5 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	106	5.44	0.92	1,120	0.86	158,843	0.92	106	5.80	0.89	942	0.85	13,405	0.89
Gender														
Male	106	5.30	0.92	759	0.85	80,099	0.92	106	5.56	0.88	569	0.84	6,906	0.88
Female	106	5.58	0.92	361	0.87	78,744	0.92	106	6.03	0.88	373	0.86	6,499	0.88
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	106	5.57	0.91	335	0.87	65,480	0.91	106	5.92	0.87	272	0.85	7,020	0.87
African American	106	5.20	0.91	314	0.81	30,199	0.91	106	5.42	0.86	336	0.79	2,337	0.87
Asian/Pacific Islander	106	5.73	0.91	193	0.90	10,755	0.91	106	6.09	0.88	614	0.88	614	0.88
Am. Indian/Alaska Native	106	5.21	0.90	1,958	0.90	1,926	0.90	106	5.93	0.80	990	0.80	990	0.80
Hispanic	106	5.32	0.91	3,242	0.86	45,210	0.91	106	5.65	0.87	283	0.84	1,874	0.87
Multiple	106	5.50	0.92	247	0.86	4,989	0.92	106	5.86	0.89	541	0.89	541	0.89
Special Instructional Needs														
Economically Disadvantaged	106	5.25	0.90	719	0.84	76,657	0.91	106	5.60	0.86	747	0.82	6,521	0.86
Not Economically Disadvantaged	106	5.61	0.91	395	0.88	82,175	0.91	106	5.99	0.87	195	0.87	6,876	0.87
English Learner	106	4.69	0.84	243	0.75	13,415	0.84	106	5.22	0.78	202	0.72	1,038	0.79
Non-English Learner	106	5.50	0.92	876	0.87	145,300	0.92	106	5.85	0.88	738	0.86	12,359	0.89
Students with Disabilities	106	4.86	0.91	1,047	0.83	22,114	0.92	106	4.95	0.87	924	0.85	2,229	0.88
Students w/o Disabilities	106	5.55	0.91	136,758	0.91	136,213	0.92	106	6.02	0.87	11,130	0.87	11,130	0.87
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	106	5.18	0.94	208	0.94	208	0.94	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	106	4.66	0.87	9,981	0.87	9,981	0.87	-	-	-	-	-	-	-

Table 8.8 Summary of Test Reliability Estimates for Subgroups: Grade 6 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	109	5.58	0.93	9,681	0.88	158,346	0.93	109	6.07	0.92	10,859	0.92	10,859	0.93
Gender														
Male	109	5.39	0.93	6,141	0.88	80,007	0.93	109	5.79	0.92	5,553	0.92	5,553	0.92
Female	109	5.77	0.93	3,540	0.89	78,339	0.93	109	6.35	0.91	5,306	0.91	5,306	0.91
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	109	5.73	0.92	3,724	0.90	258	0.93	109	6.14	0.91	5,374	0.91	5,374	0.91
African American	109	5.30	0.91	2,435	0.79	29,424	0.92	109	5.69	0.91	2,017	0.91	2,017	0.91
Asian/Pacific Islander	109	5.81	0.93	219	0.92	10,736	0.93	109	6.29	0.91	512	0.91	512	0.91
Am. Indian/Alaska Native	109	5.44	0.90	1,926	0.90	1,911	0.90	109	6.35	0.86	952	0.86	952	0.86
Hispanic	109	5.44	0.92	3,006	0.86	43,772	0.92	109	5.95	0.91	1,576	0.91	1,576	0.91
Multiple	109	5.65	0.93	232	0.88	4,876	0.93	109	6.03	0.92	403	0.92	403	0.92
Special Instructional Needs														
Economically Disadvantaged	109	5.37	0.91	5,775	0.83	73,341	0.92	109	5.92	0.90	5,527	0.90	5,527	0.90
Not Economically Disadvantaged	109	5.75	0.92	3,879	0.90	290	0.93	109	6.22	0.91	5,277	0.91	5,277	0.91
English Learner	109	4.53	0.83	909	0.66	9,286	0.85	109	5.78	0.85	935	0.85	935	0.85
Non-English Learner	109	5.64	0.93	8,768	0.89	148,830	0.93	109	6.09	0.92	9,908	0.92	9,908	0.92
Students with Disabilities	109	4.87	0.92	9,681	0.88	22,180	0.92	109	5.25	0.91	2,430	0.91	2,430	0.91
Students w/o Disabilities	109	5.71	0.93	136,601	0.92	135,605	0.93	109	6.28	0.90	8,380	0.90	8,380	0.90
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	109	5.36	0.94	241	0.94	241	0.94	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	109	4.54	0.88	9,440	0.88	9,440	0.88	-	-	-	-	-	-	-

Table 8.9 Summary of Test Reliability Estimates for Subgroups: Grade 7 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	109	5.69	0.93	9,033	0.89	158,668	0.93	109	6.44	0.91	7,249	0.91	7,249	0.91
Gender														
Male	109	5.50	0.93	5,843	0.89	79,775	0.93	109	6.10	0.91	3,860	0.91	3,860	0.91
Female	109	5.87	0.93	3,190	0.90	213	0.93	109	6.80	0.90	3,389	0.90	3,389	0.90
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	109	5.79	0.92	3,667	0.90	68,857	0.92	109	6.47	0.91	3,843	0.91	3,843	0.91
African American	109	5.48	0.91	2,311	0.84	28,651	0.92	109	6.10	0.90	1,235	0.90	1,235	0.90
Asian/Pacific Islander	109	5.80	0.93	11,028	0.92	10,977	0.93	109	6.66	0.93	179	0.93	179	0.93
Am. Indian/Alaska Native	109	5.51	0.91	1,912	0.90	1,785	0.91	109	6.63	0.85	708	0.85	708	0.85
Hispanic	109	5.58	0.92	2,640	0.87	43,221	0.92	109	6.17	0.91	1,012	0.91	1,012	0.91
Multiple	109	5.75	0.93	178	0.90	4,658	0.93	109	6.64	0.91	243	0.91	243	0.91
Special Instructional Needs														
Economically Disadvantaged	109	5.52	0.91	5,255	0.85	70,590	0.92	109	6.25	0.89	3,926	0.89	3,926	0.89
Not Economically Disadvantaged	109	5.80	0.93	3,773	0.91	267	0.93	109	6.59	0.90	3,290	0.90	3,290	0.90
English Learner	109	4.59	0.85	729	0.72	8,531	0.86	109	5.86	0.84	608	0.84	608	0.84
Non-English Learner	109	5.74	0.93	8,300	0.89	149,968	0.93	109	6.48	0.91	6,625	0.91	6,625	0.91
Students with Disabilities	109	5.04	0.92	9,033	0.89	22,223	0.93	109	5.36	0.89	1,790	0.89	1,790	0.89
Students w/o Disabilities	109	5.80	0.93	135,256	0.93	135,618	0.93	109	6.74	0.89	5,415	0.89	5,415	0.89
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	109	5.17	0.95	278	0.95	278	0.95	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	109	4.67	0.88	8,755	0.88	8,755	0.88	-	-	-	-	-	-	-

Table 8.10 Summary of Test Reliability Estimates for Subgroups: Grade 8 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	108	5.79	0.93	8,272	0.90	157,847	0.94	109	6.21	0.92	7,532	0.92	7,532	0.92
Gender														
Male	108	5.69	0.93	5,415	0.89	79,693	0.94	109	6.03	0.92	3,956	0.92	3,956	0.92
Female	108	5.88	0.93	2,857	0.90	78,154	0.94	109	6.40	0.92	3,576	0.92	3,576	0.92
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	108	5.90	0.92	3,303	0.91	70,189	0.93	109	6.22	0.92	4,052	0.92	4,052	0.92
African American	108	5.63	0.91	2,095	0.84	231	0.93	109	6.00	0.92	1,396	0.92	1,396	0.92
Asian/Pacific Islander	108	5.79	0.93	186	0.92	11,101	0.94	109	6.28	0.93	162	0.93	162	0.93
Am. Indian/Alaska Native	108	5.67	0.90	1,774	0.89	1,826	0.91	109	6.53	0.88	663	0.88	663	0.88
Hispanic	108	5.68	0.92	2,473	0.88	41,786	0.93	109	6.17	0.92	997	0.92	997	0.92
Multiple	108	5.84	0.93	176	0.91	4,602	0.94	109	6.45	0.91	245	0.91	245	0.91
Special Instructional Needs														
Economically Disadvantaged	108	5.64	0.91	4,662	0.86	68,391	0.93	109	6.19	0.91	4,008	0.91	4,008	0.91
Not Economically Disadvantaged	108	5.90	0.93	3,597	0.91	244	0.94	109	6.23	0.92	3,485	0.92	3,485	0.92
English Learner	108	4.81	0.84	672	0.72	8,009	0.86	109	5.98	0.89	561	0.89	561	0.89
Non-English Learner	108	5.84	0.93	7,595	0.90	149,740	0.94	109	6.23	0.92	6,966	0.92	6,966	0.92
Students with Disabilities	108	5.27	0.92	8,272	0.90	22,475	0.93	109	5.68	0.91	1,746	0.91	1,746	0.91
Students w/o Disabilities	108	5.88	0.93	134,042	0.92	134,891	0.94	109	6.36	0.90	5,749	0.90	5,749	0.90
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	109	5.77	0.94	276	0.94	276	0.94	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	109	5.04	0.89	7,996	0.89	7,996	0.89	-	-	-	-	-	-	-

Table 8.11 Summary of Test Reliability Estimates for Subgroups: Grade 9 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Relia-bility	Min. Sample Size	Min. Relia-bility	Max. Sample Size	Max. Relia-bility	Max. Raw Score	Avg. SEM	Avg. Relia-bility	Min. Sample Size	Min. Relia-bility	Max. Sample Size	Max. Relia-bility
Total Group	109	5.70	0.94	3,409	0.90	59,987	0.94	109	5.74	0.92	609	0.92	609	0.92
Gender														
Male	109	5.52	0.94	2,218	0.90	30,051	0.94	109	5.21	0.90	336	0.90	336	0.90
Female	109	5.86	0.94	1,191	0.90	29,936	0.94	109	6.24	0.92	273	0.92	273	0.92
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	109	5.85	0.93	1,465	0.91	25,090	0.93	109	6.04	0.93	206	0.93	206	0.93
African American	109	5.58	0.93	776	0.85	7,848	0.93	-	-	-	-	-	-	-
Asian/Pacific Islander	109	5.73	0.93	5,188	0.93	5,175	0.93	-	-	-	-	-	-	-
Am. Indian/Alaska Native	109	5.40	0.91	1,467	0.91	1,395	0.91	109	5.70	0.82	169	0.82	169	0.82
Hispanic	109	5.49	0.93	1,000	0.88	19,313	0.93	109	5.62	0.92	145	0.92	145	0.92
Multiple	109	5.81	0.94	814	0.94	849	0.94	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	109	5.48	0.93	1,620	0.86	24,422	0.93	109	5.52	0.89	365	0.89	365	0.89
Not Economically Disadvantaged	109	5.82	0.93	1,789	0.91	35,469	0.94	109	6.02	0.93	242	0.93	242	0.93
English Learner	109	4.40	0.87	3,622	0.87	3,734	0.88	109	5.36	0.81	114	0.81	114	0.81
Non-English Learner	109	5.76	0.94	3,325	0.90	56,136	0.94	109	5.80	0.92	483	0.92	483	0.92
Students with Disabilities	109	5.13	0.93	3,409	0.90	9,182	0.94	109	5.34	0.93	248	0.93	248	0.93
Students w/o Disabilities	109	5.80	0.94	50,492	0.94	50,738	0.94	109	6.02	0.91	345	0.91	345	0.91
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	109	4.64	0.90	3,350	0.90	3,350	0.90	-	-	-	-	-	-	-

Table 8.12 Summary of Test Reliability Estimates for Subgroups: Grade 10 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	109	5.95	0.93	130	0.87	85,684	0.93	109	6.06	0.93	1,119	0.93	1,119	0.93
Gender														
Male	109	5.76	0.93	1,898	0.88	42,933	0.93	109	5.66	0.92	597	0.92	597	0.92
Female	109	6.12	0.92	1,021	0.89	42,751	0.92	109	6.44	0.91	522	0.91	522	0.91
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	109	6.10	0.91	1,214	0.90	35,206	0.92	109	6.18	0.92	498	0.92	498	0.92
African American	109	5.73	0.91	733	0.84	17,063	0.91	109	5.28	0.92	189	0.92	189	0.92
Asian/Pacific Islander	109	6.19	0.92	104	0.90	7,138	0.92	-	-	-	-	-	-	-
Am. Indian/Alaska Native	109	5.66	0.89	1,445	0.89	1,531	0.90	109	6.07	0.82	170	0.82	170	0.82
Hispanic	109	5.74	0.91	793	0.87	22,745	0.92	109	6.13	0.90	200	0.90	200	0.90
Multiple	109	6.07	0.92	1,923	0.92	1,871	0.92	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	109	5.69	0.91	1,323	0.85	30,642	0.91	109	5.81	0.88	469	0.88	469	0.88
Not Economically Disadvantaged	109	6.07	0.92	1,574	0.90	53,705	0.92	109	6.14	0.93	619	0.93	619	0.93
English Learner	109	4.72	0.84	127	0.76	5,305	0.84	109	5.92	0.81	113	0.81	113	0.81
Non-English Learner	109	6.01	0.92	128	0.87	80,173	0.92	109	6.07	0.93	994	0.93	994	0.93
Students with Disabilities	109	5.32	0.92	130	0.87	12,845	0.92	109	5.75	0.93	425	0.93	425	0.93
Students w/o Disabilities	109	6.06	0.92	72,666	0.92	72,718	0.92	109	6.26	0.91	681	0.91	681	0.91
Students Taking Accommodated Forms														
A: ASL	121	4.92	0.87	130	0.87	130	0.87	-	-	-	-	-	-	-
C: Closed Caption	109	5.84	0.94	143	0.94	143	0.94	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	109	4.96	0.87	2,776	0.87	2,776	0.87	-	-	-	-	-	-	-

Table 8.13 Summary of Test Reliability Estimates for Subgroups: Grade 11 ELA/L

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Relia-bility	Min. Sample Size	Min. Relia-bility	Max. Sample Size	Max. Relia-bility	Max. Raw Score	Avg. SEM	Avg. Relia-bility	Min. Sample Size	Min. Relia-bility	Max. Sample Size	Max. Relia-bility
Total Group	109	6.03	0.92	1,731	0.86	45,337	0.92	109	6.17	0.90	865	0.90	865	0.90
Gender														
Male	109	5.76	0.92	1,150	0.86	23,271	0.92	109	5.64	0.89	466	0.89	466	0.89
Female	109	6.30	0.91	581	0.86	22,066	0.91	109	6.73	0.89	399	0.89	399	0.89
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	109	6.13	0.92	783	0.87	16,889	0.92	109	5.95	0.92	242	0.92	242	0.92
African American	109	5.85	0.91	394	0.83	8,052	0.91	109	4.81	0.90	112	0.90	112	0.90
Asian/Pacific Islander	109	6.52	0.92	2,936	0.92	2,935	0.92	-	-	-	-	-	-	-
Am. Indian/Alaska Native	109	5.91	0.89	1,334	0.89	1,337	0.89	109	6.57	0.87	361	0.87	361	0.87
Hispanic	109	5.89	0.91	455	0.85	15,527	0.91	109	6.13	0.87	131	0.87	131	0.87
Multiple	109	6.12	0.92	474	0.91	428	0.93	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	109	5.85	0.91	746	0.86	19,605	0.91	109	6.27	0.89	585	0.89	585	0.89
Not Economically Disadvantaged	109	6.16	0.92	985	0.87	25,726	0.92	109	5.89	0.92	278	0.92	278	0.92
English Learner	109	4.86	0.86	2,760	0.85	2,756	0.86	109	5.67	0.77	109	0.77	109	0.77
Non-English Learner	109	6.09	0.92	1,678	0.86	42,543	0.92	109	6.24	0.90	743	0.90	743	0.90
Students with Disabilities	109	5.17	0.90	1,731	0.86	7,309	0.91	109	5.48	0.91	325	0.91	325	0.91
Students w/o Disabilities	109	6.19	0.91	37,799	0.91	38,006	0.92	109	6.58	0.88	524	0.88	524	0.88
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	109	4.71	0.86	1,671	0.86	1,671	0.86	-	-	-	-	-	-	-

Table 8.14 Summary of Test Reliability Estimates for Subgroups: Grade 3 Mathematics

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	66	3.37	0.94	2,070	0.92	33,176	0.94	65	3.39	0.93	4,405	0.90	37,527	0.93
Gender														
Male	66	3.34	0.94	1,285	0.92	16,168	0.95	65	3.37	0.93	2,686	0.90	18,455	0.94
Female	66	3.40	0.94	785	0.92	14,921	0.94	65	3.41	0.93	1,719	0.90	19,072	0.93
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	66	3.43	0.93	705	0.92	9,510	0.95	65	3.48	0.92	905	0.90	10,291	0.93
African American	66	3.25	0.94	528	0.92	23,135	0.94	65	3.27	0.92	1,078	0.87	10,967	0.92
Asian/Pacific Islander	66	3.35	0.93	7,583	0.92	1,522	0.95	65	3.44	0.93	123	0.91	1,792	0.93
Am. Indian/Alaska Native	66	3.23	0.93	593	0.92	1,378	0.93	65	3.28	0.91	1,081	0.91	1,081	0.91
Hispanic	66	3.33	0.93	655	0.90	30,453	0.93	65	3.38	0.92	2,166	0.90	12,458	0.92
Multiple	66	3.38	0.94	3,815	0.94	779	0.96	65	3.45	0.94	841	0.94	841	0.94
Special Instructional Needs														
Economically Disadvantaged	66	3.29	0.93	1,246	0.92	53,165	0.93	65	3.33	0.92	3,611	0.89	25,866	0.92
Not Economically Disadvantaged	66	3.42	0.93	803	0.93	11,588	0.95	65	3.49	0.92	765	0.91	11,591	0.92
English Learner	66	3.25	0.92	593	0.92	8,124	0.93	65	3.34	0.91	1,872	0.90	8,147	0.91
Non-English Learner	66	3.39	0.94	1,477	0.92	21,070	0.95	65	3.40	0.93	2,532	0.90	29,346	0.94
Students with Disabilities	66	3.18	0.94	1,544	0.91	12,802	0.95	66	3.22	0.91	3,270	0.89	2,431	0.93
Students w/o Disabilities	66	3.40	0.94	526	0.92	21,778	0.94	65	3.41	0.93	1,114	0.91	34,937	0.93
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	66	3.27	0.94	29,244	0.94	30,666	0.94	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	66	3.19	0.92	4,022	0.92	4,022	0.92	66	3.23	0.92	677	0.92	677	0.92

Table 8.15 Summary of Test Reliability Estimates for Subgroups: Grade 4 Mathematics

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	66	3.50	0.93	2,002	0.92	34,021	0.94	66	3.40	0.93	4,383	0.87	34,426	0.93
Gender														
Male	66	3.46	0.94	1,297	0.92	18,864	0.94	66	3.38	0.93	2,599	0.88	16,830	0.94
Female	66	3.53	0.93	705	0.92	15,157	0.94	66	3.42	0.92	1,784	0.85	17,596	0.93
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	66	3.59	0.92	666	0.91	10,367	0.94	66	3.60	0.93	926	0.89	9,134	0.93
African American	66	3.34	0.92	6,370	0.92	20,601	0.93	66	3.19	0.90	1,237	0.82	9,563	0.91
Asian/Pacific Islander	66	3.53	0.92	9,244	0.92	1,496	0.95	66	3.65	0.94	1,640	0.94	1,640	0.94
Am. Indian/Alaska Native	66	3.27	0.92	668	0.91	1,323	0.92	66	3.16	0.90	1,081	0.90	1,081	0.90
Hispanic	66	3.40	0.92	11,443	0.92	27,565	0.92	66	3.38	0.91	2,015	0.86	12,177	0.92
Multiple	66	3.53	0.93	4,703	0.93	790	0.94	66	3.56	0.94	750	0.94	750	0.94
Special Instructional Needs														
Economically Disadvantaged	66	3.36	0.92	18,987	0.92	48,245	0.92	66	3.30	0.91	3,614	0.84	23,880	0.92
Not Economically Disadvantaged	66	3.59	0.93	714	0.92	12,162	0.95	66	3.63	0.93	749	0.91	10,485	0.93
English Learner	66	3.23	0.90	7,743	0.90	599	0.91	66	3.27	0.89	1,710	0.86	7,796	0.90
Non-English Learner	66	3.53	0.93	1,403	0.92	23,187	0.94	66	3.44	0.93	2,668	0.88	26,606	0.94
Students with Disabilities	66	3.18	0.92	13,128	0.91	11,578	0.94	66	3.14	0.89	3,540	0.86	2,529	0.93
Students w/o Disabilities	66	3.55	0.93	475	0.92	20,482	0.94	66	3.45	0.93	820	0.88	31,661	0.93
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	66	3.19	0.94	31,026	0.94	32,156	0.94	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	66	2.94	0.91	3,082	0.91	3,082	0.91	66	3.19	0.87	453	0.87	453	0.87

Table 8.16 Summary of Test Reliability Estimates for Subgroups: Grade 5 Mathematics

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	66	3.58	0.93	1,892	0.89	119,440	0.94	66	3.63	0.92	2,067	0.86	12,319	0.93
Gender														
Male	66	3.54	0.93	1,202	0.90	59,514	0.94	66	3.60	0.92	1,255	0.87	6,239	0.93
Female	66	3.61	0.93	690	0.88	59,926	0.93	66	3.66	0.92	812	0.85	6,080	0.93
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	66	3.67	0.92	658	0.88	11,140	0.94	66	3.74	0.91	703	0.86	6,604	0.92
African American	66	3.39	0.91	508	0.89	20,251	0.92	66	3.35	0.89	647	0.85	2,005	0.90
Asian/Pacific Islander	66	3.63	0.93	9,117	0.91	1,769	0.95	66	3.73	0.94	600	0.94	600	0.94
Am. Indian/Alaska Native	66	3.39	0.91	702	0.90	661	0.92	66	3.46	0.88	958	0.88	958	0.88
Hispanic	66	3.50	0.91	584	0.90	28,850	0.92	66	3.55	0.90	573	0.86	1,626	0.91
Multiple	66	3.59	0.94	4,409	0.93	840	0.94	66	3.72	0.93	497	0.93	497	0.93
Special Instructional Needs														
Economically Disadvantaged	66	3.44	0.91	1,227	0.89	48,529	0.92	66	3.46	0.89	1,563	0.85	5,674	0.90
Not Economically Disadvantaged	66	3.67	0.93	642	0.89	13,148	0.94	66	3.78	0.92	486	0.87	6,601	0.92
English Learner	66	3.18	0.88	462	0.86	6,156	0.89	66	3.26	0.86	493	0.83	826	0.88
Non-English Learner	66	3.61	0.93	1,429	0.90	33,023	0.94	66	3.66	0.92	1,572	0.87	11,482	0.93
Students with Disabilities	66	3.28	0.91	1,486	0.88	11,884	0.93	66	3.25	0.89	1,797	0.86	1,340	0.93
Students w/o Disabilities	66	3.62	0.93	405	0.92	25,315	0.94	66	3.73	0.93	263	0.87	10,934	0.93
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	66	3.40	0.93	41,510	0.93	40,258	0.93	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	66	3.04	0.84	2,815	0.84	2,815	0.84	-	-	-	-	-	-	-

Table 8.17 Summary of Test Reliability Estimates for Subgroups: Grade 6 Mathematics

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	66	3.46	0.93	874	0.90	125,724	0.94	66	3.54	0.92	1,440	0.85	9,154	0.93
Gender														
Male	66	3.42	0.94	552	0.91	62,821	0.94	66	3.48	0.92	883	0.85	4,533	0.93
Female	66	3.50	0.93	322	0.85	62,903	0.93	66	3.59	0.92	557	0.85	4,621	0.93
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	66	3.63	0.93	326	0.90	10,302	0.94	66	3.73	0.92	484	0.88	4,820	0.92
African American	66	3.15	0.91	242	0.79	21,315	0.92	66	3.04	0.87	481	0.80	1,459	0.90
Asian/Pacific Islander	66	3.78	0.94	9,394	0.93	1,473	0.95	66	4.02	0.93	471	0.93	471	0.93
Am. Indian/Alaska Native	66	3.17	0.90	403	0.88	1,426	0.91	66	3.31	0.88	852	0.88	852	0.88
Hispanic	66	3.30	0.91	244	0.89	30,857	0.92	66	3.37	0.88	366	0.76	1,173	0.91
Multiple	66	3.53	0.94	4,292	0.94	700	0.95	66	3.67	0.93	358	0.93	358	0.93
Special Instructional Needs														
Economically Disadvantaged	66	3.23	0.91	531	0.86	51,631	0.91	66	3.21	0.87	1,053	0.76	4,299	0.89
Not Economically Disadvantaged	66	3.65	0.93	326	0.92	11,716	0.94	66	3.85	0.92	381	0.90	4,812	0.93
English Learner	66	2.84	0.87	7,022	0.84	4,859	0.90	66	3.05	0.82	314	0.75	649	0.86
Non-English Learner	66	3.50	0.93	759	0.90	120,719	0.93	66	3.58	0.92	1,124	0.87	8,492	0.93
Students with Disabilities	66	3.02	0.90	14,328	0.87	12,738	0.93	66	2.96	0.89	1,306	0.86	1,078	0.92
Students w/o Disabilities	66	3.54	0.93	20,726	0.93	112,489	0.93	66	3.70	0.93	127	0.79	8,033	0.93
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	66	3.27	0.93	35,762	0.93	34,778	0.93	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	66	2.81	0.82	2,436	0.82	2,436	0.82	-	-	-	-	-	-	-

Table 8.18 Summary of Test Reliability Estimates for Subgroups: Grade 7 Mathematics

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	66	3.37	0.94	722	0.90	30,373	09.4	66	3.38	0.92	979	0.86	6,059	0.93
Gender														
Male	66	3.30	0.94	466	0.90	60,083	0.94	66	3.27	0.92	627	0.86	3,106	0.93
Female	66	3.44	0.93	256	0.89	14,273	0.94	66	3.48	0.92	352	0.85	2,953	0.92
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	66	3.50	0.93	228	0.91	9,401	0.94	66	3.54	0.92	318	0.87	3,453	0.92
African American	66	3.07	0.91	239	0.79	20,014	0.92	66	3.01	0.87	314	0.79	860	0.90
Asian/Pacific Islander	66	3.62	0.94	7,889	0.94	1,174	0.96	66	3.83	0.94	153	0.94	153	0.94
Am. Indian/Alaska Native	66	3.03	0.91	452	0.90	1,303	0.91	66	3.03	0.88	641	0.88	641	0.88
Hispanic	66	3.23	0.92	215	0.91	11,305	0.92	66	3.15	0.89	265	0.83	715	0.91
Multiple	66	3.40	0.94	3,633	0.94	513	0.95	66	3.51	0.93	214	0.93	214	0.93
Special Instructional Needs														
Economically Disadvantaged	66	3.16	0.92	482	0.86	19,793	0.92	66	3.09	0.88	757	0.81	3,034	0.90
Not Economically Disadvantaged	66	3.52	0.94	240	0.93	11,036	0.95	66	3.64	0.92	222	0.90	3,015	0.92
English Learner	66	2.68	0.87	6,062	0.85	4,590	0.89	66	2.70	0.85	213	0.84	383	0.86
Non-English Learner	66	3.41	0.94	622	0.90	25,949	0.94	66	3.43	0.92	765	0.86	5,663	0.93
Students with Disabilities	66	2.84	0.91	632	0.87	13,147	0.93	66	2.75	0.89	881	0.85	861	0.92
Students w/o Disabilities	66	3.46	0.94	106,499	0.94	19,354	0.94	66	3.57	0.92	5,163	0.92	5,163	0.92
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	66	3.08	0.94	30,373	0.94	30,480	0.94	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	66	2.61	0.82	2,391	0.82	2,391	0.82	-	-	-	-	-	-	-

Table 8.19 Summary of Test Reliability Estimates for Subgroups: Grade 8 Mathematics

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	66	3.17	0.91	674	0.84	55,740	0.92	66	3.27	0.91	899	0.82	6,055	0.92
Gender														
Male	66	3.12	0.92	445	0.84	28,417	0.92	66	3.22	0.91	525	0.82	3,135	0.93
Female	66	3.22	0.91	229	0.84	12,025	0.92	66	3.31	0.91	374	0.83	2,920	0.92
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	66	3.32	0.91	238	0.85	7,363	0.92	66	3.44	0.92	315	0.86	3,483	0.92
African American	66	2.89	0.87	227	0.83	7,424	0.88	66	2.92	0.86	294	0.71	985	0.90
Asian/Pacific Islander	66	3.56	0.94	6,753	0.93	843	0.95	66	3.76	0.93	125	0.93	125	0.93
Am. Indian/Alaska Native	66	2.90	0.85	1,652	0.84	370	0.88	66	3.03	0.88	564	0.88	564	0.88
Hispanic	66	3.04	0.89	170	0.82	10,442	0.91	66	3.06	0.86	225	0.75	675	0.90
Multiple	66	3.21	0.92	4,054	0.92	429	0.93	66	3.31	0.93	211	0.93	211	0.93
Special Instructional Needs														
Economically Disadvantaged	66	2.99	0.88	434	0.83	18,327	0.90	66	3.00	0.87	688	0.74	3,019	0.89
Not Economically Disadvantaged	66	3.34	0.92	240	0.86	8,433	0.93	66	3.54	0.92	210	0.89	3,032	0.92
English Learner	66	2.64	0.82	3,363	0.79	2,412	0.85	66	2.75	0.82	149	0.75	380	0.85
Non-English Learner	66	3.21	0.91	585	0.84	23,408	0.92	66	3.31	0.91	750	0.83	5,671	0.92
Students with Disabilities	66	2.74	0.86	586	0.78	7,139	0.90	66	2.70	0.87	805	0.83	812	0.91
Students w/o Disabilities	66	3.26	0.91	119,902	0.91	17,270	0.92	66	3.44	0.92	5,212	0.92	5,212	0.92
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	66	2.91	0.92	27,562	0.92	26,775	0.92	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	66	2.60	0.81	2,042	0.81	2,042	0.81	-	-	-	-	-	-	-

Table 8.20 Summary of Test Reliability Estimates for Subgroups: Algebra I

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	81	3.58	0.94	118	0.66	81,445	0.94	81	3.20	0.88	408	0.84	683	0.91
Gender														
Male	81	3.55	0.95	187	0.89	41,162	0.95	81	3.10	0.87	233	0.80	404	0.91
Female	81	3.61	0.94	6,286	0.93	40,283	0.94	81	3.32	0.90	175	0.87	279	0.91
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	81	3.78	0.93	4,340	0.92	34,404	0.93	81	3.73	0.92	251	0.92	251	0.92
African American	81	3.21	0.90	103	0.52	15,778	0.90	81	2.69	0.82	123	0.82	123	0.82
Asian/Pacific Islander	81	4.02	0.95	7,469	0.95	699	0.95	-	-	-	-	-	-	-
Am. Indian/Alaska Native	81	3.05	0.86	181	0.83	174	0.89	81	2.92	0.66	140	0.66	140	0.66
Hispanic	81	3.29	0.91	6,292	0.87	20,593	0.91	81	3.00	0.76	330	0.71	132	0.88
Multiple	81	3.72	0.94	197	0.93	177	0.95	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	81	3.26	0.91	7,429	0.89	29,118	0.91	81	2.96	0.78	260	0.71	371	0.83
Not Economically Disadvantaged	81	3.75	0.94	200	0.91	51,037	0.94	81	3.47	0.92	147	0.90	285	0.93
English Learner	81	2.80	0.85	3,987	0.75	3,859	0.90	81	2.93	0.67	326	0.67	326	0.67
Non-English Learner	81	3.62	0.94	114	0.67	77,534	0.94	81	3.39	0.92	564	0.92	564	0.92
Students with Disabilities	81	3.10	0.91	118	0.66	12,014	0.93	81	3.05	0.89	274	0.89	274	0.89
Students w/o Disabilities	81	3.66	0.94	9,646	0.94	69,307	0.94	81	3.23	0.81	317	0.68	388	0.92
Students Taking Accommodated Forms														
A: ASL	81	2.84	0.66	118	0.66	118	0.66	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	81	3.40	0.94	11,901	0.93	10,480	0.94	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	81	2.66	0.66	2,588	0.66	2,588	0.66	81	2.93	0.67	319	0.67	319	0.67

Table 8.21 Summary of Test Reliability Estimates for Subgroups: Geometry

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	81	3.46	0.94	7,898	0.92	58,740	0.94	81	3.11	0.84	544	0.85	544	0.85
Gender														
Male	81	3.44	0.94	4,295	0.92	29,288	0.95	81	2.91	0.82	299	0.82	299	0.82
Female	81	3.48	0.93	3,603	0.91	29,452	0.94	81	3.32	0.86	245	0.86	245	0.86
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	81	3.67	0.93	2,178	0.92	26,009	0.93	81	3.33	0.91	138	0.91	138	0.91
African American	81	3.00	0.89	1,308	0.87	7,574	0.90	-	-	-	-	-	-	-
Asian/Pacific Islander	81	4.12	0.95	347	0.94	5,702	0.95	-	-	-	-	-	-	-
Am. Indian/Alaska Native	81	2.95	0.88	124	0.83	1,382	0.89	81	3.10	0.71	198	0.71	198	0.71
Hispanic	81	3.07	0.89	3,874	0.85	16,809	0.90	81	3.02	0.85	128	0.85	128	0.85
Multiple	81	3.75	0.95	802	0.95	1,065	0.95	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	81	3.03	0.89	4,427	0.85	20,847	0.89	81	3.08	0.83	354	0.83	354	0.83
Not Economically Disadvantaged	81	3.69	0.94	3,462	0.93	36,472	0.95	81	3.16	0.90	156	0.90	156	0.90
English Learner	81	2.66	0.81	2,377	0.73	1,834	0.86	-	-	-	-	-	-	-
Non-English Learner	81	3.51	0.94	5,504	0.92	55,978	0.94	81	3.13	0.87	442	0.87	442	0.87
Students with Disabilities	81	2.89	0.90	2,241	0.84	6,450	0.92	81	2.93	0.89	212	0.89	212	0.89
Students w/o Disabilities	81	3.56	0.94	5,655	0.92	50,071	0.94	81	3.22	0.81	321	0.81	321	0.81
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	81	3.17	0.93	6,381	0.92	5,946	0.93	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	81	2.58	0.69	1,912	0.69	1,912	0.69	-	-	-	-	-	-	-

Table 8.22 Summary of Test Reliability Estimates for Subgroups: Algebra II

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	81	3.68	0.93	4,424	0.93	45,816	0.94	81	3.47	0.89	731	0.89	731	0.89
Gender														
Male	81	3.65	0.94	2,748	0.93	22,489	0.95	81	3.34	0.90	370	0.90	370	0.90
Female	81	3.70	0.93	2,089	0.92	23,327	0.93	81	3.57	0.87	361	0.87	361	0.87
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	81	3.86	0.93	1,856	0.92	20,532	0.93	81	3.74	0.90	205	0.90	205	0.90
African American	81	3.27	0.88	8,131	0.87	763	0.89	-	-	-	-	-	-	-
Asian/Pacific Islander	81	4.20	0.94	289	0.94	4,754	0.95	-	-	-	-	-	-	-
Am. Indian/Alaska Native	81	3.18	0.82	105	0.59	1,078	0.85	81	3.35	0.86	292	0.86	292	0.86
Hispanic	81	3.34	0.88	2,214	0.86	11,980	0.89	81	3.21	0.88	131	0.88	131	0.88
Multiple	81	3.85	0.94	983	0.93	835	0.95	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	81	3.30	0.88	2,493	0.87	14,850	0.89	81	3.29	0.87	453	0.87	453	0.87
Not Economically Disadvantaged	81	3.86	0.94	2,493	0.93	30,098	0.94	81	3.70	0.89	255	0.89	255	0.89
English Learner	81	2.83	0.78	1,224	0.64	1,351	0.88							
Non-English Learner	81	3.71	0.93	4,115	0.93	44,434	0.94	81	3.53	0.89	628	0.89	628	0.89
Students with Disabilities	81	3.09	0.89	1,363	0.81	5,384	0.91	81	3.18	0.83	210	0.83	210	0.83
Students w/o Disabilities	81	3.75	0.93	3,205	0.93	40,389	0.94	81	3.56	0.89	507	0.89	507	0.89
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	81	3.32	0.93	4,424	0.93	4,486	0.93	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	81	2.57	0.59	894	0.59	894	0.59	-	-	-	-	-	-	-

Table 8.23 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics I

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	81	3.22	0.83	104	0.38	572	0.91	-	-	-	-	-	-	-
Gender														
Male	81	3.31	0.92	309	0.92	309	0.92	-	-	-	-	-	-	-
Female	81	3.36	0.90	263	0.90	263	0.90	-	-	-	-	-	-	-
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	81	3.67	0.93	150	0.93	150	0.93	-	-	-	-	-	-	-
African American	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Asian/Pacific Islander	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Am. Indian/Alaska Native	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hispanic	81	3.18	0.87	368	0.87	368	0.87	-	-	-	-	-	-	-
Multiple	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	81	3.20	0.89	373	0.89	373	0.89	-	-	-	-	-	-	-
Not Economically Disadvantaged	81	3.54	0.92	199	0.92	199	0.92	-	-	-	-	-	-	-
English Learner	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Non-English Learner	81	3.37	0.91	498	0.91	498	0.91	-	-	-	-	-	-	-
Students with Disabilities	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Students w/o Disabilities	81	3.35	0.90	515	0.90	515	0.90	-	-	-	-	-	-	-
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 8.24 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics II

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	80	3.16	0.89	778	0.89	778	0.89	-	-	-	-	-	-	-
Gender														
Male	80	3.20	0.91	380	0.91	380	0.91	-	-	-	-	-	-	-
Female	80	3.12	0.88	398	0.88	398	0.88	-	-	-	-	-	-	-
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	80	3.66	0.93	171	0.93	171	0.93	-	-	-	-	-	-	-
African American	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Asian/Pacific Islander	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Am. Indian/Alaska Native	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hispanic	80	2.92	0.81	489	0.81	489	0.81	-	-	-	-	-	-	-
Multiple	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	80	2.90	0.68	465	0.68	465	0.68	-	-	-	-	-	-	-
Not Economically Disadvantaged	80	3.50	0.93	313	0.93	313	0.93	-	-	-	-	-	-	-
English Learner	80	2.73	*	-	-	-	-	-	-	-	-	-	-	-
Non-English Learner	80	3.22	0.90	664	0.90	664	0.90	-	-	-	-	-	-	-
Students with Disabilities	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Students w/o Disabilities	80	3.18	0.89	706	0.89	706	0.89	-	-	-	-	-	-	-
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	-	-	-	-	-	-	-	-	-	-	-	-	-	-

***** Cronbach alpha below 0.50, further investigation summarized at the end of section 8.4

Table 8.25 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics III

	CBT							PBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Total Group	81	2.85	0.77	177	0.77	177	0.77	-	-	-	-	-	-	-
Gender														
Male	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Female	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Unknown/Missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ethnicity														
White	-	-	-	-	-	-	-	-	-	-	-	-	-	-
African American	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Asian/Pacific Islander	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Am. Indian/Alaska Native	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hispanic	81	2.75	0.57	144	0.57	144	0.57	-	-	-	-	-	-	-
Multiple	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Special Instructional Needs														
Economically Disadvantaged	81	2.84	0.77	172	0.77	172	0.77	-	-	-	-	-	-	-
Not Economically Disadvantaged	-	-	-	-	-	-	-	-	-	-	-	-	-	-
English Learner	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Non-English Learner	81	2.90	0.79	154	0.79	154	0.79	-	-	-	-	-	-	-
Students with Disabilities	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Students w/o Disabilities	81	2.86	0.70	168	0.70	168	0.70	-	-	-	-	-	-	-
Students Taking Accommodated Forms														
A: ASL	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C: Closed Caption	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T: Text-to-Speech	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Students Taking Translated Forms														
Spanish-Language Form	-	-	-	-	-	-	-	-	-	-	-	-	-	-

8.5 Reliability Results for English Language Arts/Literacy Claims and Subclaims

PARCC developed subclaims in addition to major claims based on the Common Core State Standards. English language arts/literacy (ELA/L) has two major claims relating to Reading Complex Text and Writing. The major claim for Reading Complex Text is that students read and comprehend a range of sufficiently complex texts independently. The major claim for Writing is that students write effectively when using and/or analyzing sources. Refer to Table 8.26 for a summary of the ELA/L claims and subclaims.

Table 8.26 Descriptions of ELA/L Claims and Subclaims

English Language Arts/Literacy		
Major Claim	**Subclaim**	**Description**
Reading	Reading Literature	Students demonstrate comprehension and draw evidence from readings of grade-level, complex literary text.
Reading	Reading Information	Students demonstrate comprehension and draw evidence from readings of grade-level, complex informational text.
Reading	Reading Vocabulary	Students use context to determine the meaning of words and phrases.
Writing	Writing Written Expression	Students produce clear and coherent writing in which the development, organization, and style are appropriate to the task, purpose, and audience.
Writing	Writing Knowledge Language and Conventions	Students demonstrate knowledge of conventions and other important elements of language.

Reliability indices were calculated for each major claim and subclaim. Table 8.27 presents the average reliability estimates for all forms of the test at the specified grade and testing mode for the ELA/L tests. In order to assist in understanding the reliability estimates, the average maximum number of points for each major claim and subclaim are also provided.

The average reliabilities for the Reading Complex Text claim for grades 3 through 11 range from .86 to .90 for CBT and from .85 to .91 for PBT. The Writing claim average reliabilities are generally lower than those for the Reading claim. The average reliability across all grades ranged from .78 to .90 for CBT and ranged from .70 to .88 for PBT. The reliabilities for the Writing claim for grades 3 and 4 are based on 27 points and 33 points, respectively, and the average reliabilities for the grades 5 through 11 Writing claims are based on 36 points.

The average reliabilities of the Reading Literature subclaim scores over testing modes and grades have a median of .74, and the reliabilities vary from .63 to .86. For three grades, the Reading Literature subclaim was based on less than 25 points. The reliabilities for these ranged from .69 to .72. The reliabilities for tests with 25 or more points ranged from .63 to .86. The average reliabilities of the Reading Information subclaim scores over testing modes and grades have a median of .70, and the

reliabilities vary from .59 to .80. The number of points ranged from 13 to 28 with a median of 23. The average reliabilities of the Reading Vocabulary subclaim scores over testing modes and grades have a median of .65, and the reliabilities vary from .58 to .75. The number of points ranged from 10 to 18 with a median of 15.

The Writing Written Expression subclaim is based on 27 points for grades 3 and 33 points for grades 4 and 5. Grades 6 through 11 are based on 36 points. The median of the average reliabilities for the tests was .84 and the average reliabilities ranged from .7 to .9. The Writing Knowledge of Language and Conventions subclaims are all based on nine points. The median average reliability was .87 and ranged from .78 to .91.

Table 8.27 Average ELA/L Reliability Estimates for Total Test and Subscores

		Reading: Total		Reading: Literature		Reading: Information		Reading: Vocabulary		Writing: Total		Writing: Written Expression		Writing: Knowledge Language and Conventions	
Grade Level	Testing Mode	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability
3	CBT	45	0.86	19	0.72	16	0.65	10	0.68	36	0.83	27	0.78	9	0.85
	PBT	44	0.86	19	0.71	13	0.59	12	0.75	36	0.79	27	0.70	9	0.78
4	CBT	64	0.88	31	0.79	20	0.67	13	0.61	42	0.85	33	0.83	9	0.88
	PBT	64	0.87	26	0.71	22	0.68	16	0.68	42	0.81	33	0.74	9	0.80
5	CBT	64	0.88	26	0.73	22	0.67	16	0.73	42	0.87	33	0.84	9	0.89
	PBT	64	0.85	24	0.63	22	0.61	18	0.72	42	0.82	33	0.76	9	0.80
6	CBT	64	0.90	26	0.82	25	0.72	13	0.65	45	0.87	36	0.86	9	0.89
	PBT	64	0.91	34	0.86	14	0.62	16	0.72	45	0.84	36	0.80	9	0.84
7	CBT	64	0.88	26	0.72	26	0.77	12	0.58	45	0.88	36	0.87	9	0.91
	PBT	64	0.89	26	0.78	22	0.73	16	0.65	45	0.83	36	0.79	9	0.84
8	CBT	63	0.88	21	0.73	27	0.77	15	0.62	45	0.88	36	0.89	9	0.90
	PBT	64	0.90	32	0.82	14	0.65	18	0.71	45	0.86	36	0.85	9	0.87
9	CBT	64	0.90	25	0.79	24	0.80	15	0.65	45	0.89	36	0.90	9	0.91
	PBT	64	0.90	26	0.78	24	0.78	14	0.65	45	0.86	36	0.82	9	0.84
10	CBT	64	0.87	25	0.72	23	0.74	16	0.59	45	0.88	36	0.88	9	0.89
	PBT	64	0.88	26	0.76	24	0.75	14	0.60	45	0.88	36	0.88	9	0.89
11	CBT	64	0.88	25	0.74	27	0.79	12	0.59	45	0.87	36	0.85	9	0.86
	PBT	64	0.85	18	0.69	28	0.68	18	0.61	45	0.85	36	0.82	9	0.84

8.6 Reliability Results for Mathematics Subclaims

For mathematics, there are four subclaims related to the major claim that students are on track or ready for college and careers:

- Subclaim A: Students solve problems involving the *major content* for their grade level with connections to the Standards for Mathematical Practice.
- Subclaim B: Students solve problems involving the *additional and supporting content* for their grade level with connections to the Standards for Mathematical Practice.
- Subclaim C: Students express grade/course-level appropriate *mathematical reasoning* by constructing viable mathematical arguments and critiquing the reasoning of others, and/or attending to precision when making mathematical statements.
- Subclaim D: Students solve real-world problems with a degree of difficulty appropriate to the grade/course by applying knowledge and skills articulated in the standards and by engaging particularly in the *modeling practice*.

Reliability estimates were calculated for each subclaim for mathematics. Table 8.28 presents the average reliability estimates for mathematics subclaims by mode (CBT and PBT) and grade/subject. The sample sizes for Integrated Mathematics I, II, and III PBT were not sufficient for reliability analyses.

Mathematics Grades 3–8, Algebra I, Geometry, and Algebra II
Subclaims with greater numbers of points tend to have greater reliability estimates. The Major Content subclaim has the largest number of points for each assessment and, accordingly, has higher average reliabilities than the other three subclaims. For grades 3 through 8, Algebra I, Geometry, and Algebra II, the median of the average reliabilities for the Major Content subclaim is .85, with a range from .68 to .89. The Major Content ranged in number of points from 24 to 31 with a median of 27.5.

The median of the average reliabilities for the Additional and Supporting Content subclaim for grades 3 through 8, Algebra I, Geometry, and Algebra II is .67, with a range from .50 to .76. The Additional and Supporting Content subclaim ranged in number of points from 9 to 20 with a median of 13.

Due to the similarity in the number of items representing the Mathematics Reasoning subclaim across grades and courses, the reliabilities for Mathematics Reasoning are less variable than those for the other subclaims, and range from .61 to .77. The median of the average reliabilities for this subclaim is .71. The number of points is 14 for all grades and courses.

For the Modeling Practice subclaim, the median of the average reliabilities is .67 and range from .58 to .78. The number of points is 12 for grades 3 through 8 and 18 for all high school courses.

Integrated Mathematics I, II, and III
PBT reliabilities are not provided for the Integrated Mathematics assessments due to small sample sizes. For CBT, Integrated Mathematics I and II have moderate to fair average reliabilities across the subclaims ranging from .56 to .73. Integrated Mathematics III had low to moderate average reliabilities.

In Table 8.28 for Integrated Mathematics III, two subclaims have stratified coefficient alpha values less than .50 which prompts additional investigations. The sample consists of 177 students with a mean scale

score of 689.08 and a standard deviation of 26.85 (see Appendix Table A.12.47). This is an intact reused form. Reliability is reported for this same form in Table 8.28 of the Technical Report for 2017 with reliability of .83 for the Major Content subclaim and reliability of .73 for the Additional and Supporting Content subclaim. For spring 2017, 568 students reported a mean scale score of 722.72 with a standard deviation of 44.51 (see Appendix Table A.12.47 in the Technical Report for 2017). The spring 2018 sample results in a restriction in range which impacts the stratified alpha coefficient. Scale score reliability is not calculated for the subclaims; however, the conditional standard error of measurement (CSEM) on the theta scale is available and provides an estimate of measurement precision for each subclaim. For the Major Content subclaim the CSEM ranges from .36 to 1.78. For the Additional and Supporting Content subclaim the CSEM ranges from .45 to 2.44. The CSEM for the lowest obtainable theta value and the highest obtainable theta value were not included in the ranges.

Table 8.28 Average Mathematics Reliability Estimates for Total Test and Subscores

Grade Level	Testing Mode	Major Content		Additional & Supporting Content		Mathematics Reasoning		Modeling Practice	
		Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability
3	CBT	28	0.89	12	0.75	14	0.70	12	0.71
	PBT	27	0.87	12	0.72	14	0.67	12	0.67
4	CBT	31	0.88	9	0.69	14	0.74	12	0.60
	PBT	31	0.86	9	0.67	14	0.74	12	0.69
5	CBT	30	0.87	10	0.62	14	0.76	12	0.65
	PBT	30	0.86	10	0.57	14	0.68	12	0.66
6	CBT	26	0.87	14	0.70	14	0.72	12	0.69
	PBT	26	0.83	14	0.71	14	0.66	12	0.68
7	CBT	29	0.88	11	0.66	14	0.76	12	0.63
	PBT	29	0.85	11	0.54	14	0.72	12	0.63
8	CBT	27	0.82	13	0.67	14	0.68	12	0.60
	PBT	27	0.81	13	0.64	14	0.70	12	0.64
A1	CBT	26	0.84	15	0.72	14	0.72	18	0.78
	PBT	26	0.69	15	0.54	14	0.61	18	0.70
GO	CBT	30	0.85	19	0.76	14	0.77	18	0.74
	PBT	30	0.68	19	0.50	14	0.62	18	0.59
A2	CBT	24	0.84	20	0.75	14	0.73	18	0.70
	PBT	27	0.68	20	0.67	14	0.66	18	0.58
M1	CBT	27	0.70	20	0.59	14	0.64	18	0.56
	PBT	-	-	-	-	-	-	-	-
M2	CBT	31	0.73	17	0.62	14	0.61	18	0.70
	PBT	-	-	-	-	-	-	-	-
M3	CBT	26	*	23	*	14	0.52	18	0.60
	PBT	-	-	-	-	-	-	-	-

***** Cronbach alpha below 0.50, further investigation summarized at the end of section 8.6

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. Integrated Mathematics I and III PBT had insufficient sample sizes.

8.7 Reliability of Classification

The reliability of the classifications for the test takers was calculated using the computer program BB-CLASS (Brennan, 2004), which operationalizes a statistical method developed by Livingston and Lewis (1993, 1995). As Livingston and Lewis (1993, 1995) explain, this method uses information from the administration of one test form (i.e., distribution of scores, the minimum and maximum possible scores, the cut points used for classification, and the reliability coefficient) to estimate two kinds of statistics, decision accuracy and decision consistency. Decision accuracy refers to the extent to which the classifications of test takers based on their scores on the test form agree with the classifications made on the basis of the classifications that would be made if the test scores were perfectly reliable. Decision consistency refers to the agreement between these classifications based on two non-overlapping, equally difficult forms of the test.

Decision consistency values are always lower than the corresponding decision accuracy values, because in decision consistency, both of the classifications are subject to measurement error. In decision accuracy, only one of the classifications is based on a score that contains error. It is not possible to know which students were accurately classified, but it is possible to estimate the proportion of the students who were accurately classified. Similarly, it is not possible to know which students would be consistently classified if they were retested with another form, but it is possible to estimate the proportion of the students who would be consistently classified.

English Language Arts/Literacy

Table 8.29 provides information about the accuracy and the consistency of two types of classifications made on the basis of the summative scale scores on the grades 3 through 11 ELA/L assessments. The columns labeled “Exact level” provide the estimates of the indices based on classifications of students into one of five performance levels. The columns labeled “Level 4 or higher vs. 3 or lower” provide the estimates of the indices based on classifications of students as being either in one of the upper two levels (Levels 4 and 5) or in one of the lower three levels (Levels 1, 2, and 3). Performance Level 4 is considered the College and Career Readiness standard on the PARCC assessments.

The table shows that for classifying each student into one of the five performance levels, the proportion accurately classified ranges from .70 to .77 with a median of .74; the proportion who would be consistently classified on two different test forms ranges from .59 to .68 with a median of .64. For classifying each student as being at Level 4 or higher vs. being at Level 3 or lower, the proportion accurately classified ranges from .90 to .93 with a median of .92; the proportion who would be consistently classified this way on two different test forms ranges from .86 to .91 with a median of .88.

Table 8.29 Reliability of Classification: Summary for ELA/L

Grade Level	Testing Mode	Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
		Exact level	Level 4 or higher vs. 3 or lower	Exact level	Level 4 or higher vs. 3 or lower
3	CBT	0.72	0.91	0.62	0.87
	PBT	0.70	0.90	0.60	0.86
4	CBT	0.74	0.91	0.64	0.88
	PBT	0.70	0.90	0.59	0.86
5	CBT	0.76	0.91	0.67	0.88
	PBT	0.74	0.90	0.64	0.86
6	CBT	0.77	0.92	0.68	0.89
	PBT	0.76	0.91	0.67	0.88
7	CBT	0.74	0.92	0.64	0.89
	PBT	0.72	0.91	0.62	0.88
8	CBT	0.75	0.92	0.66	0.89
	PBT	0.74	0.92	0.64	0.88
9	CBT	0.77	0.93	0.68	0.90
	PBT	0.74	0.93	0.65	0.91
10	CBT	0.73	0.92	0.63	0.89
	PBT	0.72	0.92	0.63	0.89
11	CBT	0.73	0.92	0.63	0.88
	PBT	0.70	0.91	0.60	0.87

Table 8.30 provides more detailed information about the accuracy and the consistency of the classification of students into performance levels for ELA/L grade 3. Each cell in the 5-by-5 table shows the estimated proportion of students who would be classified into a particular combination of performance levels. The sum of the five bold values on the diagonal should equal the level of decision accuracy or consistency presented in Table 8.29. For "Level 4 and higher vs. 3 and lower" found in Table 8.29, the sum of the shaded values in Table 8.30 should equal the level of decision accuracy or consistency presented in Table 8.29. Note that the sums based on values in Table 8.30 may not match exactly to the values in Table 8.29 due to truncation and rounding.

Detailed information for all ELA/L spring results are provided in Appendix 8 Tables A.8.1 through A.8.9. Fall block results for ELA/L grades 9 through 11 are provided in the Addendum. The structure of these tables is the same as that of Table 8.30 and the values in the tables should be interpreted in the same manner.

Table 8.30 Reliability of Classification: Grade 3 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.16**	0.03	0.00	0.00	0.00	0.19
		700 – 724	0.04	**0.10**	0.05	0.00	0.00	0.18
		725 – 749	0.00	0.04	**0.13**	0.04	0.00	0.22
		750 – 809	0.00	0.00	0.05	**0.30**	0.03	0.38
		810 – 850	0.00	0.00	0.00	0.01	**0.02**	0.03
	Decision Consistency	650 – 699	**0.15**	0.04	0.01	0.00	0.00	0.20
		700 – 724	0.04	**0.08**	0.06	0.01	0.00	0.18
		725 – 749	0.01	0.04	**0.10**	0.05	0.00	0.20
		750 – 809	0.00	0.01	0.06	**0.27**	0.02	0.37
		810 – 850	0.00	0.00	0.00	0.02	**0.02**	0.05
PBT	Decision Accuracy	650 – 699	**0.15**	0.03	0.00	0.00	0.00	0.18
		700 – 724	0.04	**0.09**	0.05	0.00	0.00	0.18
		725 – 749	0.00	0.04	**0.13**	0.05	0.00	0.22
		750 – 809	0.00	0.00	0.05	**0.31**	0.03	0.40
		810 – 850	0.00	0.00	0.00	0.01	**0.01**	0.02
	Decision Consistency	650 – 699	**0.14**	0.04	0.01	0.00	0.00	0.19
		700 – 724	0.04	**0.07**	0.06	0.01	0.00	0.18
		725 – 749	0.01	0.04	**0.09**	0.06	0.00	0.21
		750 – 809	0.00	0.01	0.07	**0.27**	0.03	0.38
		810 – 850	0.00	0.00	0.00	0.02	**0.02**	0.04

Note: This table includes the same information as Table A.8.1. The sum of the five bold values on the diagonal should equal the level of decision accuracy or consistency presented in Table 8.28. For "Level 4 and higher vs. 3 and lower" presented in Table 8.27, the sum of the shaded values in Table 8.29 should equal the level of decision accuracy or consistency presented in Table 8.27. Any differences between the sums based on values in Table 8.28 and the values in Table 8.29 are due to truncation and rounding.

Mathematics

Table 8.31 provides information about the accuracy and the consistency of two types of classifications made on the basis of the summative scale scores on the mathematics assessments. For the grades 3 through 8 mathematics tests, the table shows that for classifying each student into one of the five performance levels, the proportion accurately classified ranges from .73 to .80 with a median of .77; the proportion who would be consistently classified on two different test forms ranges from .64 to .72 with a median of .68. For the six high school mathematics courses, the table shows that for classifying each student into one of the five performance levels, the proportion accurately classified ranges from .67 to .81 with a median of .74; the proportion who would be consistently classified on two different test forms ranges from .56 to .73 with a median of .65.

For classifying each student as being at Level 4 or higher vs. being at Level 3 or lower, for the grades 3 through 8 mathematics tests, the proportion accurately classified ranges from .92 to .94 with a median of .92; the proportion who would be consistently classified on two different test forms ranges from .89 to .91 with a median of .89. For the six high school mathematics courses, the proportion accurately classified as being at Level 4 or higher vs. being at Level 3 or lower ranges from .92 to .98 with a median of .94; the proportion who would be consistently classified on two different test forms ranges from .89 to .97 with a median of .92.

Appendix 8 tables A.8.10 through A.8.21 provide more detailed information about the accuracy and the consistency of the classification of students into performance levels for mathematics. Each cell in the 5-by-5 table shows the estimated proportion of students who would be classified into a particular combination of performance levels. Fall block results for Algebra I, Geometry, and Algebra II are provided in the Addendum.

Table 8.31 Reliability of Classification: Summary for Mathematics

Grade Level	Testing Mode	Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
		Exact Level	Level 4 or higher vs. 3 or lower	Exact Level	Level 4 or higher vs. 3 or lower
3	CBT	0.77	0.93	0.68	0.90
	PBT	0.76	0.93	0.67	0.90
4	CBT	0.78	0.92	0.69	0.89
	PBT	0.77	0.94	0.68	0.91
5	CBT	0.77	0.92	0.68	0.89
	PBT	0.75	0.93	0.66	0.90
6	CBT	0.77	0.93	0.68	0.89
	PBT	0.75	0.92	0.66	0.89
7	CBT	0.80	0.93	0.72	0.90
	PBT	0.77	0.92	0.67	0.89
8	CBT	0.73	0.92	0.64	0.89
	PBT	0.73	0.92	0.64	0.89
A1	CBT	0.79	0.93	0.70	0.90
	PBT	0.72	0.95	0.63	0.92
GO	CBT	0.80	0.94	0.72	0.91
	PBT	0.74	0.96	0.64	0.95
A2	CBT	0.76	0.94	0.67	0.91
	PBT	0.74	0.94	0.65	0.92
M1	CBT	0.67	0.92	0.56	0.89
	PBT	-	-	-	-
M2	CBT	0.71	0.96	0.61	0.94
	PBT	-	-	-	-
M3	CBT	0.81	0.98	0.73	0.97
	PBT	-	-	-	-

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. "--" means insufficient sample size (< 100 students).

8.8 Inter-rater Agreement

Inter-rater agreement is the agreement between the first and second scores assigned to student responses. Inter-rater agreement measurements include exact, adjacent, and nonadjacent agreement. Pearson scoring staff used these statistics as one factor in determining the needs for continuing training and intervention on both individual and group levels. Table 8.32 displays both PARCC's expectations and the actual spring 2018 agreement percentages for perfect agreement and perfect plus adjacent agreement.

Table 8.32 Inter-rater Agreement Expectations and Results

Subject	Score Point Range	Perfect Agreement Expectation	Perfect Agreement Result	Within One Point Expectation	Within One Point Result
Mathematics	0–1	90%	97%	96%*	100%
Mathematics	0–2	80%	94%	96%	100%
Mathematics	0–3	70%	93%	96%	99%
Mathematics	0–4	65%	92%	95%	98%
Mathematics	0–5	65%	90%	95%	98%
Mathematics	0–6	65%	91%	95%	97%
ELA/L	Multi-trait	65%	75%	96%	99%

Note: *A 0 or 1 score compared to a blank score will have a disagreement greater than 1 point.

Pearson's ePEN2 scoring system included comprehensive inter-rater agreement reports that allowed supervisory personnel to monitor both individual and group performance. Based on reviews of these reports, scoring experts targeted individuals for increased backreading and feedback and, if necessary, retraining. Table 8.32 shows that the actual percentages for both exact reader agreement and the percentages of agreement within one point were higher than the inter-rater agreement expectations. Refer to Section 4 for more information on hand-scoring.

Section 9: Validity

9.1 Overview

The Standards for Educational and Psychological Testing, issued jointly by the American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME] (2014), reports:

> Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations (p. 11).

The purpose of test validation is not to validate the test itself but to validate interpretations of the test scores for particular uses. Test validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the lifetime of an assessment. Every aspect of an assessment provides evidence in support of its validity (or evidence of lack of validity), including design, content specifications, item development, and psychometric characteristics. The 2017–2018 operational assessment provided an opportunity to gather evidence of validity based on both test content and on the internal structure of the tests.

Pearson applies the principles of universal design, as articulated in materials developed by the National Center for Educational Outcomes (NCEO) at the University of Minnesota (Thompson, Johnstone, and Thurlow, 2002).

9.2 Evidence Based on Test Content

Evidence based on content of achievement tests is supported by the degree of correspondence between test items and content standards. The degree to which the test measures what it claims to measure is known as construct validity. The PARCC assessments adhere to the principles of evidence-centered design, in which the standards to be measured (the Common Core State Standards) are identified, and the performance a student needs to achieve to meet those standards is delineated in the PARCC evidence statements. Test items are reviewed for adherence to universal design principles, which maximize the participation of the widest possible range of students.

Pearson and PARCC built spreadsheets at the evidence statement level that incorporate the probability statements from the test blueprints and attrition rates at committee review and data review. The basis of our entire item development is driven by the use of these item development target spreadsheets provided by PARCC. Before beginning item development, Pearson uses these target spreadsheets to develop an internal item development plan to correlate with the expectations of the test design. These are reviewed and approved by PARCC. We acknowledge that each assessment has multiple parts and each part specifies the types of tasks and standards eligible for assessment.

In addition to the PARCC evidence statements, content is aligned through the articulation of performance in the performance level descriptors. At the policy level, the performance level descriptors include policy claims about the educational achievement of students who attain a particular performance level, and a broad description of the grade-level knowledge, skills, and practices students performing at a particular achievement level are able to demonstrate. Those policy-level descriptors are the foundation for the subject- and grade-specific performance level descriptors, which, along with the PARCC evidence frameworks, guide the development of the items and tasks.

The PARCC college- and career-ready determinations (CCRD) in English language arts/literacy (ELA/L) and mathematics describe the academic knowledge, skills, and practices students must demonstrate to show readiness for success in entry-level, credit-bearing college courses and relevant technical courses. The PARCC states determined that this level means graduating from high school and having at least a 75 percent likelihood of earning a grade of "C" or better in credit-bearing courses without the need for remedial coursework. After reviewing the standards and assessment design, the PARCC Governing Board (made up of the K–12 education chiefs in PARCC states) in conjunction with the PARCC Advisory Committee on College Readiness (composed of higher education chiefs in the PARCC states), determined that students who achieve at Levels 4 and 5 on the final PARCC high school assessments are likely to have acquired the skills and knowledge to meet the definition of college- and career-readiness. To validate the determinations, PARCC conducted a postsecondary educator judgment study and a benchmark study of the SAT, ACT, National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), Programme of International Student Assessment (PISA), and Progress in International Reading Literacy Study (PIRLS) tests (McClarty, Korbin, Moyer, Griffin, Huth, Carey, and Medberry, 2015).

Gathering construct validity evidence for PARCC is embedded in the process by which the PARCC assessment content is developed and validated. At each step in the assessment development process, PARCC states involved hundreds of educators, assessment experts, and bias and sensitivity experts in review of text, items, and tasks for accuracy, appropriateness, and freedom from bias. See Section 2 for an overview of the content development process. In the early stages of development, Pearson conducted research studies to validate the PARCC item and task development approach. One such study was a student task interaction study designed to collect data on the student's experience with the assessment tasks and technological functionalities, as well as the amount of time needed for answering each task. Pearson also conducted a rubric choice study that compared the functioning of two rubrics developed to score the prose constructed-response (PCR) tasks in ELA/L. Quantitative and qualitative evidence was collected to support the use of a condensed or expanded trait scoring rubric in scoring student responses.

PARCC items and tasks were field tested prior to their use on an assessment. During the initial field test administration in 2014, PARCC states collected feedback from students, test administrators, test coordinators, and classroom teachers on their experience with the PARCC assessments, including the quality of test items and student experience. A summary of the feedback can be found at: http://parcc-assessment.org/content/uploads/2017/11/PARCCCCRDPolicyandPLDsFINAL.pdf. The feedback from that survey was used to inform test directions, test timing, and the function of online task interactions.

Performance data from the field test also informed the future development of additional items and tasks.

All item developers and item writers are provided an electronic version of PARCC accessibility guidelines and PARCC's linguistic complexity rubric. Items and passages are reviewed internally by accessibility and fairness experts trained in the principles of universal design and who become well versed in PARCC's accessibility guidelines. Items received internal review for alignment to PARCC evidence tables, task generation model, item selection guidelines, and accessibility and fairness reviews.

An important consideration when constructing test forms is recognition of items that may introduce construct-irrelevant variance. Such items should not be included on test forms to help ensure fairness to all subgroups of test takers. PARCC convened bias and sensitivity committees to review all items. Additionally, content experts facilitated reviews of all items. All reviewers were trained using PARCC bias and sensitivity guidelines, and the guidelines were used to review items and ELA/L passages. Accommodations were made available based on individual need documented in the student's approved IEP, 504 Plan, or if required by the PARCC member state, an English Learner (EL) Plan. An accessibility specialist worked in consultation with the PARCC accessibility specialist to review forms and determine which forms should be used for students with accommodations.

The ELA/L and mathematics operational test forms, as described in Section 2, were carefully constructed to align with the test blueprints and specifications that are based on the Common Core State Standards (CCSS). During the fall of 2017, content experts representing various PARCC states, along with other content experts, held a series of meetings to review the operational forms for ELA/L and mathematics. These meetings provided opportunity to evaluate tests forms in their entirety and recommend changes. Requested item replacements were accommodated to the extent possible while striving to maintain the integrity of the various linking designs required for the operational test analyses. Psychometricians were available throughout this process to provide guidance with regard to implications of item replacements for the linking and statistical requirements.

Further information regarding the PARCC assessment college- and career-ready content standards, performance level descriptors, and accessibility features and accommodations is provided at http://parcc-assessment.org/college-career-ready.

9.3 Evidence Based on Internal Structure

Analyses of the internal structure of a test typically involve studies of the relationships among test items and/or test components (i.e., subclaims) in the interest of establishing the degree to which the items or components appear to reflect the construct on which a test score interpretation is based (AERA, APA, & NCME, 2014, p. 16). The term construct is used here to refer to the characteristics that a test is intended to measure; in the case of the PARCC operational tests, the characteristics of interest are the knowledge and skills defined by the test blueprint for ELA/L and for mathematics.

The PARCC assessments provide a full summative test score, Reading claim score, and Writing claim score as well as ELA/L subclaim and mathematics subclaim scores. The goal of reporting at this level is to

provide criterion-referenced data to assess the strengths and weaknesses of a student's achievement in specific components of each content area. This information can then be used by teachers to plan for further instruction, to plan for curriculum development, and to report progress to parents. The results can also be used as one factor in making administrative decisions about program effectiveness, teacher effectiveness, class grouping, and needs assessment.

9.3.1 Intercorrelations

The ELA/L full summative tests comprise two claim scores, Reading (RD) and Writing (WR), and five subclaim scores—Reading Literature (RL), Reading Information (RI), Reading Vocabulary (RV), Writing Written Expression (WE), and Writing Knowledge Language and Conventions (WKL). The RD claim score is a composite of RL, RI, and RV. The writing claim score, a composite of WE and WKL, comprises only PCR items, and the same PCR items are in each subclaim. The ELA/L operational test analyses were performed by evaluating the separate trait scores of WE and WKL, and for some PCR items also RL or RI; therefore, the trait scores were used for the intercorrelations.

The mathematics full summative tests have four subclaim scores—Major Content (MC), Mathematical Reasoning (MR), Modeling Practice (MP), and Additional and Supporting Content (ASC).

High total group internal consistencies as well as similar reliabilities across subgroups provide additional evidence of validity. High reliability of test scores implies that the test items within a domain are measuring a single construct, which is a necessary condition for validity when the intention is to measure a single construct. Refer to Section 8 for reliability estimates for the overall population, subgroups of interest, as well as for claims and subclaims for ELA/L and subclaims for mathematics.

Another way to assess the internal structure of a test is through the evaluation of correlations among scores. These analyses were conducted between the ELA/L Reading and Writing claim scores and the ELA/L subclaims (RL, RI, RV, WE, and WKL) and between the mathematics subclaims. If these components within a content area are strongly related to each other, this is evidence of unidimensionality.

A series of tables are provided to summarize the results for the spring 2018 administration.[6] Tables 9.1 through 9.9 present the Pearson correlations observed between the ELA/L Reading and Writing claim scores and subclaim scores for each grade; correlations are reported separately for online (CBT) and paper (PBT) versions of the tests. The tables provide the weighted average intercorrelations by averaging the intercorrelations computed for all the core operational forms of the test separately for the CBT and PBT tests within each grade level. The total sample size across all forms is provided in the upper triangle portion of the tables. The subclaim reliabilities (from Section 8) are reported along the diagonal. The WR, WE, and WKL scores tended to be highly correlated; this is expected given that these three intercorrelations are based on the trait scores from the same Writing items. RL, RI, and RV, all subclaims of Reading, are moderately to highly correlated. Additionally, the WR claim and the WE and WKL

[6] Addendum 9 provides a summary of results for the fall 2017 administration.

subclaims are moderately correlated with RD subclaims (of RL, RI, and RV). These moderate to high ELA/L intercorrelations amongst the subclaims are sufficiently high to provide evidence that the ELA/L tests are unidimensional. The moderate intercorrelations among the subclaims and claims suggest the claims may be sufficient for individual student reporting.

The intercorrelations and reliability estimates for mathematics are provided in Tables 9.10 through 9.21. The mathematics intercorrelations are moderate. The main observable pattern in the mathematics intercorrelations is that the MC subclaim generally has slightly higher correlations with the ASC, MR, and MP subclaims; the intercorrelations amongst the ASC, MR, and MP subclaims are usually slightly lower. The mathematics intercorrelations are sufficiently high to suggest that the mathematics tests are likely to be unidimensional with some minor secondary dimensions.

Additionally, the ELA/L and mathematics correlations for the two modes, PBT and CBT, displayed similar patterns of intercorrelations, with the CBT intercorrelations generally, but not always, being slightly larger than their PBT counterparts. Overall, these findings suggest that the structure of the PBT assessments and CBT assessments are similar. The sample sizes for Integrated Mathematics I, II, and III PBTs were below 100 and thus are not reported here.

Table 9.1 Average Intercorrelations and Reliability between Grade 3 ELA/L Subclaims

	CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL	
RD	0.86	296,409	296,409	296,409	296,409	296,409	296,409	RD	0.86	42,411	42,411	42,411	42,411	42,411	42,411	
RL	0.91	0.72	296,409	296,409	296,409	296,409	296,409	RL	0.91	0.71	42,411	42,411	42,411	42,411	42,411	
RI	0.87	0.68	0.65	296,409	296,409	296,409	296,409	RI	0.83	0.64	0.59	42,411	42,411	42,411	42,411	
RV	0.85	0.67	0.63	0.68	296,409	296,409	296,409	RV	0.88	0.68	0.62	0.75	42,411	42,411	42,411	
WR	0.72	0.67	0.65	0.56	0.83	296,409	296,409	WR	0.75	0.70	0.65	0.63	0.79	42,411	42,411	
WE	0.71	0.66	0.65	0.55	0.99	0.78	296,409	WE	0.74	0.68	0.64	0.61	0.99	0.70	42,411	
WKL	0.66	0.61	0.60	0.52	0.92	0.85	0.85	WKL	0.70	0.64	0.60	0.60	0.91	0.83	0.78	

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.2 Average Intercorrelations and Reliability between Grade 4 ELA/L Subclaims

	CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL	
RD	0.88	304,640	304,640	304,640	304,640	304,640	304,640	RD	0.87	40,580	40,580	40,580	40,580	40,580	40,580	
RL	0.94	0.79	304,640	304,640	304,640	304,640	304,640	RL	0.90	0.71	40,580	40,580	40,580	40,580	40,580	
RI	0.86	0.71	0.67	304,640	304,640	304,640	304,640	RI	0.88	0.68	0.68	40,580	40,580	40,580	40,580	
RV	0.83	0.68	0.62	0.61	304,640	304,640	304,640	RV	0.87	0.68	0.66	0.68	40,580	40,580	40,580	
WR	0.74	0.70	0.69	0.56	0.85	304,640	304,640	WR	0.76	0.70	0.69	0.62	0.81	40,580	40,580	
WE	0.74	0.70	0.69	0.55	0.99	0.83	304,640	WE	0.75	0.69	0.68	0.61	0.99	0.74	40,580	
WKL	0.69	0.66	0.63	0.53	0.94	0.89	0.88	WKL	0.73	0.67	0.65	0.60	0.93	0.88	0.80	

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.3 Average Intercorrelations and Reliability between Grade 5 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.88	333,568	333,568	333,568	333,568	333,568	333,568	RD	0.85	14,613	14,613	14,613	14,613	14,613	14,613
RL	0.91	0.73	333,568	333,568	333,568	333,568	333,568	RL	0.86	0.63	14,613	14,613	14,613	14,613	14,613
RI	0.87	0.69	0.67	333,568	333,568	333,568	333,568	RI	0.85	0.60	0.61	14,613	14,613	14,613	14,613
RV	0.88	0.71	0.66	0.73	333,568	333,568	333,568	RV	0.89	0.64	0.64	0.72	14,613	14,613	14,613
WR	0.74	0.68	0.69	0.59	0.87	333,568	333,568	WR	0.74	0.66	0.65	0.62	0.82	14,613	14,613
WE	0.73	0.67	0.68	0.58	0.99	0.84	333,568	WE	0.73	0.65	0.65	0.61	0.99	0.76	14,613
WKL	0.72	0.66	0.66	0.58	0.95	0.92	0.89	WKL	0.70	0.62	0.61	0.60	0.94	0.90	0.80

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.4 Average Intercorrelations and Reliability between Grade 6 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.90	333,216	333,216	333,216	333,216	333,216	333,216	RD	0.91	11,001	11,001	11,001	11,001	11,001	11,001
RL	0.94	0.82	333,216	333,216	333,216	333,216	333,216	RL	0.96	0.86	11,001	11,001	11,001	11,001	11,001
RI	0.90	0.75	0.72	333,216	333,216	333,216	333,216	RI	0.82	0.70	0.62	11,001	11,001	11,001	11,001
RV	0.84	0.70	0.65	0.65	333,216	333,216	333,216	RV	0.89	0.77	0.64	0.72	11,001	11,001	11,001
WR	0.73	0.71	0.70	0.53	0.87	333,216	333,216	WR	0.74	0.71	0.72	0.60	0.84	11,001	11,001
WE	0.73	0.70	0.69	0.52	1.00	0.86	333,216	WE	0.73	0.70	0.71	0.59	1.00	0.80	11,001
WKL	0.73	0.70	0.69	0.52	0.97	0.95	0.89	WKL	0.73	0.70	0.71	0.59	0.96	0.94	0.84

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.5 Average Intercorrelations and Reliability between Grade 7 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.88	331,251	331,251	331,251	331,251	331,251	331,251	RD	0.89	7,343	7,343	7,343	7,343	7,343	7,343
RL	0.91	0.72	331,251	331,251	331,251	331,251	331,251	RL	0.93	0.78	7,343	7,343	7,343	7,343	7,343
RI	0.92	0.74	0.77	331,251	331,251	331,251	331,251	RI	0.90	0.73	0.73	7,343	7,343	7,343	7,343
RV	0.81	0.64	0.65	0.58	331,251	331,251	331,251	RV	0.86	0.72	0.67	0.65	7,343	7,343	7,343
WR	0.77	0.71	0.74	0.56	0.88	331,251	331,251	WR	0.76	0.72	0.71	0.59	0.83	7,343	7,343
WE	0.76	0.71	0.73	0.55	1.00	0.87	331,251	WE	0.75	0.72	0.71	0.59	1.00	0.79	7,343
WKL	0.76	0.70	0.73	0.55	0.97	0.95	0.91	WKL	0.74	0.71	0.69	0.58	0.96	0.94	0.84

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.6 Average Intercorrelations and Reliability between Grade 8 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.88	331,553	331,553	331,553	331,553	331,553	331,553	RD	0.90	7,626	7,626	7,626	7,626	7,626	7,626
RL	0.90	0.73	331,553	331,553	331,553	331,553	331,553	RL	0.95	0.82	7,626	7,626	7,626	7,626	7,626
RI	0.93	0.75	0.77	331,553	331,553	331,553	331,553	RI	0.84	0.72	0.65	7,626	7,626	7,626	7,626
RV	0.83	0.64	0.67	0.62	331,553	331,553	331,553	RV	0.89	0.77	0.66	0.71	7,626	7,626	7,626
WR	0.76	0.73	0.73	0.53	0.88	331,553	331,553	WR	0.73	0.70	0.71	0.57	0.86	7,626	7,626
WE	0.75	0.72	0.73	0.52	1.00	0.89	331,553	WE	0.73	0.70	0.71	0.57	1.00	0.85	7,626
WKL	0.75	0.72	0.73	0.53	0.98	0.96	0.90	WKL	0.73	0.70	0.71	0.57	0.98	0.96	0.87

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.7 Average Intercorrelations and Reliability between Grade 9 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.90	127,308	127,308	127,308	127,308	127,308	127,308	RD	0.90	641	641	641	641	641	641
RL	0.93	0.79	127,308	127,308	127,308	127,308	127,308	RL	0.93	0.78	641	641	641	641	641
RI	0.92	0.78	0.80	127,308	127,308	127,308	127,308	RI	0.91	0.76	0.78	641	641	641	641
RV	0.85	0.70	0.68	0.65	127,308	127,308	127,308	RV	0.86	0.72	0.70	0.65	641	641	641
WR	0.81	0.78	0.77	0.59	0.89	127,308	127,308	WR	0.74	0.73	0.69	0.56	0.86	641	641
WE	0.80	0.77	0.77	0.59	1.00	0.90	127,308	WE	0.74	0.73	0.69	0.55	1.00	0.82	641
WKL	0.80	0.77	0.77	0.60	0.98	0.96	0.91	WKL	0.72	0.70	0.67	0.55	0.98	0.96	0.84

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.8 Average Intercorrelations and Reliability between Grade 10 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.87	186,647	186,647	186,647	186,647	186,647	186,647	RD	0.88	1,159	1,159	1,159	1,159	1,159	1,159
RL	0.91	0.72	186,647	186,647	186,647	186,647	186,647	RL	0.92	0.76	1,159	1,159	1,159	1,159	1,159
RI	0.90	0.73	0.74	186,647	186,647	186,647	186,647	RI	0.91	0.74	0.75	1,159	1,159	1,159	1,159
RV	0.83	0.65	0.65	0.59	186,647	186,647	186,647	RV	0.83	0.66	0.65	0.60	1,159	1,159	1,159
WR	0.78	0.74	0.74	0.56	0.88	186,647	186,647	WR	0.78	0.75	0.73	0.57	0.88	1,159	1,159
WE	0.77	0.73	0.73	0.56	1.00	0.88	186,647	WE	0.77	0.74	0.73	0.56	1.00	0.88	1,159
WKL	0.78	0.73	0.73	0.57	0.98	0.97	0.89	WKL	0.77	0.73	0.72	0.56	0.98	0.97	0.89

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.9 Average Intercorrelations and Reliability between Grade 11 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.88	101,559	101,559	101,559	101,559	101,559	101,559	RD	0.85	895	895	895	895	895	895
RL	0.91	0.74	101,559	101,559	101,559	101,559	101,559	RL	0.87	0.69	895	895	895	895	895
RI	0.93	0.75	0.79	101,559	101,559	101,559	101,559	RI	0.87	0.64	0.68	895	895	895	895
RV	0.79	0.62	0.64	0.59	101,559	101,559	101,559	RV	0.86	0.65	0.60	0.61	895	895	895
WR	0.75	0.71	0.73	0.49	0.87	101,559	101,559	WR	0.77	0.72	0.70	0.58	0.85	895	895
WE	0.75	0.70	0.73	0.49	1.00	0.85	101,559	WE	0.77	0.72	0.70	0.58	1.00	0.82	895
WKL	0.74	0.70	0.72	0.49	0.98	0.97	0.86	WKL	0.74	0.71	0.68	0.55	0.98	0.96	0.84

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.10 Average Intercorrelations and Reliability between Grade 3 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.89	305,106	305,106	305,106	MC	0.87	42,794	42,794	42,794
ASC	0.81	0.75	305,106	305,106	ASC	0.80	0.72	42,794	42,794
MR	0.78	0.72	0.70	305,106	MR	0.78	0.72	0.67	42,794
MP	0.74	0.68	0.70	0.71	MP	0.73	0.68	0.69	0.67

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.11 Average Intercorrelations and Reliability between Grade 4 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.88	312,951	312,951	312,951	MC	0.86	40,807	40,807	40,807
ASC	0.75	0.69	312,951	312,951	ASC	0.74	0.67	40,807	40,807
MR	0.78	0.67	0.74	312,951	MR	0.74	0.67	0.74	40,807
MP	0.72	0.63	0.72	0.60	MP	0.70	0.63	0.71	0.69

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.12 Average Intercorrelations and Reliability between Grade 5 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.87	340,771	340,771	340,771	MC	0.86	14,709	14,709	14,709
ASC	0.71	0.62	340,771	340,771	ASC	0.69	0.57	14,709	14,709
MR	0.79	0.65	0.76	340,771	MR	0.76	0.61	0.68	14,709
MP	0.74	0.62	0.69	0.65	MP	0.76	0.63	0.70	0.66

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.13 Average Intercorrelations and Reliability between Grade 6 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.87	334,362	334,362	334,362	MC	0.83	11,037	11,037	11,037
ASC	0.77	0.70	334,362	334,362	ASC	0.76	0.71	11,037	11,037
MR	0.78	0.70	0.72	334,362	MR	0.75	0.69	0.66	11,037
MP	0.73	0.66	0.71	0.69	MP	0.72	0.66	0.67	0.68

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.14 Average Intercorrelations and Reliability between Grade 7 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.88	315,954	315,954	315,954	MC	0.85	7,322	7,322	7,322
ASC	0.75	0.66	315,954	315,954	ASC	0.66	0.54	7,322	7,322
MR	0.81	0.70	0.76	315,954	MR	0.77	0.60	0.72	7,322
MP	0.79	0.68	0.75	0.63	MP	0.75	0.61	0.73	0.63

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.15 Average Intercorrelations and Reliability between Grade 8 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.82	256,214	256,214	256,214	MC	0.81	7,477	7,477	7,477
ASC	0.75	0.67	256,214	256,214	ASC	0.74	0.64	7,477	7,477
MR	0.74	0.66	0.68	256,214	MR	0.73	0.64	0.70	7,477
MP	0.70	0.64	0.66	0.60	MP	0.72	0.64	0.71	0.64

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.16 Average Intercorrelations and Reliability between Algebra I Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.84	215,867	215,867	215,867	MC	0.69	1,161	1,161	1,161
ASC	0.81	0.72	215,867	215,867	ASC	0.63	0.54	1,161	1,161
MR	0.76	0.71	0.72	215,867	MR	0.65	0.61	0.61	1,161
MP	0.78	0.73	0.73	0.78	MP	0.69	0.64	0.70	0.70

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.17 Average Intercorrelations and Reliability between Geometry Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.85	129,428	129,428	129,428	MC	0.68	681	681	681
ASC	0.81	0.76	129,428	129,428	ASC	0.64	0.50	681	681
MR	0.76	0.73	0.77	129,428	MR	0.66	0.51	0.62	681
MP	0.76	0.72	0.78	0.74	MP	0.67	0.56	0.68	0.59

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.18 Average Intercorrelations and Reliability between Algebra II Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.84	122,392	122,392	122,392	MC	0.68	826	826	826
ASC	0.78	0.75	122,392	122,392	ASC	0.72	0.67	826	826
MR	0.72	0.73	0.73	122,392	MR	0.68	0.70	0.66	826
MP	0.75	0.73	0.71	0.70	MP	0.68	0.67	0.71	0.58

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.19 Average Intercorrelations and Reliability between Integrated Mathematics I Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.70	750	750	750	MC				
ASC	0.67	0.59	750	750	ASC				
MR	0.69	0.69	0.64	750	MR				
MP	0.57	0.56	0.53	0.56	MP				

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Integrated Mathematics I PBT had insufficient sample sizes. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.20 Average Intercorrelations and Reliability between Integrated Mathematics II Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.73	853	853	853	MC				
ASC	0.67	0.62	853	853	ASC				
MR	0.68	0.61	0.61	853	MR				
MP	0.73	0.65	0.68	0.70	MP				

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Integrated Mathematics II PBT had an insufficient sample size. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

Table 9.21 Average Intercorrelations and Reliability between Integrated Mathematics III Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.41	186	186	186	MC				
ASC	0.32	0.12	186	186	ASC				
MR	0.48	0.40	0.52	186	MR				
MP	0.58	0.48	0.68	0.60	MP				

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Integrated Mathematics III PBT had an insufficient sample size. Please refer to Appendix 12.1 (Form Composition) for information about the number of items and number of score points in each subclaim.

9.3.2 Reliability

Additionally, the reliability analyses presented in Section 8 of this technical report provide information about the internal consistency of the PARCC assessments. Internal consistency is typically measured via correlations amongst the items on an assessment and provides an indication of how much the items measure the same general construct. The reliability estimates, computed using coefficient alpha (Cronbach, 1951), are presented in Tables 8.1 and 8.2 and are along the diagonals of Tables 9.1 through 9.21.[7] The average reliabilities for ELA/L and mathematics PARCC assessments range from .77 up to .94. Tables 8.5 through 8.13 summarize test reliability for groups of interest for ELA/L grades 3 through 11, and Tables 8.14 through 8.25 summarize test reliability for groups of interest for mathematics grades/subjects. Along with the subclaim intercorrelations, the reliability estimates indicate that the items within each PARCC assessment are measuring the same construct and provide further evidence of unidimensionality.

9.3.3 Local Item Dependence

In addition to the intercorrelations for ELA/L and mathematics, local item independence was evaluated. Local independence is one of the primary assumptions of IRT that states the probability of success on one item is not influenced by performance on other items, when controlling for ability level. This implies that ability or theta accounts for the associations among the observed items. Local item dependence (LID) when present essentially overstates the amount of information predicted by the IRT model. It can exert other undesirable psychometric effects and represents a threat to validity since other factors besides the construct of interest are present. Classical statistics are also affected when LID is present since estimates of test reliability like IRT information can be inflated (Zenisky, Hambleton, & Sireci, 2003).

The LID issue affects the choice of item scoring in IRT calibrations. Specifically, if evidence suggests these items indeed have local dependence, then it might be preferable to sum the item scores into clusters or testlets as a method of minimizing LID. However, if these items do not appear to have strong local item dependence, then retaining the scores as individual item scores in an IRT calibration is preferred since more information concerning item properties is retained. During the initial operational administration of the PARCC assessments in spring 2015, a study that included two methods of investigating the presence of LID was conducted. A description of the methods along with study findings are summarized below.

First, analyses of the internal consistency in items and testlets were conducted under classical test theory (Wainer & Thissen, 2001) as a way to evaluate the degree of LID. Two estimates of Cronbach's alpha (Cronbach, 1951) were compared based on individual items in a test and those clustered into testlets. Cronbach's alpha is formulated as:

[7] Section 8 provides information on the computations of the reliability estimates.

$$\alpha = \frac{k}{k-1}\frac{\sum_{i \neq i'} \sigma_{ii'}}{\sigma_T^2}, \quad (9\text{-}1)$$

where k is the total number of items, $\sigma_{ii'}$ is the covariance of items i and i' ($i \neq i'$), and σ_T^2 is the variance of total scores. To compute an alpha coefficient, sample standard deviations and variances are substituted for the $\sigma_{ii'}$ and σ_T^2. The alpha for the total test based on individual items is compared with those that form testlets based on larger subparts. If the item-level configuration has appreciably higher levels of internal consistency compared with the testlets, LID may be present.

For IRT-based methods, local dependence can be evaluated using statistics such as Q_3 (Yen, 1984). The item residual is the difference between observed and expected performance. The Q_3 index is the correlation between residuals of each item pair defined as

$$\begin{aligned} d_i &= (O - \hat{E}), \\ Q_3 &= r(d_i, d_{i'}) \end{aligned} \quad (9\text{-}2)$$

where O is the observed score and $\hat{E}$ is the expected value of O under a proposed IRT model and the index is defined as the correlation between the two item residuals.

LID manifests itself as a residual correlation that is nonzero and large. For Q_3, LID can be either positive or negative. Positive (negative) LID indicates that performance is higher (lower) than expectation. The residual Q_3 correlation matrix can be inspected to determine if there are any blocks of locally dependent items (e.g., perhaps blocks of items belonging to the same reading passage). For Q_3, the null hypothesis is that local independence holds. The expected value of Q_3 is -1/n-1 where n is the number of items such that the statistic shows a small negative bias. As a rule of thumb, item pairs with moderate levels of LID for Q_3 are |0.20| or greater. Significant levels of LID are present when the statistic is greater than |0.40|. An alternative is to use the Fisher r to z transformation and evaluate the resulting p values.

For the LID comparisons, the following eight test levels administered in spring 2015 were selected:

1. Grade 4 for span 3–5 in ELA/L,
2. Grade 4 for span 3–5 in mathematics,
3. Grade 7 for span 6–8 in ELA/L,
4. Grade 7 for span 6–8 in mathematics,
5. Grade 10 for span 9–11 in ELA/L,
6. Integrated Mathematics II for Integrated Mathematics I–III,
7. Algebra I, and
8. Algebra II.

One spring 2015 CBT form for each of the eight tests was selected that was roughly at the median in terms of test difficulty. For ELA/L, reading items were summed according to passage assignment. For mathematics, items were summed according to subclaims. Cronbach's alpha was computed for the entire forms using the two different approaches as described above, one involving calculations at the

item level and the second utilizing scores on summed items (i.e., testlets). Further description of the data is given in Table 9.22.

To cross-validate the internal consistency analysis, the Q_3 statistic was computed from spring CBT data based on grade 4 ELA/L and Integrated Mathematics II items. All items in the pool at that test level were included. The CBT item pool for grade 4 ELA/L contained 125 items while Integrated Mathematics II had 77 items.

The results for the internal consistency analysis are shown in Figure 9.1. In every instance, the item-level Cronbach's alpha is higher than in the testlet configuration. The greatest difference was for Algebra II, which showed a difference of 0.07. Although this was not unexpected, the magnitude of the differences in the respective alpha coefficients in general do not suggest a concerning level of LID. Table 9.23 shows the summary for the Q_3 values. Figures 9.2 and 9.3 show graphs of the distribution of Q_3 values. Most of the Q_3 values were small and negative, again suggesting that LID is not at a level of concern.

For these two test levels, the difference in the alpha coefficients was 0.03 and was consistent with the low values of Q_3.

In summary, this investigation did not find evidence for the existence of pervasive LID. The results of both the internal consistency analyses and Q_3 methods support a claim of minimal LID. For a multiple-choice-only test containing four reading passages with 5 to 12 items associated with a reading passage, Sireci, Thissen, and Wainer (1991) reported that testlet alpha was approximately 10 percent lower than the item-level coefficient. In comparison, PARCC tests have complex test structures and exhibited smaller differences in alpha coefficients. In addition, the median Q_3 values presented in Table 9.23 centered around the expectation of -1/n-1.

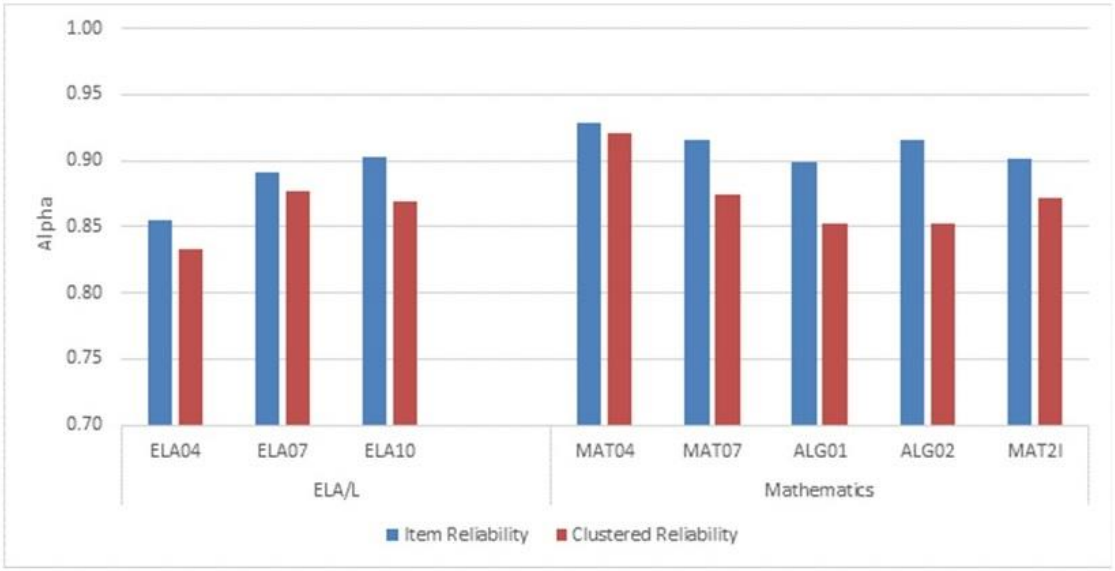

Figure 9.1 Comparison of Internal Consistency by Item and Cluster (Testlet)

Table 9.22 Conditions used in LID Investigation and Results

Content	Grade	N Valid	N Complete	Percent Incomplete	No. Items	No. Tasks	Item Rel.	Task Rel.
				ELA/L				
ELA/L	4	13,660	13,518	1.04	31	5	0.86	0.83
ELA/L	7	12,757	12,685	0.56	41	7	0.89	0.88
ELA/L	10	3,097	3,033	2.07	41	7	0.90	0.87
				Mathematics				
Math	4	10,332	10,255	0.75	53	4	0.93	0.92
Math	7	10,295	10,188	1.04	50	6	0.92	0.87
Math	A1	5,072	4,885	3.69	52	6	0.90	0.85
Math	A2	4,982	4,769	4.28	54	6	0.92	0.85
Math	M2	2,708	2,645	2.33	51	6	0.90	0.87

Note: A1 = Algebra I, A2 = Algebra II, M2 = Integrated Mathematics II.

Table 9.23 Summary of Q_3 Values for ELA/L Grade 4 and Integrated Mathematics II (Spring 2015)

Min.	Q_1	Median	Mean	Q_3	Max.	*SD*
			ELA/L Grade 4			
-0.138	-0.047	-0.031	-0.031	-0.017	0.279	0.030
			Integrated Mathematics II			
-0.160	-0.038	-0.017	-0.019	0.001	0.280	0.032

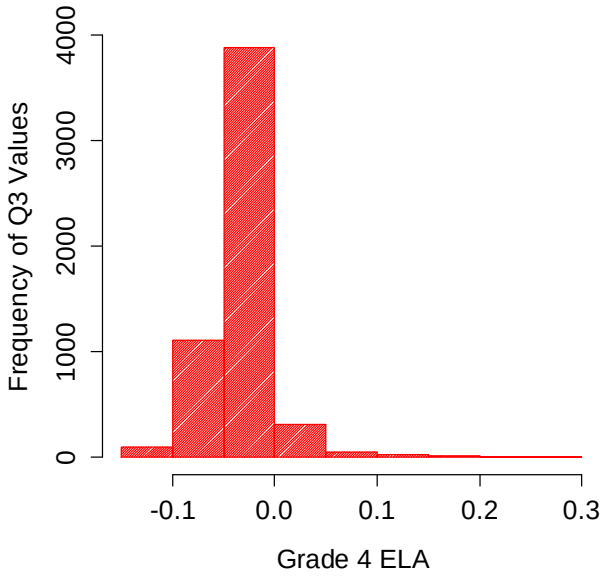

Figure 9.2 Distribution of Q_3 Values for Grade 4 ELA/L (Spring 2015)

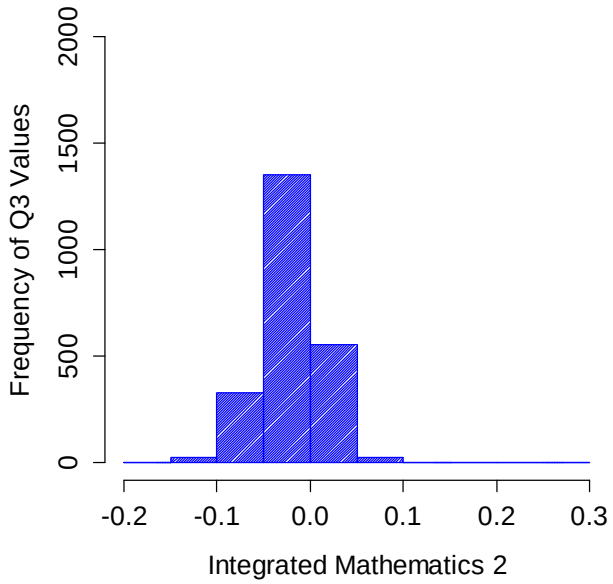

Figure 9.3 Distribution of Q_3 Values for Integrated Mathematics II (Spring 2015)

9.4 Evidence Based on Relationships to Other Variables

Empirical results concerning the relationships between scores on a test and measures of other variables external to the test can also provide evidence of validity when these relationships are found to be consistent with the definition of the construct that the test is intended to measure. As indicated in the AERA, APA, and NCME standards (2014), the variables investigated can include other tests that measure the same construct and different constructs, criterion measures that scores on the test are expected to predict, as well as demographic characteristics of test takers that are expected to be related and unrelated to test performance.

The relationship of the scores across the ELA/L and mathematics assessments was evaluated using correlational analyses. Tables 9.24 through 9.29 present the Pearson correlations observed between the ELA/L scale scores and the mathematics scale scores for each grade; the correlations are reported separately for online (CBT) and paper (PBT) versions of the tests. For grades 3 through 8, students must have a valid test score for both ELA/L and mathematics at the same grade level to be included in the tables. These tables provide the correlation in the lower triangle and the sample size is provided in the upper triangle. In computing the correlations between a particular pair of ELA/L and mathematics tests, test takers must have taken both tests via the same mode in spring 2018. Across both modes, ELA/L, Reading (RD), and Writing (WR) are moderately to highly correlated with mathematics; the correlations range from .61 up to .80 for grades 3 through 8. These correlations suggest that the ELA/L and mathematics tests are assessing different content. The higher intercorrelations between the ELA/L, Reading (RD), and Writing (WR) scores suggest stronger internal relationships when compared to the correlations with the mathematics content area.

The ELA/L and mathematics correlations for the high school tests are presented in Tables 9.30 through 9.32. Because students in high school can take the mathematics courses in different years (e.g., one student may take Algebra I in grade 9 while another student may take Algebra I in grade 10), the high school mathematics scores were correlated with several of the ELA/L grades (e.g., Algebra I correlated with both grades 9 and 10). Only correlations for pairings with total sample sizes of at least 100 are shown in the tables. Across both modes of grades 8 through 11, ELA/L, Reading (RD), and Writing (WR) scores have correlations with high school mathematics tests that range from .38 to .76. Correlations between high school mathematics scores and corresponding ELA/L scores demonstrate low to moderate correlations.

Table 9.24 Correlations between ELA/L and Mathematics for Grade 3

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		295,747	295,747	295,747	ELA/L		42,185	42,185	42,185
RD	0.95		295,747	295,747	RD	0.95		42,185	42,185
WR	0.88	0.73		295,747	WR	0.89	0.73		42,185
MA	0.79	0.76	0.70		MA	0.77	0.75	0.67	

Note: ELA/L = English language arts/literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 3 ELA/L score *and* a valid grade 3 mathematics score to be included in this table.

Table 9.25 Correlations between ELA/L and Mathematics for Grade 4

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		304,145	304,145	304,145	ELA/L		40,235	40,235	40,235
RD	0.95		304,145	304,145	RD	0.96		40,235	40,235
WR	0.88	0.73		304,145	WR	0.89	0.73		40,235
MA	0.80	0.78	0.70		MA	0.76	0.75	0.64	

Note: ELA/L = English language arts/literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 4 ELA/L score *and* a valid grade 4 mathematics score to be included in this table.

Table 9.26 Correlations between ELA/L and Mathematics for Grade 5

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		332,564	332,564	332,564	ELA/L		14,478	14,478	14,478
RD	0.95		332,564	332,564	RD	0.95		14,478	14,478
WR	0.87	0.72		332,564	WR	0.87	0.71		14,478
MA	0.77	0.76	0.66		MA	0.73	0.72	0.61	

Note: ELA/L = English language arts/literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 5 ELA/L score *and* a valid grade 5 mathematics score to be included in this table.

Table 9.27 Correlations between ELA/L and Mathematics for Grade 6

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		331,613	331,613	331,613	ELA/L		10,902	10,902	10,902
RD	0.95		331,613	331,613	RD	0.95		10,902	10,902
WR	0.86	0.71		331,613	WR	0.85	0.70		10,902
MA	0.80	0.79	0.66		MA	0.77	0.77	0.63	

Note: ELA/L = English language arts/literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 6 ELA/L score *and* a valid grade 6 mathematics score to be included in this table.

Table 9.28 Correlations between ELA/L and Mathematics for Grade 7

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		312,685	312,685	312,685	ELA/L		7,226	7,226	7,226
RD	0.95		312,685	312,685	RD	0.95		7,226	7,226
WR	0.90	0.75		312,685	WR	0.87	0.71		7,226
MA	0.79	0.79	0.68		MA	0.78	0.77	0.65	

Note: ELA/L = English language arts/literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 7 ELA/L score *and* a valid grade 7 mathematics score to be included in this table.

Table 9.29 Correlations between ELA/L and Mathematics for Grade 8

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		253,208	253,208	253,208	ELA/L		7,391	7,391	7,391
RD	0.93		253,208	253,208	RD	0.94		7,391	7,391
WR	0.89	0.70		253,208	WR	0.90	0.72		7,391
MA	0.75	0.74	0.63		MA	0.75	0.75	0.63	

Note: ELA/L = English language arts/literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 8 ELA/L score *and* a valid grade 8 mathematics score to be included in this table.

Table 9.30 Correlations between ELA/L and Mathematics for High School

ELA/L	CBT						ELA/L	PBT					
	A1	GO	A2	M1	M2	M3		A1	GO	A2	M1	M2	M3
8	0.71	0.55	0.56				8	0.76					
	(63,901)	(8,693)	(635)					(108)					
9	0.71	0.70	0.69	0.70			9	0.61					
	(83,394)	(28,719)	(6,652)	(302)				(463)					
10	0.55	0.62	0.69	0.69	0.60		10	0.55	0.60	0.64			
	(17,047)	(72,744)	(40,863)	(277)	(549)			(143)	(457)	(171)			
11	0.46	0.51	0.58	0.64	0.70	0.54	11			0.68			
	(2,499)	(9,167)	(56,504)	(102)	(190)	(163)				(563)			

Note: ELA/L = English language arts/literacy, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. The correlations are provided with the sample sizes, below in parentheses. Shaded cells indicate pairings with sample sizes less than 100.

Table 9.31 Correlations between ELA/L Reading and Mathematics for High School

RD	CBT						RD	PBT					
	A1	GO	A2	M1	M2	M3		A1	GO	A2	M1	M2	M3
8	0.70	0.57	0.58				8	0.71					
	(63,901)	(8,693)	(635)					(108)					
9	0.69	0.70	0.68	0.70			9	0.55					
	(83,394)	(28,719)	(6,652)	(302)				(463)					
10	0.53	0.62	0.69	0.68	0.59		10	0.45	0.55	0.59			
	(17,047)	(72,744)	(40,863)	(277)	(549)			(143)	(457)	(171)			
11	0.46	0.53	0.59	0.65	0.72	0.55	11			0.66			
	(2,499)	(9,167)	(56,504)	(102)	(190)	(163)				(563)			

Note: RD = Reading, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. The correlations are provided with the sample sizes, below in parentheses. Shaded cells indicate pairings with sample sizes less than 100.

Table 9.32 Correlations between ELA/L Writing and Mathematics for High School

WR	CBT						WR	PBT					
	A1	GO	A2	M1	M2	M3		A1	GO	A2	M1	M2	M3
8	0.61	0.42	0.42				8	0.69					
	(63,901)	(8,693)	(635)					(108)					
9	0.61	0.61	0.61	0.58			9	0.58					
	(83,394)	(28,719)	(6,652)	(302)				(463)					
10	0.47	0.52	0.60	0.59	0.50		10	0.57	0.53	0.59			
	(17,047)	(72,744)	(40,863)	(277)	(549)			(143)	(457)	(171)			
11	0.38	0.40	0.46	0.48	0.62	0.46	11			0.58			
	(2,499)	(9,167)	(56,504)	(102)	(190)	(163)				(563)			

Note: WR = Writing, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. The average correlations are provided with the sample sizes, below in parentheses. Shaded cells indicate pairings with sample sizes less than 100.

9.5 Evidence from the Special Studies

Several research studies were conducted to provide additional validity evidence for the PARCC's goals of assessing more rigorous academic expectations, helping to prepare students for college and careers, and providing information back to teachers and parents about their students' progress toward college and career readiness. Some of the special studies conducted include:

- content alignment studies,
- a benchmarking study,
- a longitudinal study of external validity,
- a mode comparability study, and
- a device comparability study.

The following paragraphs briefly describe each of these studies.

9.5.1 Content Alignment Studies

In 2016, content of the ELA/L assessments at grades 5, 8, and 11 and the Algebra II and Integrated Mathematics II assessments were evaluated to determine how well the PARCC assessments were aligned to the Common Core State Standards (CCSS; Doorey, & Polikoff, 2016, Schultz, Michaels, Dvorak, & Wiley, 2016). These content alignment studies were conducted by the Fordham Institute for grades 5 and 8 and by Human Resources Research Organization (HumRRO) for the high school assessments. Both of these studies used the same methodology by having content experts review the assessment items and answers (for the constructed-response items the rubrics were reviewed). The content experts then judged how well the items aligned to the CCSS, the depth of knowledge of the items, and the accessibility of the items to all students, including English learners and students with disabilities. The authors of both studies noted that the content experts reviewing the assessments were required to be familiar with the CCSS but could not be employed by participating organizations or be the writers of the CCSS. Therefore, an effort was made to eliminate any potential conflicts of interest.

The content studies had the individual content experts review and rate each item; then as a group the content experts came to a consensus on the final ratings for the content alignment, depth of knowledge, and accessibility to all students. In addition to the ratings, the content experts were asked to make comments that provided an explanation of their ratings; these comments were then used by the full group of content experts to provide narrative comments regarding the overall ratings and to provide feedback and recommendation about the assessment programs.

The PARCC assessment program was rated as *Excellent Match* for ELA/L content and depth and *Good Match* for mathematics content and depth for grades 5 and 8. However, for grade 11 ELA/L content was rated as *Excellent Match* but depth was rated as *Limited/Uneven Match*. The high school mathematics assessments were rated at *Excellent Match* for content and *Good Match* for depth.

The content studies noted some weaknesses and strengths of the PARCC assessments. For ELA/L, it was noted that the assessments include complex texts, a range of cognitive demands, and have a variety of

item types. Furthermore, the ELA/L "assessments require close reading, assess writing to sources, research, and inquiry, and emphasize vocabulary and language skills" (Doorey & Polikoff, 2016). The grade 11 ELA/L assessment had a smaller range of depth and included items assessing the higher-demand cognitive level. A weakness of the ELA/L assessments is the lack of a listening and speaking component. It was also suggested that the ELA/L assessments could be enhanced by the inclusion of a research task that requires the use of two or more sources of information.

The strengths of the mathematics assessments include assessments that are aligned to the major work for each grade level. While the grade 5 assessment includes a range of cognitive demand, the grade 8 assessment includes a number of higher-demand items and may not fully assess the standards at the lowest level of cognitive demand. It was suggested that the grade 5 assessment could include more focus on the major work and the grade 8 assessment could include items at the lowest cognitive demand level. Additionally, the reviewers noted that some of the mathematics items should be carefully reviewed for editorial and mathematical accuracy.

The high school report noted that the PARCC assessment program incorporates a number of accessibility features and test accommodations for students with disabilities and for English learners. Furthermore, the PARCC assessments included items designed to accommodate the needs of students with disabilities.

In 2017, HumRRO conducted a study to evaluate the quality and alignment of ELA/L and mathematics assessments for grades 3, 4, 6, and 7 (Shultz, Norman-Dvorak, Chen, 2017). This alignment study followed a similar methodology as the 2016 study. Reviewers were asked to determine the extent to which items were aligned to the CCSS, using fully, partially, or not aligned as the rating categories. Ratings were averaged to determine overall alignment. For ELA/L, 99.6 percent of grade 3 and 4 items, 95.5 percent of grade 6 items, and 94.6 percent of grade 7 items were fully aligned. For mathematics, 92.0 percent of grade 3, 91.1 percent of grade 4 items, 83.1 percent of grade 6 items, and 94.0 percent of grade 7 items were fully aligned. The majority of the items that did not fall into fully aligned were considered partially aligned to the standards. CCSS are designed to be measured by multiple items, so items that aligned to multiple CCSS received a partially aligned rating. The overall item-to-CCSS alignment was captured by a holistic alignment rating that indicated if an item captured the identified standards as a set. Holistic ratings (either yes or no) were found by averaging review ratings across clusters for items that included more than one standard. For ELA, for all four grades, at least 93 percent of items had a holistic alignment rating of yes to indicate that the identified standards captured the skills or knowledge required. For mathematics, grade 6 had the lowest percentage for the holistic alignment rating of yes (84.8 percent), and grade 7 had the highest (96.3 percent). Overall the alignment study suggests that the identified CCSS capture the knowledge and skills required in the items.

In addition to the alignment study, HumRRO also evaluated the CCSSO criteria for content and depth for ELA/L and mathematics grades 3, 4, 6, and 7, as well as the cognitive complexity levels of these same grades. There are five criteria for ELA/L content: close reading, writing, vocabulary and language skills, research and inquiry, and speaking and listening. Reviewers were asked to rate the content as *Excellent, Good, Limited/Uneven*, or *Weak Match*. For grades 3, 4, 5, and 7, the ELA/L assessments received a

composite rating of *Excellent Match* for assessing the content needed for college and career readiness. There are four criteria for ELA/L depth: text quality and types, complexity of texts, cognitive demand, and high-quality items and item variety. All grades in this study received a composite rating of *Good Match* for depth. For mathematics content, the composite rating is based on two criteria: focus and concepts, procedures, and applications. Grades 3, 4, and 6 received a composite content rating of *Good Match*, and grade 7 received a composite content rating of *Excellent Match*. The mathematics composite depth rating is based on three criteria: connecting practice to content, cognitive demand, and high-quality items and item variety. All grades in the study were rated as *Excellent Match* at assessing the depth needed to successfully meet college and career readiness.

Finally, the 2017 HumRRO study looked at cognitive complexity of the items on ELA/L and mathematics at grades 3, 4, 6, and 7. Reviewers indicated their agreement with the intended cognitive complexity ratings provided by PARCC of low, medium, or high. The results indicated that the reviewers generally agreed with the distribution of complexity levels assigned by PARCC. There were differences in agreements in ELA/L language cluster and a few exceptions to agreement in math, particularly at grade 6, where there was disagreement in the ratings at the medium complexity level for two domains and the high complexity level for one domain. For grade 7, there was agreement across low, medium, and high in all domains.

9.5.2 Benchmarking Study

The purpose of the PARCC benchmarking study (McClarty, Korbin, Moyer, Griffin, Huth, Carey, and Medberry, 2015) was to provide information that would inform the PARCC performance level setting (PLS) process. PARCC used an evidence-based standard setting approach (EBSS; McClarty, Way, Porter, Beimers, & Miles, 2013) to establish the performance levels for its assessments. In EBSS, the threshold scores for performance levels are set based on a combination of empirical research evidence and expert judgment. This benchmarking study provided one source of empirical evidence to inform the PARCC college- and career-readiness performance level (i.e., Level 4). The study findings were provided to PARCC's pre-policy standard-setting committee. The charge of this committee was to suggest a reasonable range for the percentage of students meeting or exceeding the PARCC Level 4 threshold score and therefore considered college- and career-ready. Section 11.3.2 of this report provides more information about the PARCC pre-policy meeting. For the PARCC benchmarking study, external information was analyzed to provide information about the Level 4 threshold scores for the grade 11 ELA/L, Algebra II, and Integrated Mathematics III assessments, the grade 8 ELA/L and mathematics assessments, and the grade 4 ELA/L and mathematics assessments. The PARCC assessments and Level 4 expectations were compared with comparable assessments and expectations for the Programme of International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), National Assessment of Educational Progress (NAEP), ACT, SAT, the Michigan Merit Exam, and the Virginia End-of-Course exams. For each external assessment, the best-matched performance level was determined and the percentage of students reaching that level across the nation and in the PARCC states was determined. Across all grades and subjects, the data indicated approximately 25 to 50 percent of students were college- and career-ready or on track to readiness based on PARCC's Level 4 expectations.

For details on how the benchmarking study was used during the standard setting process, refer to Section 11 of this technical report.

9.5.3 Longitudinal Study of External Validity of PARCC Performance Levels (Phase 1)

In 2016–2017, the first phase of a two-part external validity study of claims about the alignment of PARCC Level 4 to college readiness was completed (Steedle, Quesen, and Boyd, 2017) using PARCC assessment scores from the 2014–2015 and 2015–2016 academic years. Associations between PARCC performance levels and college-readiness benchmarks established by the College Board and ACT were used to study the claim that students who achieve Level 4 have a .75 probability of attaining at least a C in entry-level, credit-bearing, postsecondary coursework. Regression estimates measured the relationship between the PARCC and external test scores. The PARCC Level 4 benchmark was used to estimate the expected score on an external test, and vice versa. Assessment scores were dichotomized for additional analyses. Cross-tabulation tables provided classification agreement among tests. Logistic regression modeled the relationship between students' PARCC scores and their probabilities of meeting the external assessment benchmark, and vice versa.

These methods were used to make the following comparisons in mathematics: Algebra I and PSAT10 Math; Geometry and PSAT10 Math; Algebra II and PSAT10 Math; Algebra II and PSAT/NMSQT Math; Algebra II and SAT Math; and Algebra II and ACT Math. The classification agreement (meeting the benchmark on both tests or not meeting the benchmark on both tests) ranged from 62.5 percent to 86.5 percent. The overall trend indicated that students who met the benchmark on a PARCC mathematics assessment were likely to meet or exceed the benchmark on an external test (probabilities ranged from .509 to .886). However, students who met the benchmark on the external test had relatively low probabilities of meeting the PARCC mathematics benchmark (.097 to .310).

The following comparisons were made in English language arts/literacy: grade 9 and PSAT10 evidence-based reading and writing (EBRW); grade 10 and PSAT10 EBRW; grade 10 and PSAT/NMSQT EBRW; grade 10 and SAT EBRW; grade 11 and PSAT/NMSQT EBRW; grade 11 and SAT EBRW; grade 11 and ACT English; and grade 11 and ACT reading. In the majority of comparisons, the trend in ELA/L results was similar to mathematics. The classification agreements ranged from 67.3 percent to 79.7 percent. Students meeting the PARCC ELA/L benchmark had probabilities between .667 and .825 of meeting the benchmark on the external assessment. However, a student taking the external test had lower probabilities of meeting the benchmark on the PARCC ELA/L assessments (.326 to .513).

Overall, results indicated that a student meeting the benchmark on the PARCC test had a high probability of making the benchmark on the external test, but the converse did not hold for students meeting the benchmark on the external test, for the majority of comparisons. These results suggest that meeting the PARCC benchmark is an indicator of academic readiness for college. However, it may be that students who meet the PARCC benchmark have a greater than .75 probability of earning a C or higher in first-year college courses.

Phase 1 is a preliminary study using indirect comparisons; therefore, there are limitations to interpretations. Phase 2 of this study was to occur in 2018 and use longitudinal data including academic

performance in entry-level college courses for students who took PARCC assessments during high school. Currently, this study is on hold due to challenges obtaining student academic data from entry-level college courses and/or matching the data to the PARCC student scores.

9.5.4 2017 Mode and Device Comparability Studies

PARCC assessments have been operational since the 2014–2015 school year. In addition to the traditional paper format, the assessments were available for online administration via a variety of electronic devices, including desktop computers, laptop computers, and tablets. PARCC's research agenda includes several studies evaluating the interchangeability of scale scores across modes and devices.

This report describes a two-pronged study consisting of a mode comparability analysis and a device comparability analysis. In the mode comparability analysis, scores arising from the paper administration were compared to those arising from any type of online administration. In the device comparability analysis, online scores arising from tests administered using a tablet are compared with online scores arising from any other type of electronic administration where a tablet was not present (i.e., laptops, desktops, Chromebooks).

The goal of this study was threefold: 1) to investigate whether assessment items were of similar difficulty across the levels of conditions for each analysis (i.e., paper and online for the mode comparability analysis and tablet and non-tablet for the device comparability analysis); 2) to determine whether the psychometric properties of test scores were similar across the levels of conditions for each analysis; and 3) to determine whether overall test performance was similar across the levels of conditions for each analysis.

This study examined performance on 12 PARCC assessments, split evenly between mathematics and ELA/L. Students were matched on demographic variables as well as the score from the PARCC assessment in the same content area in the prior year, creating comparable samples that allowed for an unbiased comparison of performance across different conditions.

The results of the mode comparability analysis were mixed and found to be consistent with prior research. The item means suggested that items were of similar difficulty on paper and online modes. Only two items were flagged for mode effects, both of which were on the mathematics assessments. C-level differential item functioning (DIF) was present in both analyses. All the items flagged for C-level DIF in the mathematics assessments favored the online students, whereas the majority of items flagged for C-level DIF in the ELA/L assessments favored the paper students. An examination of test reliability displayed comparable reliability values between the two modes; none of the test forms were flagged for mode effects with respect to test reliability. The test-level adjustment analysis as well as the change of the paper students' performance levels after the adjustment constants were applied to the paper students' scores indicated that more scale scores were adjusted downward than were adjusted upward on the paper test form for each assessment except grades 5 and 7 mathematics. However, all adjustments were less than the minimum standard error of Theta except for grade 11 ELA/L, which was

the same as the minimum standard error of Theta. Therefore the adjustments are within measurement precision for each assessment.

The results of the device comparability study revealed consistent evidence supporting the comparability between the tablet condition (TC) and the non-tablet condition (NTC). Specifically, the item means suggested that items were similarly difficult for the TC and NTC, and none of the items were flagged for device effects. The DIF analysis revealed that none of the items had C-level DIF. Consistent with the findings at the item level, an examination of test reliability indicated that the TC and NTC test forms were similarly reliable and that none of the test forms were flagged for device effects. Furthermore, the test-level adjustment analysis as well as the change of the students' performance levels after the adjustment constants were applied did not indicate strong evidence of device effects.

The generalizability of the findings from this study may be limited due to the small sample size of both the paper students (for mode comparability) and the tablet students (for device comparability) at the high-school grades; however, it appears that high-quality matching supports the internal validity of this study's findings. For mode and device comparability, there were little to no items flagged for mode or device effects, the psychometric properties of test scores were similar across assessment conditions, and any adjustments to student performance for the paper or tablet condition were within measurement precision.

9.6 Evidence Based on Response Processes

As noted in the AERA, APA, and NCME Standards (2014), additional support for a particular score interpretation or use can be provided by theoretical and empirical evidence indicating that test takers are using the intended response processes when responding to the items in a test. This type of evidence may be gathered from interacting with test takers in order to understand what processes underlie their item responses. Evidence may also be derived from feedback provided by test proctors/teachers involved in the administration of the test and raters involved in the scoring of constructed-response items. Evidence may also be gathered by evaluating the correct and incorrect responses to short constructed-response items (e.g., items requiring a few words to respond) or by evaluating the response patterns to multi-part items.

PARCC has undertaken research investigating the quality of the items, tasks, and stimuli, focusing on whether students interact with items/tasks as intended, whether they were given enough time to complete the assessments, and the degree to which PARCC scoring rubrics allow accurate and reliable scoring. In addition, PARCC has examined the accessibility of the test for students with disabilities and English learners. This research has included examining students' understanding of the format of the assessments and the use of technology.

One such study conducted involved a series of four component studies that were conducted to evaluate the usability and effect of a drawing tool for online mathematics items. The purpose of these studies was to determine if results could support the use of the drawing tool, which is a way to expand students' ability to demonstrate their understanding and reasoning, thereby enhancing accessibility and

construct validity of the assessment. This goal is in keeping with guidance from the Common Core State Standards (CCSS) and the National Council of Teachers of Mathematics (NCTM) that students should have multiple paths and tools available to express their responses. Additionally, the drawing tool was intended to boost comparability across modes.

The first two studies (Brandt, Bercovitz, McNally, & Zimmerman, 2015; Brandt, Bercovitz, & Zimmerman, 2015) focused on evaluating the usability of the tool itself both in the general population and among students with low-vision and fine motor impairment disabilities. During these studies, detailed information regarding the functionality of the tool was collected and it was determined that the items should be tested operationally.

The third and fourth studies (Steedle & LaSalle, 2016; Minchen, LaSalle, & Boyd, 2017) involved evaluating the effect of the tool in the context of PARCC operational assessments. The third study was conducted in grade 3 and the fourth study was conducted in grades 4 and 5. To evaluate the drawing tool in context, a set of items were studied by field testing them with and without the drawing tool. The drawing tool version of each item was randomly assigned to students so that comparisons could be made. The goal was to explore the impact of the drawing tool on item performance. In general, the results showed that the drawing tool usually did not have a significant impact on performance or item statistics. Items with access to the drawing tool, however, did show longer response times for grades 4 and 5, prompting a limitation to be placed on the number of drawing tool items in each unit.

Several other PARCC research efforts have investigated questions relevant to response processes evidence. Descriptions of the research conducted can be found online.[8]

9.7 Interpretations of Test Scores

The PARCC ELA/L and mathematics scores are expressed as scale scores (both total scores and claim scores), along with performance levels to describe how well students met the academic standards for their grade level. Additionally, information on specific skills (the subclaims) is also provided and is reported as *Below Expectations*, *Nearly Meets Expectations*, and *Meets or Exceeds Expectations*. On the basis of a student's total score, an inference is drawn about how much knowledge and skill in the content area the student has acquired. The total score is also used to classify students in terms of the level of knowledge and skill in the content area as students progress in their K–12 education. These levels are called performance levels and are reported as:

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

[8] Various PARCC research is described at: https://parcc-assessment.org/research/

Students classified as either Level 4 or Level 5 are meeting or exceeding the grade level expectations. PARCC has developed performance level descriptors (PLDs) to assist with the understanding and interpretations of the ELA/L and mathematics scores (http://parcc-assessment.org/performance-levels/). Additionally, resource information is available online to educators, parents, and students (http://parcc-assessment.org/resources), which includes information on understanding and interpreting the ELA/L and mathematics score reports (https://www.isbe.net/Documents/Score_Report_Interpretation_Guide_Spring_2017.pdf http://understandthescore.org).

Section 12 of this technical report provides more information on the scale scores and the subclaim scores.

9.8 Evidence Based on the Consequences to Testing

The consequence of testing should also be investigated to support the validity evidence for the use of the PARCC assessments as the standards note that tests are usually administered "with the expectation that some benefit will be realized from the intended use of the scores" (AERA, APA, & NCME, 2014). When this is the case, evidence that the expected benefits accrue will provide support for the intended use of the scores. Evidence of the consequence of testing will also accrue with the continued implementation of the CCSS and the continued administration of the PARCC assessments.

Consequences of the PARCC tests may vary by state or by school district. For example, some states may require "passing" the PARCC assessments as one of several criteria for high school graduation, while other states/districts may not require students to "pass" the PARCC assessments for high school graduation. Additionally, some school districts may use the PARCC scores along with other information such as school grades and teacher recommendations for placing students into special programs (e.g., remedial support, gifted and talented program) or for course placement (e.g., Algebra I in grade 8). Because the consequences for the PARCC assessments can vary by each state, it is suggested that each PARCC member state provide school districts, teachers, parents, and students with information on how to interpret and use the PARCC scores. Additionally, the states should monitor how PARCC scores are used to ensure that the scores are being used as intended by PARCC.

9.9 Summary

In this section of the technical report, several aspects of validity were included, such as validity evidence based on content, the internal structure of the assessments, relationships across the content assessments, and evidence from special studies.

The PARCC item development process involved educators, assessment experts, and bias and sensitivity experts in review of text, items, and tasks for accuracy, appropriateness, and freedom from bias. PARCC conducted several studies during the item development process to evaluate the item development process (e.g., technological functionalities, answer time required, and student experiences). Additionally, items were field tested prior to the initial operational administration, and data and feedback from students, test administrators, and classroom teachers was used to improve the

operational administration of the items and to inform future item development. The multiple item and form reviews conducted by educators and studies to evaluate item administration help to ensure the integrity of the PARCC assessments.

The intercorrelations of the subclaims, the reliability analyses, and the local item dependence analyses indicated that the ELA/L and the mathematics assessments are both essentially unidimensional. Furthermore, the correlations between ELA/L and mathematics indicated that the two assessments are measuring different content. Also, the patterns of correlations for the CBT and PBT assessments were similar, indicating that the structure of the assessments were similar across the two modes.

Several studies were conducted as part of the PARCC assessment program (e.g., benchmarking study, content evaluation/alignment studies, longitudinal study, and mode and device comparability studies). The benchmarking study was conducted in support of the standard setting meeting. This study indicated students performing at or above Level 4 could be considered to be college- and career-ready or on track to readiness.

The content evaluation/alignment studies performed by the Fordham Institute and HumRRO indicate that the PARCC assessments are good to excellent matches to the CCSS in terms of content and depth of knowledge. Thus, the PARCC assessments are assessing the college- and career-readiness standards. However, the reports noted that the PARCC program could improve by adding a wider range of depth of knowledge to some of the assessments. The reports also suggested enhancing the ELA/L assessments by including a research task that requires the use of two or more sources of information.

In the longitudinal study of external validity, associations between PARCC performance levels and college-readiness benchmarks established by the College Board and ACT were used to study the claim that students who achieve Level 4 have a .75 probability of attaining at least a C in entry-level, credit-bearing, postsecondary coursework. In the first phase of the study, the relationship between PARCC and external tests was studied. Overall, results indicated that a student meeting the benchmark on the PARCC test had a high probability of making the benchmark on the external test, but the converse did not hold for students meeting the benchmark on the external test, for the majority of comparisons. These results suggest that meeting the PARCC benchmark is an indicator of academic readiness for college. In the next phase of the study, the relationship between PARCC scores and performance in first-year college courses will be explored.

The mode comparability study indicated that the comparability across modes was inconsistent across content domains and grade levels. The results of the mode comparability analysis were mixed and found to be consistent with prior research. The results of the device comparability study revealed consistent evidence supporting the comparability between the tablet condition (TC) and the non-tablet condition (NTC). In both the mode and device comparability studies, there were little to no items flagged for mode or device effects, the psychometric properties of test scores were similar across assessment conditions, and any adjustments to student performance for the paper or tablet condition were within measurement precision.

In addition to the validity information presented in this section of the technical report, other information in support of the uses and interpretations of the PARCC scores appear in the following sections:

Section 5 presents information regarding student characteristics for the spring administration of the ELA/L and mathematics administration.

Section 6 provides information concerning the test characteristics based on classical test theory.

Section 7 provides information regarding the differential item functioning (DIF) analyses.

Section 8 provides information on the test reliability (total test score and for subclaims) and includes information on the interrater reliability/agreement.

Section 12 provides detailed information concerning the scores that were reported and the cut scores for ELA/L and mathematics.

The technical report addendum provides the test taker characteristics and test reliability (total test score and for subclaims) for the 2017 fall block administration.

Section 10: IRT Calibration and Scaling in Operational Year Four

10.1 Overview

Multiple operational core forms were administered for each grade in English language arts/literacy (ELA/L) and mathematics assessments. The purpose of the item response theory (IRT) calibration and scaling was to place all operational items for a single grade/subject onto a common scale. For the ELA/L computer-based tests (CBTs), the IRT parameters were post-equated. This section describes procedures used to calibrate and scale the post-equated PARCC operational assessments. Because ELA/L paper-based tests (PBTs) and all mathematics tests were pre-equated, much of the discussion in this section will not apply; however, the parameters used to construct the conversion tables for these tests are presented in this section.

In this section of the technical report, the following topics related to IRT calibration and scaling are discussed:

Calibration:

10.2 IRT Data Preparation
10.3 Description of the Calibration Process
10.4 Model Fit Evaluation Criteria
10.5 Items Excluded from Score Reporting

Scaling:

10.6 Scaling Parameter Estimates
10.7 Items Excluded from Linking Sets
10.8 Correlations and Plots of Scaling Item Parameter Estimates
10.9 Scaling Constants
10.10 Summary Statistics and Distributions from IRT Analyses

10.2 IRT Data Preparation

10.2.1 Overview

Post-equating was performed on an early sample of the student data. The Bureau of Indian Education (BIE), the District of Columbia (DC), the Department of Defense Education Activity (DoDEA), and four states participated in the spring administration: Illinois (IL), Maryland (MD), New Jersey (NJ), and New Mexico (NM). Each state tested over multiple weeks. Student data were monitored to determine when the early equating sampling criteria were met. Student data were evaluated for the following:

1. Overall n-count, form count, and item count
2. Demographic representation
3. State representation
4. Summative scale score distribution

Based on the early equating report, using data from the prior spring PARCC administration, approximately 25 percent (75,000–90,000) of the online student data were sufficient for post-equating selected tests in grades 3 through 8. Approximately 40 percent (70,000–90,000) of the online student data were sufficient for post-equating selected tests in grades 9 through 11. The larger percentage of the student data for high school assessments was due to the high percentage of students removed for not meeting attemptedness criteria and the need to obtain student responses for each score category for the more difficult items. The results from the research study were used to determine criteria for sample size and acceptable differences between the baseline demographic distributions and the sample demographic distributions for the post-equating sample.

The resulting early equating samples for the spring administration exceeded state representation, exceeded the sample size criteria, met criteria for most of the demographic groups, and met criteria for the prior grade's PARCC performance level distributions for most of the grades/subjects. Table 10.1 lists the equating sample sizes for all tests that were post-equated.

All student response data in the early equating samples for operational items were used to create the IRT sparse data matrices for the concurrent calibration. IRT sparse data matrices combine student data across forms within administration mode. When duplicate records for a single student existed, the record with the largest raw score was included in the data file (and the other record was excluded). No student was included more than one time in the IRT sparse data matrices files.

10.2.2 Student Inclusion/Exclusion Rules

The following are the IRT valid case criteria. These criteria are the same as the student inclusion/exclusion rules used to evaluate and filter data prior to conducting the operational item analysis (IA) and differential item functioning (DIF) analyses (steps 1–5). The rules were agreed upon with PARCC and applied to the scored data used in the IRT calibration.

1. All records with an invalid form number were excluded.
2. All records that were flagged as "void" were excluded.
3. Records in which the student attempted fewer than 25 percent of the items in any unit were excluded. An item was deemed "not attempted" if it had a value of "M" (item omitted) or "Z" (item "spoiled," do not score) in the scored item response block. For example, if there were 25 items on a form and two were flawed ("Z"), those two items were not included in the numerator or denominator of the percentage-attempted calculation.
4. For students with more than one valid record, the record with the higher raw score was chosen. If the raw scores were the same, the record with the higher attemptedness rate across all operational units was chosen.
5. Records for students with administration issues or anomalies were excluded.

10.2.3 Items Excluded from IRT Sparse Matrices

Pearson conducted an initial scoring and key check. Items identified by Pearson as "spoiled" (also referred to as "do not use (DNU)") were listed and excluded from the analyses. When the IRT sparse data matrices were created, all items were included in the files unless they were marked as "spoiled" by Pearson.

10.2.4 Omitted, Not Reached, and Not Presented Items

In the student data files, "Z" was used to represent "spoiled" or "not presented" items and "M" was used to represent omitted items. Item response scores for omits were recoded as "0" in the IRT sparse matrix files *unless* the omitted item was a "not reached" item. Not reached items are omitted items at the end of the test or unit—items that the student probably did not reach or try to answer. Not reached items were recoded from "M" to "N" (i.e., not presented) in the IRT sparse matrix files, if all items from that point to the end of the form are "M" or "Z." Not reached items were counted as missing or no response, and therefore did not contribute to the item statistics.

10.2.5 Quality Control of the IRT Sparse Matrix Data Files

The IRT sparse data matrices were created by the primary analysts and replicators from Pearson and HumRRO. The matrices were checked for quality and accuracy by comparing the number of students (counts), item category frequencies, and item statistics (e.g., average item score values) between Pearson and HumRRO. Since the same inclusion rules for students were used, all counts, category frequencies, and statistics for all items matched. All discrepancies in counts were resolved. The programs used to create the IRT statistics were independent, so the QC procedure involved parallel computing. Table 10.1 shows the counts and number of items in the CBT IRT sparse data matrices for each grade in ELA/L.

Table 10.1 Counts and Number of Items in the ELA/L IRT Calibration Files

Grade	Count	Items
3	204,307	47
4	258,143	59
5	242,986	56
6	216,642	60
7	236,106	61
8	198,863	58
9	105,382	63
10	138,073	60
11	85,338	58

10.3 Description of the Calibration Process

The IRT calibrations were performed only on the ELA/L CBT tests. The form-to-form linking is established through internal and external common items selected during test construction to represent the blueprint.

10.3.1 Two-Parameter Logistic/Generalized Partial Credit Model

The operational IRT analyses were conducted by both Pearson and HumRRO. The operational items in the IRT sparse data matrix were concurrently calibrated with the two-parameter logistic/generalized partial credit model (2PL/GPC: Muraki, 1992). The 2PL/GPC is denoted

$$p_{im}(\theta_j) = \frac{exp\left[\sum_{k=0}^{m} Da_i(\theta_j - b_i + d_{ik})\right]}{\sum_{v=0}^{M_i-1} exp\left[\sum_{k=0}^{v} Da_i(\theta_j - b_i + d_{ik})\right]} \quad (10\text{-}1)$$

where $a_i(\theta_j - b_i + d_{i0}) \equiv 0$; $p_{im}(\theta_j)$ is the probability of a test taker with θ_j getting score m on item i; D is the IRT scale constant (1.7); a_i is the discrimination parameter of item i; b_i is the item difficulty parameter of item i; d_{ik} is the k^{th} step deviation value for item i; M_i is the number of score categories of item i with possible item scores as consecutive integers from zero to $M_i - 1$; v sequences through each response category through $M_i - 1$. IRT calibrations might also use a guessing parameter in special cases, if needed.

10.3.2 Treatment of Prose Constructed-Response (PCR) Tasks

The prose constructed-response (PCR) tasks were calibrated at the trait score level (and not as aggregated scores). To address the issue of local independence related to PCR items, a single-calibration "model" approach was used. When sample sizes were large (i.e., greater than 10,000 test takers), the data were manipulated using random assignment, by selecting one of the two traits for each PCR item for each student. Then one calibration was run so that all trait parameters were independently estimated. When sample sizes were smaller (i.e., field-test samples), a multiple-calibration "model" approach was used. In this alternative approach, the same data set was calibrated two times, each trait represented in one of the two data sets for all students. Then the PCR traits were scaled onto the base scale using non-PCR items as anchor items. These two trait calibration approaches addressed the issue of local dependence while allowing for the accurate calculation of claim scores and the proper weighting of traits in the summative scale scores.

10.3.3 IRT Item Exclusion Rules (Before Calibration)

In addition to checking IRT data for accuracy, Pearson conducted item analyses (IA) to identify items that were not performing as expected and should be considered for removal from calibration and score reporting. The following are the criteria Pearson used to flag extremely problematic items to be dropped from calibration. All "non-spoiled" items were included in the IRT data matrices; however, the IRTPRO calibration software (Cai, Thissen, & du Toit, 2011) control files were used to exclude from calibration items flagged for the following reasons:

1. A weighted polyserial correlation less than 0.0
2. An average item score of 0.0
3. 100 percent of the students having the same item score, such as:
   a. 100 percent omitted the item
   b. 100 percent received the same score
   c. 100 percent of the responses were at the same score after collapsing score categories due to low frequencies, or
   d. 100 percent of the responses were not presented or not reached
4. Insufficient sample sizes for the selected IRT model combinations (i.e., 300 for the 2PL/GPC)
5. High omit rates (i.e., greater than 50 percent) on one or more forms (usually an indication that an item may not be functioning correctly on all forms)

A master list of all problematic items before and after calibration was maintained and all flagged and potentially flawed items were brought to the PARCC Priority Alert Task Force (consisting of New Meridian Corporation and participating State Leads for PARCC member states) for content and statistical reviews. Ultimately, the decisions about whether to keep or exclude an item from score reporting was made by the PARCC Priority Task Force.

10.3.4 IRTPRO Calibration Procedures and Convergence Criteria

The data were calibrated concurrently across forms using the 2PL/GPC model combination. The primary goal was to place the operational item data within each content area and grade/subject on a common difficulty scale. The following are the steps used to calibrate the operational item response data:

1. Using the IRT sparse data matrices, concurrent calibrations were conducted using commercially available IRTPRO for Windows (version 4.2) on CBT data within each grade/subject.
2. IRTPRO Calibration Settings: The logistic partial credit model was specified using the scale constant of 1.0. The prior distributions for latent traits were set to a mean of zero and a standard deviation of one. The number of quadrature points used in the estimation was set to 49. And the slope starting value was set/updated before each run.
3. Each IRTPRO run was inspected for convergence and for any unexpected item-parameter estimates. The PRIORS command in IRTPRO provided a prior on IRT parameters to constrain the calibration so that convergence was more likely. Specifically, option “Guessing[0]” indicated that the prior is placed on the lower asymptote for the 3-PL model, and a normal distribution for the priors with mean of -1.4 and standard deviation 1. For these items, an inspection of item-level statistics and modal-data fit plots were sufficient to ensure that item parameters were acceptable if convergence was reached. Item information functions from the IRTPRO output may also be reviewed. Pearson verified that the maximum number of EM (expectation-maximization) cycles was not reached (which indicated the program did not converge).

4. To convert IRTPRO item parameters to the commonly used logistic parameter presentation (called new item parameters), the following formula was used since IRTPRO uses 1.0 for a scaling constant. There was no need to transfer *b*- and *c*-parameters from IRTPRO output. Please note that all unscaled and scaled item parameters were kept on the theta scale. For 2PL models:

   New *a*-parameter: $a_{new} = \frac{a_{irtpro}}{1.7}$ (10-2)

5. Pearson reported any need for item-calibration decisions, including convergence issues and extreme parameter estimates, along with proposed resolutions, to the Priority Alert Task Force. Anticipated resolutions included fixing the slope parameters to a minimum .10 value, fixing the guessing parameter to a rational value (1 divided by number of options), and fixing the difficulty parameters at an upper or lower bound, depending on the nature of the problem. If extreme *b*-parameter values were observed (e.g., > 100) and the *a*-parameter values for these items were low (i.e., < 0.10), it was recommended that the prior for the *a*-parameter be set to 0.5.

6. Dropping an item from further processing or dropping an item and rerunning IRTPRO was performed only if it was needed after communication with HumRRO and the Priority Alert Task Force.

7. Inspection of model-data fit plots was helpful in deciding parameter constraints and acceptability of parameter fit. Documentation of each step, after resolution of any issues, was provided by Pearson to New Meridian Corporation, HumRRO, and Measured Progress.

10.3.5 Calibration Quality Control

To ensure IRT calibrations and conversion tables were produced accurately, HumRRO replicated the IRT calibrations and the generation of the score conversion tables. Both Pearson and HumRRO used the same calibration software, IRTPRO. Meetings were held, as needed, so that Pearson and HumRRO could provide status reports and discuss issues related to the IRT work. Measured Progress performed independent quality control comparisons between the Pearson and HumRRO item parameter estimates to identify any differences.

Specifically, Measured Progress completed the following quality control analyses/comparisons:

1. Made sure all items were treated the same way (e.g., if Pearson collapsed a category, made sure HumRRO collapsed the category in the same way for the item)

2. Compared IRT item parameter estimates by Pearson and HumRRO (i.e., IRT *a*-, *b*-, and *d*-parameter estimates)

3. Compared the scaling constants for the common item linking sets

4. Compared scaled CBT parameter estimates generated by Pearson and HumRRO
5. Compared all conversion tables produced by Pearson and HumRRO to make sure they were accurate

Measured Progress prepared reports documenting their findings. Exact matches were found between all Pearson and HumRRO conversion tables before scores were reported.

10.4 Model Fit Evaluation Criteria

The usefulness of IRT models is dependent on the extent to which they effectively reflect the data. As discussed by Hambleton, Swaminathan, and Rogers (1991), "The advantages of item response models can be obtained only when the fit between the model and the test data of interest is satisfactory. A poorly fitting IRT model will not yield invariant item and ability parameters" (p. 53).

After convergence was achieved for each IRT data set, the IRT model fit was evaluated by doing the following:

1. Calculating the Q_1 statistic and comparing it to a criterion score
2. Calculating the G^2 statistic and comparing it to a criterion score
3. Reviewing graphical output for all items

The Q_1 statistic (Yen, 1981) was used as an index of correspondence between observed and expected performance. To compute Q_1, first the estimated item parameters and student response data (along with observed item scores) were used to estimate student ability ($\hat{\theta}$). Next, expected performance was computed for each item using students' ability estimates in combination with estimated item parameters. Differences between expected item performance and observed item performance were then compared at 10 intervals across the range of student achievement (with approximately the same number of students per interval). Q_1 was computed as a ratio involving expected and observed item performance. Q_1 is interpretable as a chi-squared (χ^2) statistic, which can be compared to a critical chi-squared value to make a statistical inference about whether the data (observed item performance) were consistent with what might be observed if the IRT model was true (expected item performance). Q_1 is not directly comparable across different item types because items with different numbers of IRT parameters have different degrees of freedom (*df*). For that reason, a linear transformation (to a *Z*-score, Z_{Q_1}) was applied to Q_1. This transformation also made item fit results easier to interpret and addressed the sensitivity of Q_1 to sample size.

To evaluate item fit, Yen's Q_1 statistic was calculated for all items. Q_1 is a fit statistic that compares observed and expected item performance. MAP (maximum *a posteriori*) estimates from IRTPRO were used as student ability estimates. For dichotomous items, Q_1 was computed as

$$Q_{1i} = \sum_{j=1}^{J} \frac{N_{ij}(O_{ij} - E_{ij})^2}{E_{ij}(1 - E_{ij})} \qquad (10\text{-}3)$$

where N_{ij} was the number of examinees in interval (or group) *j* for item *i*, O_{ij} was the observed proportion of the students for the same cell, and E_{ij} was the expected proportions of the students for the same interval. The expected proportion was computed as

$$E_{ij} = \frac{1}{N_{ij}} \sum_{a \in j}^{N_{ij}} P_i(\hat{\theta}_a) \quad (10\text{-}4)$$

where $P_i(\hat{\theta}_a)$ was the item characteristic function for item *i* and students *a*. The summation is taken over students in interval *j*.

The generalization of Q_1 for items with multiple response categories is

$$Gen\ Q_{1i} = \sum_{j=1}^{10} \sum_{k=1}^{m_i} \frac{N_{ij}(O_{ikj} - E_{ikj})^2}{E_{ikj}} \quad (10\text{-}5)$$

Where

$$E_{ikj} = \frac{1}{N_{ij}} \sum_{a \in j}^{N_{ij}} P_{ik}\ (\hat{\theta}_a) \quad (10\text{-}6)$$

Both Q_1 and generalized Q_1 results were transformed to ZQ_1 and were compared to a criterion $ZQ_{1,crit}$ to determine acceptable fit. The conversion formula was

$$ZQ_1 = \frac{Q_1 - df}{\sqrt{2df}} \quad (10\text{-}7)$$

And

$$ZQ_{1,crit} = \frac{N}{1500} * 4 \quad (10\text{-}8)$$

where *df* is the degrees of freedom. The degrees of freedom is equal to the number of independent cells less the number of independent item parameters. For example, the degrees of freedom for polytomous items equals [10 × (number of score categories–1) – number of independent item parameters]. For the GPCM, the number of independent item parameters equals 1 (for the *a* parameter) plus the number of step values (e.g., for an item scored 0, 1, 2, 3: there are 3 independent step values—the *b* parameter is simply the mean of the step values and is not, therefore, independent).

If Q_1 is found to be excessively sensitive (i.e., a large number of items are flagged for poor fit, even if their item fit plots look reasonable), a likelihood-ratio chi-squared statistic may be computed for each item (Muraki and Bock, 1997):

$$G_i^2 = 2\sum_{w=1}^{W_i}\sum_{h=1}^{m_i} r_{wih} \ln \frac{r_{wih}}{N_{wi}P_{ih}(\bar{\theta}_w)} \quad (10\text{-}9)$$

where r_{wih} is the observed frequency of the h^{th} categorical response to item *i* in interval *w*, N_{wi} is the number of students in interval *w* for item *i*, $P_{ih}(\bar{\theta}_w)$ is the expected probability of observing the h^{th} categorical response to item *i* for the mean θ in interval *w*, and W_i is the number of intervals remaining after neighboring intervals are merged, if necessary, to avoid expected values, $N_{wi}P_{ih}(\bar{\theta}_w)$, less than 5. To conduct a standard hypothesis test, the number of degrees of freedom is equal to the number of intervals, W_i, multiplied by $m_i - 1$.

As an alternative to a traditional hypothesis test, the "contingency coefficient" (effect size; Barton & Huynh, 2003) was computed:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \quad (10\text{-}10)$$

In this formula, G_i^2 was substituted for χ^2, and *N* is the sample size on which the IRT parameters were estimated. According to Cohen (1988, pp. 224-225), values of *C* below .10 are considered insignificant, .10+ small, .287+ medium, and .447+ large. A threshold of .35 is recommended (i.e., flag items for which C ≥ .35).

An item fit-plot was created for each item. Item-fit plots show observed and expected average scores for each interval (e.g., figure below).

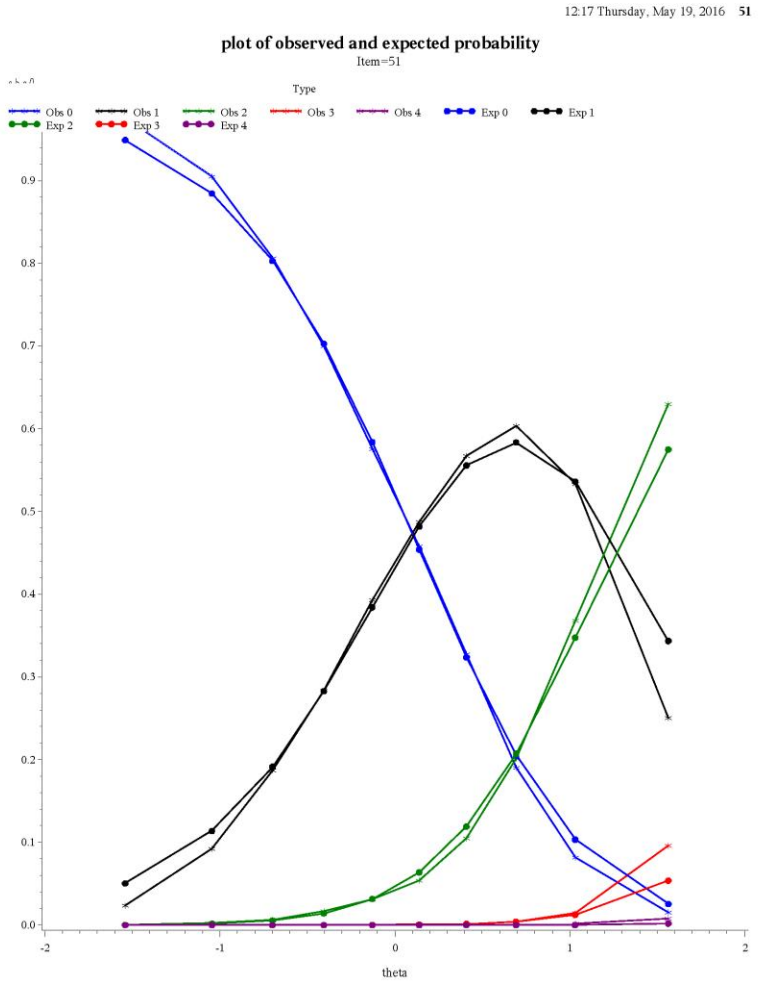

Figure 10.1 An example ELA/L 5-Category Item, 2 PL/GPC Model, n-count 44,658, Q_1=1266.64, ZQ_1=147.21 and a criterion ZQ_1,crit= 237.02

10.5 Items Excluded from Score Reporting

As mentioned previously, after calibration and model fit evaluation were completed, a master list of all problematic items was compiled and potentially flawed items were brought to the PARCC Priority Alert Task Force. The Task Force reviewed each item, its content, and the statistical properties, and made decisions about whether to include the item in the operational scores. Sometimes, an item was rejected because it appeared to have content issues, and sometimes an item was excluded because it had unreasonable IRT parameters or showed extremely poor IRT model fit. Ultimately the decision about whether to keep or exclude each flagged item was made by the Task Force.

10.5.1 Item Review Process

The following are the types of problematic items that were brought to the PARCC Priority Alert Task Force for evaluation and an "include or exclude" determination was made:

1) Extremely difficult items (e.g., an item with a p-value less than 0.02)
2) Items with low *a*-parameter estimates (e.g., slope less than 0.10)
3) Items flagged for subgroup DIF

Again, the primary goal was to minimize the number of items dropped from the operational test forms. An equally important goal was to not advantage or disadvantage any test takers.

10.5.2 Count and Percentage of Items Excluded from Score Reporting

All items that did not have IRT item parameter estimates or were DNU'd after the IRT calibration were excluded from the student operational scores and the conversion tables used for score reporting. For ELA/L and mathematics, at most 2 percent of the items were excluded from score reporting for all grades/subjects. Figure 10.2 demonstrates an item that was flagged during the calibrations and item fit review for poor statistics (*a*-parameter=0.06 and *b*-parameter=15.75) and poor fit (e.g. Q_1= 2682.75, ZQ_1= 457.17 and a criterion $ZQ_{1,crit}$= 268.33). This item was reviewed by the Priority Alert Task Force and removed from scoring.

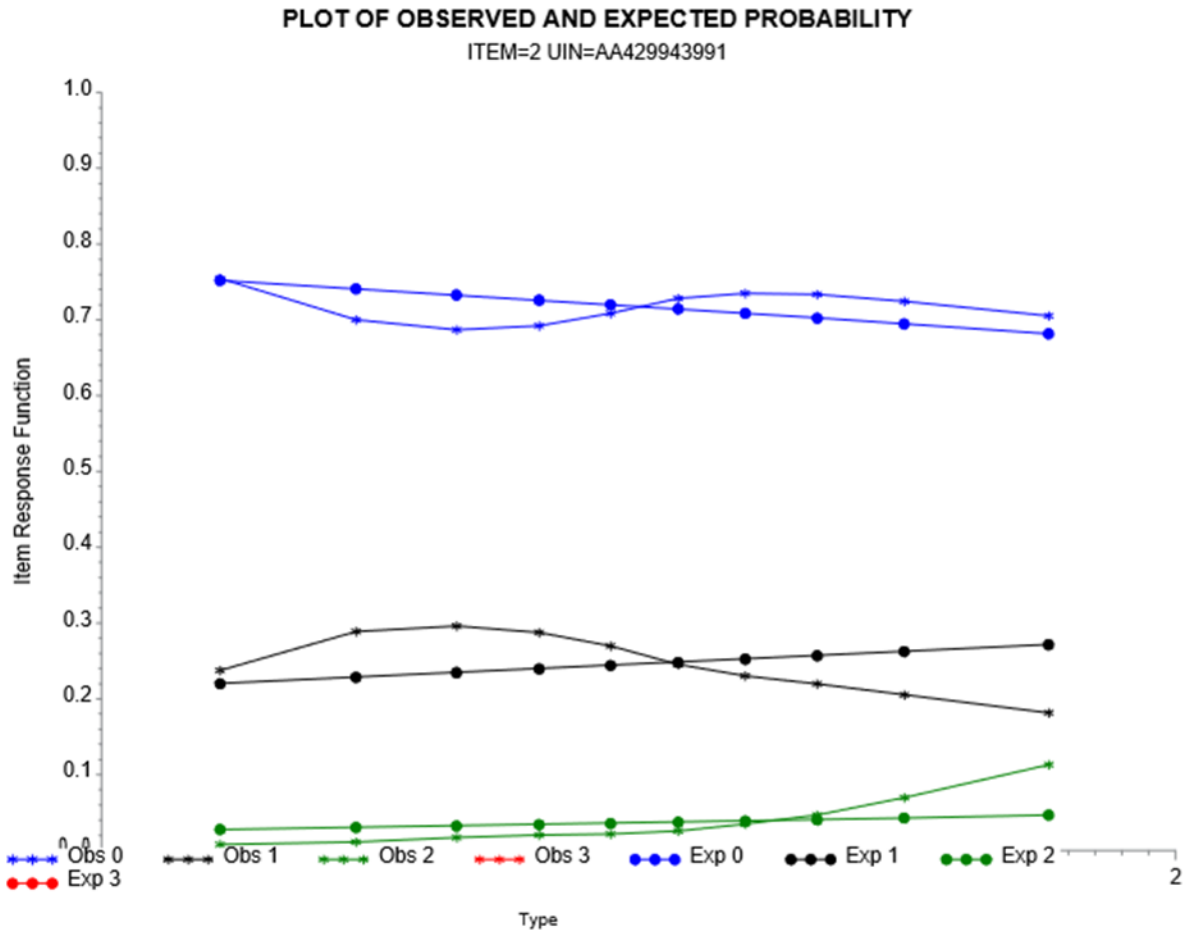

Figure 10.2 An example ELA/L 3-Category Item, 2 PL/GPC Model, n-count 100,622

Table 10.2 presents the count and percentage of CBT items excluded from IRT calibration along with the reasons the items were excluded.

Table 10.2 Number and Percentage of ELA/L Items Excluded from IRT Calibration

Grade	Total *n* of CBT Items	*n* of CBT Items Excluded	Percent Excluded	Reason Excluded			
				Small Sample Size	Poor IA Stats	Did Not Calibrate	Other
3	47	1	2%		Yes		Poor IRT stats
4	59	0	0%				
5	56	0	0%				
6	60	0	0%				
7	61	0	0%				
8	58	1	2%		Yes		
9	63	0	0%				
10	60	0	0%				
11	58	0	0%				

10.6 Scaling Parameter Estimates

Year-to-year linking was performed on all ELA/L CBTs to transform IRT parameters to the base IRT scale. The linking analyses included common-item sets. The linking methodology was based on the Stocking and Lord (1983) test characteristic curve scale transformation method.

Year-to-year linking transforms IRT parameters from different years (or administrations) onto the same underlying IRT scale. Due to the PARCC test redesign and updates to the trait scoring for the PCRs, the spring 2016 online IRT scale was established as the base IRT scale.

The entire item bank was previously transformed to the base IRT scale through a common-item linking set consisting of spring 2016 operational online items that were operational or field-test items on the spring 2015 online administration. The details for this procedure can be found in the 2016 PARCC Technical Report.

HumRRO not only conducted independent calibrations of item response data using IRTPRO scaling software, they also used STUIRT (Kim & Kolen, 2004) software to transform their IRTPRO item parameter estimates onto the IRTPRO scales for each grade/subject. HumRRO's scaling constants were compared to those generated by Pearson and found to exactly match. As described in Section 10.3.4, Measured Progress independently verified that the same items were excluded from the linking sets, and compared transformed parameter estimates by Pearson and HumRRO.

10.7 Items Excluded from Linking Sets

Linking is an iterative process. Robust *Z* (Huynh & Meyer, 2010) and Weighted Root Mean Square Difference (WRMSD) were used to identify outlier items in the linking sets. The following rules were used to identify items for possible exclusion from the linking sets:

1. Exclude an item from the common-item set if different amounts of collapsing resulted in a different number of response categories across modes or versions.
2. Flag and potentially exclude an item from the common-item set if the online or paper weighted polyserial correlation, based on the item analysis, was less than 0.10.
3. Exclude items dropped by the PARCC Priority Alert Task Force (i.e., due to content or parameter estimation issues).
4. Exclude an item if the scoring rules changed.

After removing items due to item performance issues as described above, the following steps were performed:

5. Implement the Robust *Z* approach to see if any common items are flagged.
6. Run the initial Stocking and Lord procedure using the STUIRT software.
7. Calculate WRMSD and check to see if any common items exceed the threshold.
8. Re-run STUIRT after removing the items flagged by Robust *Z* and WRMSD.
9. Compare the slopes and intercepts from steps 2 and 4.

SAS code was developed to calculate WRMSD, Robust *Z*, to compare the item characteristic curves (ICCs) across modes, and to identify items for possible removal from the linking sets. Table 10.3 lists the flag criteria for the WRMSD. (*Note: these values were originally developed as part of the 2014 PARCC field-test analysis.*).

Table 10.3 WRMSD Flagging Criteria for Inspection and Possible Removal of Linking Items

Categories	Points	WRMSD/ Points	WRMSD
2	1	0.100	0.100
3	2	0.075	0.150
4	3	0.075	0.225
5	4	0.075	0.300
6	5	0.075	0.375
7	6	0.075	0.450
>=8	>= 7	0.090	0.999

When inspecting items flagged for exclusion from the linking sets, content representation was also considered to avoid removing large numbers of items from the same subclaim. Table 10.4 presents the total number of common items, items excluded from the year-to-year linking sets, and items kept in the

linking sets for each grade for ELA/L. The final number of linking items ranged from 15 (in grade 4) to 22 (in grade 5). Grades 3, 9, and 11 had the largest number of items removed from the linking sets due to Robust *Z* for the *a*-parameter and *b*-parameter, some of which were also flagged for high WRMSD.

Table 10.4 Number of ELA/L Items Excluded from the Year-to-Year Linking Sets

Grade	Total *n* of Common Items	Number Excluded	Final Number in Linking Set	Number of Excluded Items by Reason for Exclusion			
				Low Polyserial	Robust *Z* IRT_a	Robust *Z* IRT_b	High WRMSD
3	21	4	17	0	0	4	1
4	18	3	15	0	2	1	0
5	22	0	22	0	0	0	0
6	21	2	19	0	2	0	0
7	21	3	18	0	0	3	1
8	20	3	17	0	3	0	0
9	24	4	20	0	2	2	0
10	20	2	18	0	0	2	0
11	22	5	17	0	0	5	1

*Note: WRMSD did not flag any additional items for removal from the common item sets.

10.8 Correlations and Plots of Scaling Item Parameter Estimates

Once the final group of items for each linking set was determined, the *a*- and *b*-parameter estimates were plotted and the correlation between the *a*-parameter estimates and the *b*-parameter estimates were calculated. Table 10.5 presents the number of linking items, score points of the linking items, and the correlation of the *a*- and *b*-parameter estimates across years.

Table 10.5 Number of Items, Number of Points, and Correlations for ELA/L Year-to-Year Linking Items

	Number		Parameter Correlations	
Grade	**Items**	**Points**	***a*-**	***b*-**
3	17	40	0.9819	0.9994
4	15	38	0.9955	0.9975
5	22	52	0.9932	0.9967
6	19	47	0.9901	0.9945
7	18	45	0.9730	0.9966
8	17	43	0.9990	0.9951
9	20	46	0.9987	0.9798
10	18	45	0.9810	0.9805
11	17	43	0.9830	0.9496

Figures 10.3 to 10.4 are a selection of plots of the *a*- and *b*-parameter estimates for linking items for the year-to-year linking. ELA/L grade 8 (Figures 10.3 and 10.4) is an example of the year-to-year linking. For each plot, the *x*-axis is the original (reference) parameter and the *y*-axis is the new parameter after applying the scaling constants.

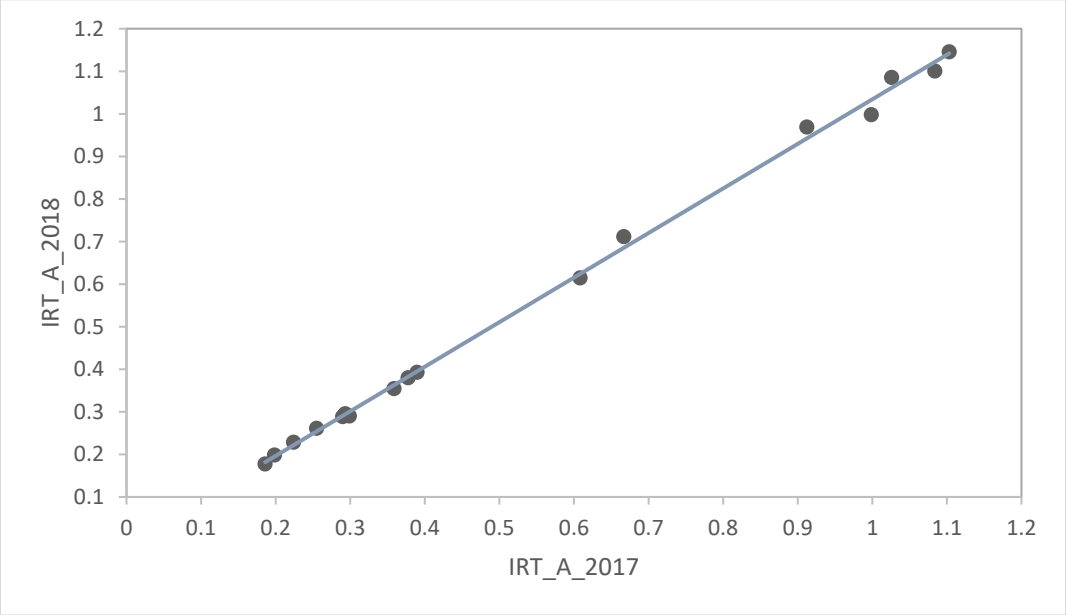

Figure 10.3 ELA/L Grade 8 Transformed New *a*- vs. Reference *a*-Parameter Estimates for Year-to-Year Linking

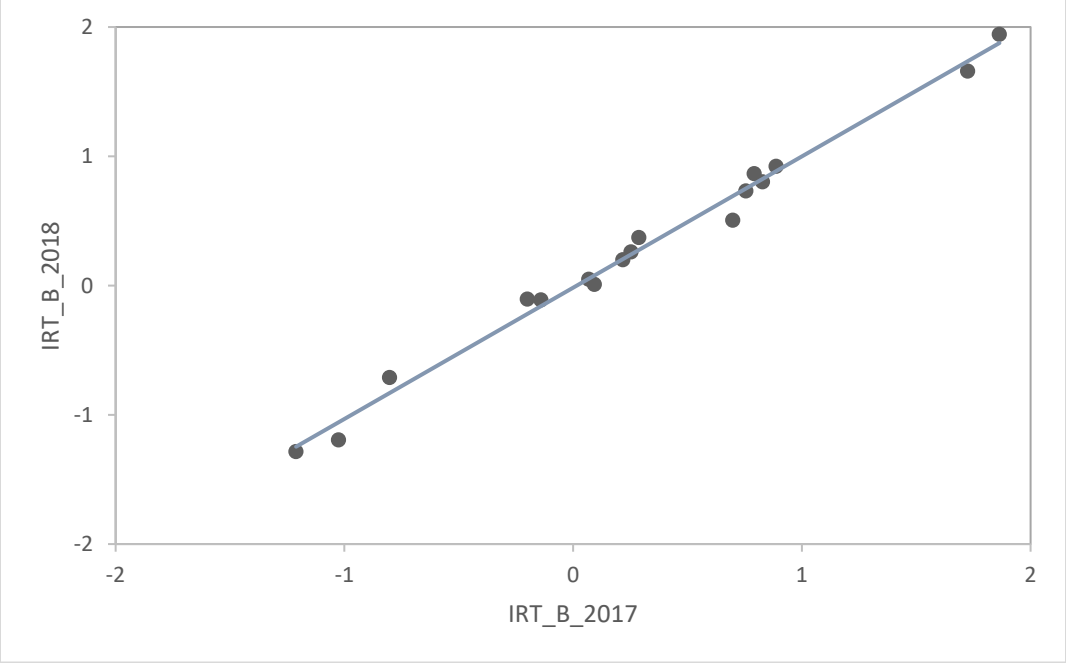

Figure 10.4 ELA/L Grade 8 Transformed New *b*- vs. Reference *b*-Parameter Estimates for Year-to-Year Linking

10.9 Scaling Constants

Table 10.6 presents the slope and intercept scaling constants for ELA/L for the year-to-year linking, derived from STUIRT (Kim & Kolen, 2004) using the Stocking and Lord (1983) test characteristic curve procedure. The slopes and intercepts are similar. The slopes range from 1.0435 to 1.0883, and the intercepts range from -0.0964 to 0.2713.

Table 10.6 Scaling Constants Spring 2017 to Spring 2018 for ELA/L

	Spring 2017 to Spring 2018	
Grade/Subject	**Slope**	**Intercept**
3	1.0435	0.0919
4	1.0795	0.1565
5	1.0883	0.1074
6	1.0533	0.1248
7	1.0727	0.1204
8	1.0447	-0.0964
9	1.0665	0.2713
10	1.0500	0.1833
11	1.0494	0.0692

10.10 Summary Statistics and Distributions from IRT Analyses

Tables 10.7 through 10.15 present summary statistics for the IRT (*b*- and *a*-) parameter estimates, the standard errors (SEs) of the parameter estimates, and the IRT model fit values (chi-square and adjusted fit) for ELA/L CBTs. For all other tests, which were pre-equated, the summary statistics for the parameter estimates and the administration from which the parameter was first obtained are presented. The information is provided by content area (ELA/L and mathematics) and by mode (CBT and PBT) for all items at each grade level or course. The summary statistics shown include the total number of items and score points, along with the mean, standard deviation (SD), minimum, and maximum.

10.10.1 IRT Summary Statistics for English Language Arts/Literacy

Table 10.7 shows the *b*- and *a*-parameter estimates for the ELA/L CBT assessments. Table 10.8 presents the standard errors (SE) of the estimates, and Table 10.9 provides model fit information for the ELA/L CBT assessments. IRT summary statistics are provided in Appendix 10 for ELA/L for all items, reading-only, and writing-only. Table 10.10 shows the summary of the parameters estimates, and Table 10.11 shows the source year for the item statistics for each of the ELA/L PBT assessments that were pre-equated.

Table 10.7 CBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

			Summary of *b* Estimates				Summary of *a* Estimates			
Grade	No. of Items	No. of Score Points	Mean	SD	Min	Max	Mean	SD	Min	Max
3	46	102	0.56	0.99	-1.35	3.13	0.55	0.27	0.16	1.07
4	59	131	0.39	1.19	-5.92	2.01	0.46	0.23	0.13	0.95
5	56	125	0.59	1.10	-1.70	3.59	0.46	0.27	0.13	1.02
6	60	135	0.59	0.91	-1.54	2.67	0.49	0.22	0.20	1.03
7	61	137	0.46	1.02	-2.29	4.13	0.48	0.31	0.14	1.27
8	57	129	0.39	0.98	-1.42	4.56	0.47	0.28	0.15	1.24
9	63	141	0.52	0.74	-0.99	2.39	0.53	0.33	0.17	1.44
10	60	135	0.90	0.97	-0.90	4.03	0.45	0.31	0.13	1.24
11	58	131	0.95	0.77	-0.67	2.80	0.47	0.23	0.17	1.10

Table 10.8 CBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade

			SE of *b* Estimates				SE of *a* Estimates			
Grade	No. of Items	No. of Score Points	Mean	SD	Min	Max	Mean	SD	Min	Max
3	46	102	0.005	0.002	0.002	0.011	0.005	0.002	0.002	0.011
4	59	131	0.003	0.002	0.001	0.008	0.003	0.002	0.001	0.008
5	56	125	0.004	0.002	0.002	0.009	0.004	0.002	0.002	0.009
6	60	135	0.004	0.002	0.002	0.010	0.004	0.002	0.002	0.010
7	61	137	0.004	0.003	0.002	0.012	0.004	0.003	0.002	0.012
8	57	129	0.004	0.003	0.002	0.013	0.004	0.003	0.002	0.013
9	63	141	0.006	0.004	0.003	0.019	0.006	0.004	0.003	0.019
10	60	135	0.005	0.003	0.002	0.015	0.005	0.003	0.002	0.015
11	58	131	0.006	0.003	0.003	0.017	0.006	0.003	0.003	0.017

Table 10.9 CBT IRT Model Fit for All Items for ELA/L by Grade

			G^2				Q_1			
Grade	No. of Items	No. of Score Points	Mean	SD	Min	Max	Mean	SD	Min	Max
3	46	102	2705.6	2157.7	523.0	8368.7	2575.7	2092.2	513.3	8253.3
4	59	131	3760.9	3306.8	587.0	14921.7	3731.2	3514.4	584.9	16584.2
5	56	125	2916.0	2198.0	713.2	9495.7	2833.6	2155.9	686.0	9588.9
6	60	135	2938.5	2513.7	598.2	13431.5	2777.8	2360.1	556.1	11539.3
7	61	137	2105.7	1670.9	251.8	8072.7	2030.2	1648.0	241.0	7673.3
8	57	129	2677.9	2206.6	150.1	9687.4	2574.3	2249.4	146.7	11232.3
9	63	141	1359.6	1127.8	232.4	5524.6	1309.0	1165.9	221.9	5277.8
10	60	135	1933.3	1617.9	185.1	9157.8	1837.2	1581.0	186.5	8347.6
11	58	131	1021.0	658.5	169.2	3545.0	933.8	605.5	151.1	3488.6

Table 10.10 PBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

			Summary of *b* Estimates				Summary of *a* Estimates			
Grade	No. of Items	No. of Score Points	Mean	SD	Min	Max	Mean	SD	Min	Max
3	25	56	0.73	1.05	-1.08	3.44	0.55	0.28	0.11	1.02
4	34	76	0.55	0.83	-1.08	1.77	0.45	0.23	0.13	0.95
5	34	76	0.95	0.99	-1.34	3.59	0.42	0.25	0.13	0.99
6	34	77	0.37	0.82	-1.09	1.86	0.50	0.22	0.21	1.02
7	34	77	0.42	0.77	-1.70	1.60	0.49	0.30	0.17	1.23
8	34	77	0.20	0.97	-1.42	2.38	0.50	0.24	0.17	1.15
9	34	77	0.48	0.70	-1.04	2.06	0.49	0.24	0.18	1.22
10	34	77	1.05	0.93	-0.43	4.03	0.45	0.32	0.13	1.19
11	34	77	1.30	1.00	-0.67	4.21	0.43	0.23	0.17	1.10

Table 10.11 PBT IRT Parameter Distribution by Year for All Items for ELA/L by Grade

Grade	ALL	2014	2015	2016	2017	2018
3	25	0	3	1	8	13
4	34	1	5	12	2	14
5	34	0	2	9	3	20
6	34	0	10	2	8	14
7	34	0	4	16	2	12
8	34	0	4	9	8	13
9	34	0	5	19	4	6
10	34	0	0	6	3	25
11	34	0	12	8	1	13

10.10.2 IRT Summary Statistics for Mathematics

Tables 10.12 and 10.14 show the *b*- and *a*-parameter estimates for the CBT and PBT mathematics assessments, respectively. Tables 10.13 and 10.15 show the source year for the item statistics for each of the assessments. IRT summary statistics are provided in Appendix 10 for mathematics for all items, single-select multiple-choice items, constructed-response items, and subclaims.

Table 10.12 CBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

			Summary of *b* Estimates				Summary of *a* Estimates			
Grade	No. of Items	No. of Score Points	Mean	SD	Min	Max	Mean	SD	Min	Max
3	86	132	-0.20	1.09	-2.91	2.09	0.76	0.22	0.31	1.29
4	80	132	-0.12	1.04	-2.61	2.81	0.69	0.20	0.22	1.36
5	80	132	0.10	1.22	-3.06	4.53	0.64	0.23	0.23	1.31
6	76	132	0.42	1.01	-2.55	2.75	0.70	0.24	0.24	1.35
7	76	132	0.73	1.00	-1.39	2.93	0.74	0.26	0.25	1.39
8	72	132	0.95	1.19	-1.83	3.97	0.60	0.22	0.24	1.18
A1	82	157	1.38	1.08	-1.16	4.16	0.66	0.26	0.11	1.29
GO	84	160	1.15	0.93	-1.49	3.53	0.70	0.26	0.23	1.68
A2	81	161	1.43	0.96	-1.41	3.42	0.62	0.27	0.20	1.26
M1	42	81	1.22	0.91	-0.50	2.80	0.59	0.19	0.20	1.07
M2	41	80	1.90	1.45	-0.74	5.99	0.53	0.27	0.12	1.18
M3	40	81	1.27	1.08	-2.27	4.28	0.59	0.27	0.16	1.27

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 10.13 CBT IRT Parameter Distribution by Year for All Items for Mathematics by Grade/Subject

Grade	ALL	2014	2015	2016	2017
3	86	4	26	17	39
4	80	3	28	21	28
5	80	3	24	23	30
6	76	1	22	21	32
7	76	0	22	22	32
8	72	3	21	21	27
A1	82	1	30	24	27
A2	81	1	10	19	51
GO	84	1	27	19	37
M1	42	1	6	12	23
M2	41	0	0	41	0
M3	40	0	0	40	0

Table 10.14 PBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

			Summary of *b* Estimates				Summary of *a* Estimates			
Grade	No. of Items	No. of Score Points	Mean	SD	Min	Max	Mean	SD	Min	Max
3	42	65	-0.30	1.17	-3.03	1.58	0.77	0.23	0.28	1.29
4	40	66	-0.12	1.13	-2.69	1.93	0.73	0.18	0.33	1.09
5	40	66	0.19	0.93	-1.43	2.65	0.66	0.24	0.22	1.31
6	38	66	0.38	0.92	-2.25	1.73	0.70	0.24	0.26	1.35
7	38	66	0.65	0.99	-1.66	2.24	0.73	0.27	0.28	1.39
8	36	66	0.80	1.13	-2.29	2.35	0.64	0.22	0.24	1.08
A1	42	81	1.27	1.10	-1.16	3.13	0.64	0.24	0.21	1.16
GO	43	81	1.03	1.07	-1.98	2.68	0.65	0.29	0.23	1.68
A2	41	81	1.52	0.86	-0.44	3.42	0.60	0.25	0.24	1.08
M1	42	81	1.19	0.94	-0.77	3.36	0.51	0.19	0.20	0.99
M2	42	80	1.67	1.14	-0.50	4.87	0.52	0.24	0.18	1.15
M3	39	80	1.31	1.33	-2.27	4.61	0.56	0.26	0.13	1.24

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 10.15 PBT IRT Parameter Distribution by Year for All Items for Mathematics by Grade/Subject

Grade	ALL	2014	2015	2016	2017
3	42	2	15	6	19
4	40	0	18	13	9
5	40	1	13	10	16
6	38	0	14	10	14
7	38	0	13	9	16
8	36	1	11	10	14
A1	42	2	13	14	13
A2	41	0	8	18	15
GO	43	3	16	15	9
M1	42	1	14	13	14
M2	42	0	22	20	0
M3	39	0	14	25	0

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Section 11: Performance Level Setting

11.1 Performance Standards

Performance standards relate levels of performance on an assessment directly to what students are expected to learn. This is done by establishing threshold scores that distinguish between performance levels. Performance level setting (PLS) is the process of establishing these threshold scores that define the performance levels for an assessment.

11.2 Performance Levels and Policy Definitions

For the PARCC assessments, the performance levels are

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

More detailed descriptions of each performance level, known as policy definitions, are:

Level 5: Exceeded expectations
Students performing at this level **exceed academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3–10: Students performing at this level **exceed academic expectations** for the knowledge, skills, and practices contained in the standards for English language arts/literacy (ELA/L) or mathematics assessed at their grade level. They are **academically well prepared** to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L Grade 11: Students performing at this level **exceed academic expectations** for the knowledge, skills, and practices contained in the mathematics and ELA/L standards assessed at grade 11. They are very likely to engage successfully in entry-level, credit-bearing courses in mathematics and ELA/L, as well as technical courses requiring an equivalent command of the content area. Students performing at this level are exempt from having to take and pass placement tests in two- and four-year public institutions of higher education designed to determine whether they are academically prepared for such courses without need for remediation.

Level 4: Met expectations
Students performing at this level **meet academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3–10: Students performing at this level **meet academic expectations** for the knowledge, skills, and practices contained in the standards for ELA/L or mathematics assessed at their grade level. They are **academically prepared** to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L Grade 11: Students performing at this level **meet academic expectations** for the knowledge, skills, and practices contained in mathematics and ELA/L at grade 11. They are very likely to engage successfully in entry-level, credit-bearing courses in mathematics and ELA/L, as well as technical courses requiring an equivalent command of the content area. Students performing at this level are exempt from having to take and pass placement tests in two- and four-year public institutions of higher education designed to determine whether they are academically prepared for such courses without need for remediation.

Level 3: Approached expectations
Students performing at this level **approach academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3–10: Students performing at this level **approach academic expectations** for the knowledge, skills, and practices contained in the standards for ELA/L or mathematics assessed at their grade level. They are likely prepared to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L Grade 11: Students performing at this level **approach academic expectations** for the knowledge, skills, and practices contained in the ELA/L and mathematics standards assessed at grade 11. They are likely to engage successfully in entry-level, credit-bearing courses in mathematics and ELA/L, as well as technical courses requiring an equivalent command of the content area. **Students performing at Level 3 are strongly encouraged to continue to take challenging high school coursework in English and mathematics through graduation.** Postsecondary institutions are encouraged to use additional information about students performing at Level 3, such as course completion, course grades, and scores on other assessments to determine whether to place them directly into entry-level courses.

Level 2: Partially met expectations
Students performing at this level **partially meet academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3–10: Students performing at this level **partially meet academic expectations** for the knowledge, skills, and practices contained in the standards for ELA/L or mathematics assessed at their grade level. They will likely need academic support to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L Grade 11: Students performing at this level **partially meet academic** expectations for the knowledge, skills, and practices contained in the ELA/L and mathematics standards assessed at grade 11. They will likely need academic support to engage successfully in entry-level, credit-bearing courses, and technical courses requiring an equivalent command of the content area. Students performing at this level are not exempt from having to take and pass placement tests designed to determine whether they are academically prepared for such courses without the need for remediation in two- and four-year public institutions of higher education.

Level 1: Did not yet meet expectations
Students performing at this level **do not yet meet academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3–10: Students performing at this level **do not yet meet academic expectations** for the knowledge, skills, and practices contained in the standards for ELA/L or mathematics assessed at their grade level. They will need academic support to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L Grade 11: Students performing at this level **do not yet meet academic expectations** for the knowledge, skills, and practices contained in the ELA/L and mathematics standards assessed at grade 11. They will need academic support to engage successfully in entry-level, credit-bearing courses in college algebra, introductory college statistics, and technical courses requiring an equivalent level of mathematics. Students performing at this level are not exempt from having to take and pass placement tests in two- and four-year public institutions of higher education designed to determine whether they are academically prepared for such courses without need for remediation.

11.3 Performance Level Setting Process for the PARCC Assessment System

One of the main objectives of the PARCC assessment system is to provide information to students, parents, educators, and administrators as to whether students are on track in their learning for success after high school, defined as college- and career-readiness. To set performance levels associated with this objective, PARCC used the evidence-based standard setting (EBSS) method (Beimers, Way, McClarty, & Miles, 2012) for the PARCC PLS process. The EBSS method is a systematic method for combining various considerations into the process for setting performance levels, including policy considerations, content standards, educator judgment about what students should know and be able to demonstrate, and research to support PARCC's policy goals related to college- and career-readiness. A defined multistep process was used to allow a diverse set of stakeholders to consider the interaction of these elements in recommending performance level threshold scores for each PARCC assessment.

The seven steps of the EBSS process that were followed in order to establish performance standards for the PARCC assessments are:

Step 1: Define outcomes of interest and policy goals
Step 2: Develop research, data collection, and analysis plans
Step 3: Synthesize the research results
Step 4: Conduct pre-policy meeting
Step 5: Conduct performance level setting (PLS) meetings with panels
Step 6: Conduct reasonableness review with post-policy panel
Step 7: Continue to gather evidence in support of standards

A summary of key components within these steps is provided below. Additional detail about each step in the PARCC PLS process is provided in the *PARCC Performance Level Setting Technical Report*.

11.3.1 PARCC Research Studies

PARCC conducted two research studies in support of their policy goals—the benchmarking study and the postsecondary educators' judgment (PEJ) study. The benchmarking study included a review of the literature relative to college- and career-readiness as well as consideration of the percentage of students obtaining a level equivalent to college- and career-readiness on a set of external assessments (e.g., ACT, SAT, NAEP). The PEJ study involved a group of nearly 200 college faculty reviewing items on the Algebra II and ELA/L grade 11 PARCC assessments and making judgments about the level of performance needed on each item to be academically ready for an entry-level college-credit bearing course in mathematics or ELA/L. Additional detail[9] about the benchmarking study can be found in the *PARCC Performance Level Setting Technical Report* as well as in the *PARCC Benchmarking Study Report*. Additional detail about the PEJ study can be found in the *PARCC Performance Level Setting Technical Report* as well as in the *Postsecondary Educators' Judgment Study Final Report*.

11.3.2 PARCC Pre-Policy Meeting

Prior to the PLS meetings, a pre-policy meeting was convened to determine reasonable ranges that would be shown to panelists during the high school PLS meetings. Pre-policy meeting participants included representatives from both K–12 and higher education who served in roles such as commissioner/superintendent, deputy/assistant commissioner, state board member, director of assessment, director of academic affairs, senior policy associate, and so on. The reasonable ranges recommended by the pre-policy meeting defined the minimum and maximum percentage of students that would be expected to be classified as college- and career-ready. The pre-policy meeting participants reviewed the test purpose, how the performance standards will be used, and the results of the research studies to provide the recommendations for the reasonable ranges without viewing any student performance data.

11.3.3 Performance Level Setting Meetings

The task of the PLS committee was to recommend four threshold scores that would define the five performance levels for each PARCC assessment. PARCC solicited nominations from all states that had administered the PARCC assessments in 2014–2015 for panelists to serve on the PLS committees. Nominations were solicited both from state departments of public education (K–12) and higher education (primarily for participation on the high school panels). When selecting panelists, PARCC placed an emphasis on those educators who had content knowledge as well as experience with a variety of student groups and attempted to balance the panels in terms of state representation.

PARCC used an extended modified Angoff (Yes/No) method to collect educator judgments on the PARCC items. This method asked panelists to review each item on a reference form of the PARCC assessment and to make the following judgment:

[9] More information can be requested online from http://www.parcconline.org/assessments/test-design/research.

How many points would a borderline student at each performance level likely earn if they answered the question?

This extension to the Yes/No standard setting method (Plake, Ferdous, Impara, & Buckendahl, 2005) allowed for incorporation of the multipoint PARCC items by asking educators to evaluate (Yes or No) whether a borderline student would earn the maximum number of points on an item, a lesser number of points on an item, or no points on the item. In the case of a single point or multiple-choice item, this task simplifies to the standard Yes/No method.

After receiving training on the PLS procedure, panelists participated in three rounds of judgments for each assessment. Within each round, panelists were asked to consider the items in the test form, starting with the performance-based assessment (PBA) component and then the end-of-year (EOY) component. Each panelist made a judgment for the Level 2 performance level, followed by judgments for the Level 3 performance level, the Level 4 performance level, and the Level 5 performance level, in this order. The panelists entered their item judgments for each round by completing an online item judgment survey. Educator judgments were summed across items to create an estimated total score on the reference form for each performance level threshold. Feedback data relative to panelist agreement, student performance on the items, and student performance on the test as a whole were provided in between each of the three rounds of judgment. Panelists were shown the pre-policy reasonable ranges prior to making their Round 1 judgments and again as feedback data following each round of judgment.

A dry-run of the PARCC PLS meeting process was held for grade 11 ELA/L and Algebra II in order to evaluate the implementation of the PLS method with the innovative characteristics of the PARCC assessments. These content areas were selected because they combined all the various aspects of the PARCC assessments, including the various types of items, scoring rules, and performance level decisions. The dry-run PLS meetings provided the opportunity to implement and evaluate multiple aspects of the operational plan for the actual PLS meeting, including pre-work, meeting materials, data analysis and feedback, and staff and panelist functions. The results of the dry-run PLS meeting were used to implement improvements in the process for the operational PLS meetings. Additional information about the methods and results of the dry-run PLS meeting is available in the full report in the *PARCC Performance Level Setting Dry-Run Meeting Report*.

The PLS meetings for the PARCC assessments were conducted during three one-week sessions. The dates of the twelve PLS committee meetings that were conducted are shown in Table 11.1.

Additional information about the methods and results of the PLS meetings is available in the *Performance Level Setting Technical Report*.

Table 11.1 PARCC PLS Committee Meetings and Dates

Dates	Committees by Subjects and Grades
July 27–31, 2015	Algebra I/Integrated Mathematics I
	Geometry/Integrated Mathematics II
	Algebra II/Integrated Mathematics III
	Grade 9 English Language Arts/Literacy
	Grade 10 English Language Arts/Literacy
	Grade 11 English Language Arts/Literacy
August 17–21, 2015	Grades 7 & 8 Mathematics
	Grades 7 & 8 English Language Arts/Literacy
August 24–28, 2015	Grades 3 & 4 Mathematics
	Grades 5 & 6 Mathematics
	Grades 3 & 4 English Language Arts/Literacy
	Grades 5 & 6 English Language Arts/Literacy

11.3.4 PARCC Post-Policy Reasonableness Review

Performance standards for all PARCC assessments were recommended by PLS committees and reviewed by the PARCC Governing Board and (for the Algebra II, Integrated Mathematics III, and ELA/L grade 11 assessments) the Advisory Committee on College Readiness as part of a post-policy reasonableness review. This group reviewed both the median threshold score recommendations from each committee and the variability in the threshold scores as represented by the standard error of judgment (SEJ) of the committee. Adjustments to the median threshold scores that were within 2 SEJ were considered to be consistent with the PLS panels' recommendation.

In addition to voting to adopt the performance standards based on the committees' recommendations, this group also voted to conduct a shift in the PARCC performance levels to better meet the intended inferences about student performance. Holding the college- and career-ready (or on-track) expectations (i.e., the current level 4) constant, performance levels above this expectation were combined and performance levels below this expectation were expanded to create the final system of performance levels with three below and two above the college- and career-ready (or on-track) expectation. The shift in performance levels was accomplished using a scale anchoring process that involved two primary steps. In the first step, the top two performance levels, above college- and career-ready (or on-track), were combined into a single performance level and an additional performance level below college- and career-ready (or on-track) was created by empirically determining the midpoint between the existing two levels. In the second step, the performance level descriptors (PLDs) were updated using items that discriminated student performance well at this level to create a PLD aligned with the new empirically determined performance level. At this same time, PLDs for all performance levels were reviewed for consistency and continuity. Members of the original PLS committees were recruited to participate in this process. Additional information about this process can be found in the *Performance Level Setting Technical Report*.

Section 12: Scale Scores

PARCC assessments are designed to measure and report results in categories called master claims and subclaims. Master claims (or simply "claims") are at a higher level than subclaims with content representing multiple subclaims contributing to each claim outcome.

Four scale scores were reported for PARCC assessments.[10] A full summative scale score was reported for each mathematics assessment. A summative scale score and separate claim scores for Reading and Writing were reported for each English language arts/literacy (ELA/L) assessment. PARCC reports results according to five performance levels that delineate the knowledge, skills, and practices students are able to demonstrate:

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

Subclaim outcomes describe student performance for content-specific subsets of the item scores contributing to a particular claim. For example, Written Expression and Knowledge of Conventions subclaim outcomes are reported along with Writing claim scores. Subclaim outcomes are reported as *Below Expectations*, *Nearly Meets Expectations*, or *Meets or Exceeds Expectations*.

12.1 Operational Test Content (Claims and Subclaims)

A claim is a statement about student performance based on how students respond to test questions. PARCC tests are designed to elicit evidence from students that supports valid and reliable claims about the extent to which they are college and career ready or on track toward that goal and are making expected academic gains based on the Common Core State Standards (CCSS).

The number of items associated with each claim and subclaim outcome vary depending on test subject and grade. The item types vary in terms of the number of points associated with them, so that both the number of items and the number of points are important in evaluating the quality of a claim or subclaim score.

12.1.1 English Language Arts/Literacy

Table 12.1[11] includes the number of items and the number of points by subclaim and claim for ELA/L grade 3. Corresponding information is provided in Appendix 12.1 for all ELA/L grades.

[10] Addendum 12 presents a summary of results on scale scores for the fall 2017 administration.
[11] Table A.12.1 in Appendix 12.1 is identical to Table 12.1.

Table 12.1 Form Composition for ELA/L Grade 3

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	9 – 12	19 – 25
	Reading Informational Text	9 – 12	19 – 25
	Vocabulary	4 – 7	8 – 14
	Claim Total	22	46
Writing			
	Written Expression	2	27
	Knowledge of Conventions	1	9
	Claim Total	3	36
SUMMATIVE TOTAL		23	82

Note: Each prose constructed-response (PCR) trait is identified as a separate item in this table for the two writing subclaims and, in some cases, either the Reading Literary Text or the Reading Informational Text subclaim.

Each ELA/L form contains items of varying types. The prose constructed-response (PCR) traits contribute to different claims and the aggregate of the traits contributes to the summative scale score. The following details the number of possible points and the associated subclaims for the three PCR tasks:

- Literary Analysis Task
- Research Simulation Task
- Narrative Writing Task

The Literary Analysis Task and the Research Simulation Task are scored for two traits: Reading Comprehension and Written Expression, and Knowledge of Conventions. The Narrative Writing Task is scored for two traits: Written Expression and Knowledge of Conventions. All traits are initially scored as either 0–3 or 0–4; the Written Expression traits are multiplied by 3 (or weighted) to increase their contribution to the total score, making possible subclaim scores 0, 3, 6, and 9, or 0, 3, 6, 9, and 12. The maximum possible points for ELA/L PCR items are provided in Table 12.2.

Table 12.2 Contribution of Prose Constructed-Response Items to ELA/L

		Possible Points		
Grade	**Score**	**Literary Analysis Task**	**Research Simulation Task**	**Narrative Writing Task**
3	Reading	3	3	0
	Written Expression	9	9	9
	Knowledge of Conventions	3	3	3
	Total	15	15	12
4–5	Reading	4	4	0
	Written Expression	12	12	9
	Knowledge of Conventions	3	3	3
	Total	19	19	12
6–11	Reading	4	4	0
	Written Expression	12	12	12
	Knowledge of Conventions	3	3	3
	Total	19	19	15

12.1.2 Mathematics

Table 12.3[12] includes the numbers of items and points associated with subclaim scores for mathematics grade 3, as an example of the composition of the mathematics tests.

Table 12.3 Mathematics Form Composition for Grade 3

	Subclaims	**Number of Items**	**Number of Points**
Mathematics			
	Major Content	26	30
	Additional & Supporting Content	10	10
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		43	66

Because there is substantial variation in the composition of the tests, corresponding information is provided in the tables in Appendix 12.1 for all mathematics grades/courses.

12.2 Establishing the Reporting Scales

PARCC reporting scales designate student performance into one of five performance levels[13] with Level 1 indicating the lowest level of performance and Level 5 indicating the highest level of performance.

[12] Table A.12.10 in Appendix 12.1 is identical to Table 12.3.

[13] Section 11 provides an overview of the performance level setting process, and detailed information can be found in the *Performance Level Setting Technical Report*.

Threshold or cut scores associated with performance levels were initially expressed as raw scores on the performance level setting (PLS) forms approved by the PARCC Governing Board.

A scale score task force was assembled by PARCC, which made recommendations about how threshold levels would be represented on the reporting scale.

12.2.1 Full Summative Score Scale and Performance Levels

There are 201 defined full summative scale score points for both ELA/L and mathematics, ranging from 650 to 850. A scale score of 700 is associated with minimum Level 2 performance, and a scale score of 750 is associated with minimum Level 4 performance. The threshold for summative performance levels on the scale score metric recommended by the scale score task force is described in Table 12.4.

Table 12.4 Defined Summative Scale Scores

	Lowest Obtainable Scale Score (LOSS)	Level 2	Level 4	Highest Obtainable Scale Score (HOSS)
Full Summative	650	700	750	850

For spring 2015, scale scores were defined for each test as a linear transformation of the theta (θ_{2015}) scale. The theta values associated with the Level 2 and Level 4 performance levels were identified using the test characteristic curve associated with the performance level setting form. With Levels 2 and 4 scale scores fixed at 700 and 750, respectively, the relationship between theta (θ_{2015}) and scale scores ($Scale\ Score_{2015}$) was established as

$$Scale\ Score_{2015} = A_{2015} * \theta_{2015} + B_{2015} \quad (12\text{-}1)$$

where A_{2015} is the slope and B_{2015} is the intercept. The slope and intercept were established as

$$A_{2015} = \frac{750 - 700}{\theta_{2015\_Level\ 4} - \theta_{2015\_Level\ 2}} \quad (12\text{-}2)$$

and

$$B_{2015} = 750 - A_{2015} * \theta_{2015\_Level\ 4} \quad (12\text{-}3)$$

As indicated by these formulas, the slope and intercept for the summative scale scores were based on the theta scale, and by default the IRT parameter scale, established in 2015. Since the spring 2016 IRT parameter scale is the base scale for the IRT parameters, the scaling constants A_{2015} and B_{2015} were updated in order to continue reporting performance levels, summative scale scores, claim scores, and subclaim performance levels on the same scale as 2015. Maintaining the 2015 scale allows for prior year scores to be compared to current and future scores, and it maintains the performance levels cut scores.

New scaling constants for the summative scale score were needed for the linear transformation of the theta scale (θ_{2016}) to the 2015 reporting scale ($Scale\ Score_{2015}$):

$$Scale\ Score_{2015} = SA_{2016} * \theta_{2016} + SB_{2016} \quad (12\text{-}4)$$

The slope ($slope_{2015\_to\_2016}$) and intercept ($intercept_{2015\_to\_2016}$) generated during the year-to-year linking defined the linear relationship between the 2015 theta scale (θ_{2015}) and the 2016 theta scale (θ_{2016}). These values were included in the scale score formula, and the formulas were used to solve for the slope (SA_{2016}) and (SB_{2016}) intercept for 2016.

The slope (A_{2016}) was updated using the following formula:

$$SA_{2016} = A_{2015}/slope_{2015\_to\_2016} \quad (12\text{-}5)$$

where A_{2015} is the current scale score multiplicative constant, $slope_{2015\_to\_2016}$ is the multiplicative coefficient from the year-to-year linking, and SA_{2016} is the scale score slope constant for 2016 and beyond.

The intercept (B_{2016}) was updated using the following formula:

$$SB_{2016} = B_{2015} - A_{2016} * intercept_{2015\_to\_2016} \quad (12\text{-}6)$$

where B_{2015} is the current scale score additive constant, A_{2016} is the updated scale score slope, and (SB_{2016}) is the scale score intercept constant for 2016 and beyond.

In addition, new scaling constants for the reading and writing claim scales were needed. The same formulas were applied by replacing the slope (A_{2015}) and intercept (B_{2015}) with the reading claim slope and intercept and the writing claim slope and intercept.

A and *B* values resulting from these calculations as well as the theta values associated with the threshold performance levels are included in the 2015–2016 technical report. Also, the 2015–2016 technical report includes raw to scale score conversion tables for the performance level setting forms.

12.2.2 ELA/L Reading and Writing Claim Scale

There are 81 defined scale score points possible for Reading, ranging from 10 to 90. A scale score of 30 is associated with minimum Level 2 performance, and a scale score of 50 is associated with minimum Level 4 performance. There are 51 defined scale score points possible for Writing, ranging from 10 to 60. A scale score of 25 is associated with minimum Level 2 performance, and a scale score of 35 is associated with minimum Level 4 performance. Not all possible scale scores may be realized in a scoring table. The threshold Reading and Writing performance levels on the scale score metric recommended by the scale score task force are described in Table 12.5.

Table 12.5 Defined Scaled Scores for Reading and Writing Claim Scores

	Lowest Obtainable Scale Score	Level 2	Level 4	Highest Obtainable Scale Score
Reading	10	30	50	90
Writing	10	25	35	60

As with the full summative scores, scale scores for Reading and Writing were defined for each test as a linear transformation of the IRT theta (θ) scale. The same IRT theta scale was used for Reading and Writing as was used for the ELA/L full summative scores. The theta values associated with the Level 2 and Level 4 performance levels were identified using the test characteristic curve associated with the performance level setting form. As with the full summative scores, the relationship between theta and scale scores was established with Level 2 and Level 4 theta scores and the corresponding predefined scale scores. The formulas used for this are provided in Table 12.6.

Table 12.6 Calculating Scaling Constants for Reading and Writing Claim Scores

Reading	Writing
$Scale = A_R * \Theta + B_R$	$Scale = A_W * \Theta + B_W$
$A_R = \frac{50 - 30}{\Theta_{Level\ 4} - \Theta_{Level\ 2}}$	$A_W = \frac{35 - 25}{\Theta_{Level\ 4} - \Theta_{Level\ 2}}$
$B_R = 50 - A * \Theta_{Level\ 4}$	$B_W = 35 - A * \Theta_{Level\ 4}$

A and *B* values resulting from these calculations are included in Appendix 12.2.

12.2.3 Subclaims Scale

The Level 4 cut is defined as *Meets or Exceeds Expectations* because high school students at Level 4 or above are likely to have the skills and knowledge to meet the definition of career and college readiness. The Level 3 cut is defined as *Nearly Meets Expectations*. Subclaim outcomes center on the Level 3 and Level 4 performance levels and are reported at three levels:

- *Below Expectations*;
- *Nearly Meets Expectations*; or
- *Meets or Exceeds Expectations*.

The subclaim performance levels are designated through the IRT theta (θ) scale for the items associated with a particular subclaim. The theta values and corresponding raw scores associated with the Level 3 and Level 4 performance levels were identified using the test characteristic curve. Students earning a raw subclaim score equal to or greater than the Level 4 threshold were designated as *Meets or Exceeds Expectations*. Students not earning a raw subclaim score equal to or greater than the Level 3 threshold were designated as *Below Expectations*. Other students whose raw subclaim score fell between the Level 3 and 4 thresholds were designated as *Nearly Meets Expectations*.

12.3 Creating Conversion Tables

A PARCC conversion table relates the number of points earned by a student on the ELA/L full summative score, the mathematics full summative score, the Reading claim score, or the Writing claim score to the corresponding scale score for the test form administered to that student. An IRT inverse test characteristic curve (TCC) approach is used to develop the relationship between point scores and IRT ability estimates (θs). In carrying out the calculations, estimates of item parameters and thetas are substituted for parameters in the formulas in each step.

Step 1: Calculate the expected item score (i.e., estimated item true score) for every scale score in the selected range (determined by LOSS, HOSS, and scale score increment) based on the generalized partial credit model for both dichotomous and polytomous items:

$$s_i(\theta_j) = \sum_{m=0}^{M_i-1} m p_{im}(\theta_j), \tag{12-7}$$

$$p_{im}(\theta_j) = \frac{\exp\left[\sum_{k=0}^{m} Da_i(\theta_j - b_i + d_{ik})\right]}{\sum_{v=0}^{M_i-1} \exp\left[\sum_{k=0}^{v} Da_i(\theta_j - b_i + d_{iv})\right]}, \tag{12-8}$$

where $a_i(\theta_j - b_i + d_{i0}) \equiv 0$; $s_i(\theta_j)$ is the expected item score for item i on a scale score, θ_j; $p_{im}(\theta_j)$ is the probability of a test taker with θ_j getting score m on item i; m_i is the number of score categories of item i with possible item scores as consecutive integers from 0 to m_i -1; D is the IRT scale constant (1.7); a_i is a slope parameter; $b_i$is a location parameter reflecting overall item difficulty; d_{ik} is a location parameter incrementing the overall item difficulty to reflect the difficulty of earning score category k; v is the number of score categories.

Step 2: Calculate the expected (weighted) test score for every scale score in the selected range:

$$T_j = \sum_{i=1}^{I} w_i s_i(\theta_j), \tag{12-9}$$

where T_j is the expected (weighted) test score on a scale score, θ_j; w_i is the item weight for item i (e.g., with w_i = 2, a dichotomous item is scored as 0 or 2, and a three-category item is scored as 0, 2, or 4); I is the total number of items in a test form.

Step 3: Calculate the estimated conditional standard error of measurement (CSEM) for each scale score in the selected range:

$$\text{CSEM}_j = \frac{1}{\sqrt{\sum_{i=1}^{I} L_i(\theta_j)}}, \tag{12-10}$$

$$L_i(\theta_j) = (Da_i)^2[s_{i2}(\theta_j) - s_i^2(\theta_j)], \tag{12-11}$$

$$s_{i2}(\theta_j) = \sum_{m=0}^{M_i-1} m^2 p_{im}(\theta_j), \tag{12-12}$$

where $L_i(\theta_j)$ is the estimated item information function for item i on scale score θ_j.

Step 4: Match every raw score with a scale score. θ_j is the scale score for a raw score r_h, if T_j - r_h is minimum across all T_j.

Figure 12.1 contains TCCs, estimated CSEM curves, and estimated information (INF) curves for ELA/L grade 3.[14] The curves in each figure are for the two core online forms (O1 and O2), one core paper form (P1), and the accommodated forms A(O) or A(P), which may be either online (O) or paper (P). The average difficulty of each form is reported and the curves are reported on the theta scale. Vertical dotted lines indicate the performance level cuts on the theta scale. For ELA/L grade 3, all forms had very similar TCCs. CSEM and information curves were also similar. Appendix 12.3 contains TCC, CSEM, and INF curves for all ELA/L grades and all mathematics grades/courses.

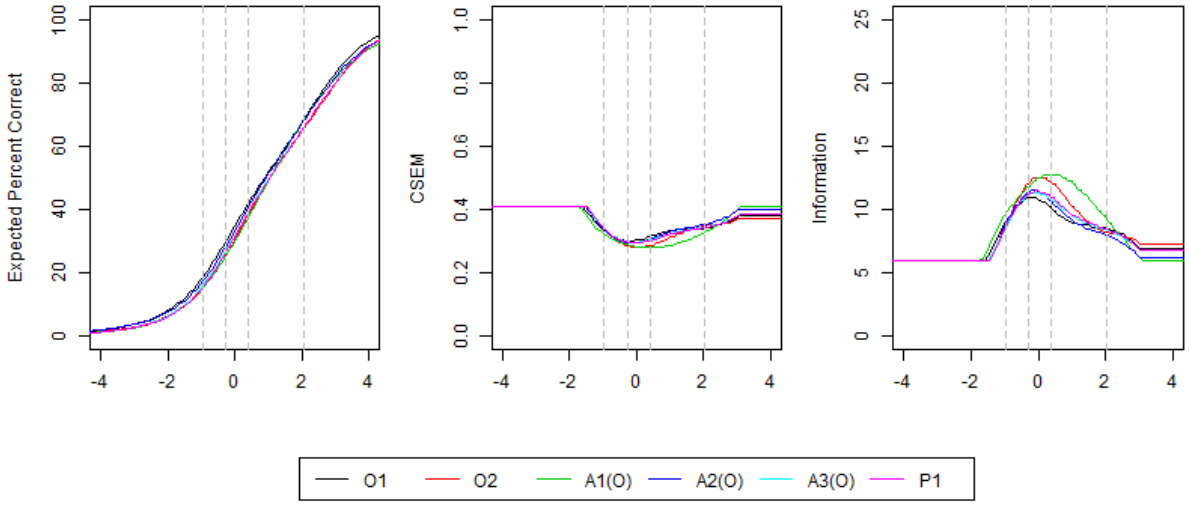

Figure 12.1 Test Characteristic Curves, Conditional Standard Error of Measurement Curves, and Information Curves for ELA/L Grade 3

[14] Grade 3 TCC, CSEM, and INF curves are also included in Appendix Figure A.12.1.

12.4 Score Distributions

12.4.1 Score Distributions for ELA/L

All Students

Figures 12.2 through 12.4 graphically represent the distributions of scale scores for grades 3 through 11 ELA/L full summative, Reading, and Writing, respectively. The vertical axis of each graph, labeled "Density," represents the proportion of students earning the scale score point indicated along the horizontal axis. For the summative distributions, the y-axis ranges from 0 to 0.015 and the x-axis from 650 to 850. For the Reading distributions, the y-axis ranges from 0 to 0.05 and the x-axis from 10 to 90. For the Writing distributions, the y-axis ranges from 0 to 0.10 and the x-axis from 10 to 60.

The distributions of the ELA/L summative scale scores were fairly symmetrical and centered around the Level 4 cut score (i.e., 750), with the exception of grades 10 and 11, whose distributions were centered around the same location but whose shapes were slightly more irregular.

Reading scale scores tended to be centered around or slightly below the Level 4 cut score of 50 and were slightly more irregular than the summative scale scores. Distributions tended to be fairly symmetric.

Writing scale score distributions were noticeably less smooth than Reading or ELA/L full summative distributions due to peaks related to the weighting of the Written Expression portion of the PCR tasks and a noticeable proportion of students at the lowest obtainable scale score (LOSS). Due to the weighting of the Written Expression trait, multiple Writing scale score values are not likely to be obtained resulting in multiple peaks across the range of the Writing scale score. A noticeable proportion of students earned the LOSS of ten in Writing across all ELA/L grades. Students with zero raw score points on the written portion of the assessment are automatically assigned the LOSS value of a scale. Writing items are embedded exclusively in PCR tasks, which tended to be difficult. The Written Expression trait also tended to be the most difficult of the PCR traits.

Across the ELA/L grades, the graphs indicate zero students obtaining scale scores in the range of eleven to seventeen. As noted in Section 12.2.2, the scale score task force selected ten as the LOSS. This value was selected to be consistent with the Reading LOSS and reduce truncation at the lower ends of the scale. However, the scale is defined by the theta values associated with the Level 2 and Level 4 performance levels. All other scale score values are identified through a theta-to-scale score linear transformation applying the scaling constants (Table 12.6). For Writing, the lowest theta estimate associated with raw scores ranging from one to two are linearly transformed to scale score values in the range of seventeen to nineteen. Whereas, the Reading lowest theta estimates associated with raw scores ranging from one to two are linearly transformed to scale score values in the range of ten to eleven. The gap in the proportion of students at the scale scores between the LOSS value of ten and the scale score values around seventeen to nineteen is an artifact of scale score task force selecting the LOSS value of ten.

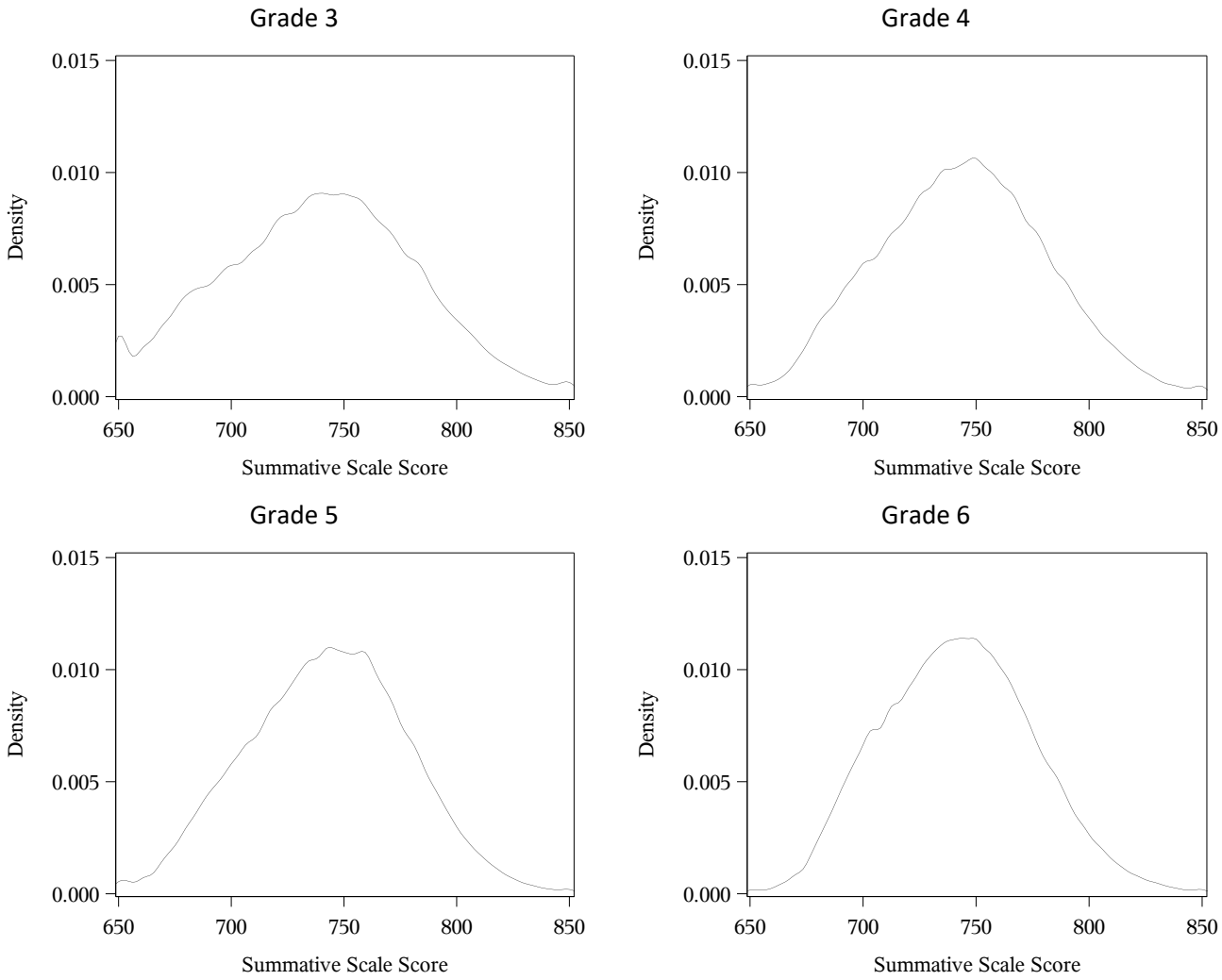

Figure 12.2 Distributions of ELA/L Scale Scores: Grades 3–11

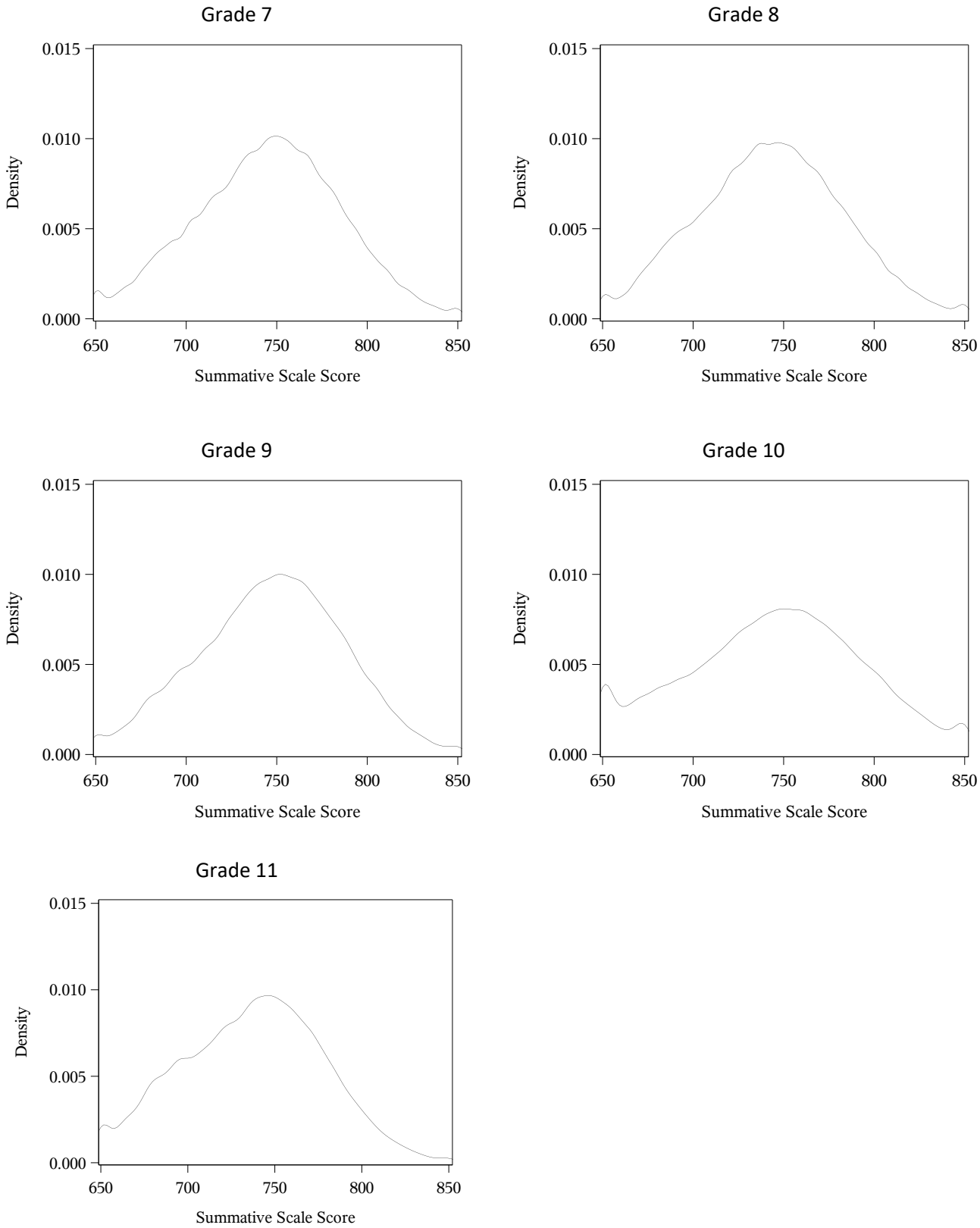

Figure 12.2 (continued) Distributions of ELA/L Scale Scores: Grades 3–11

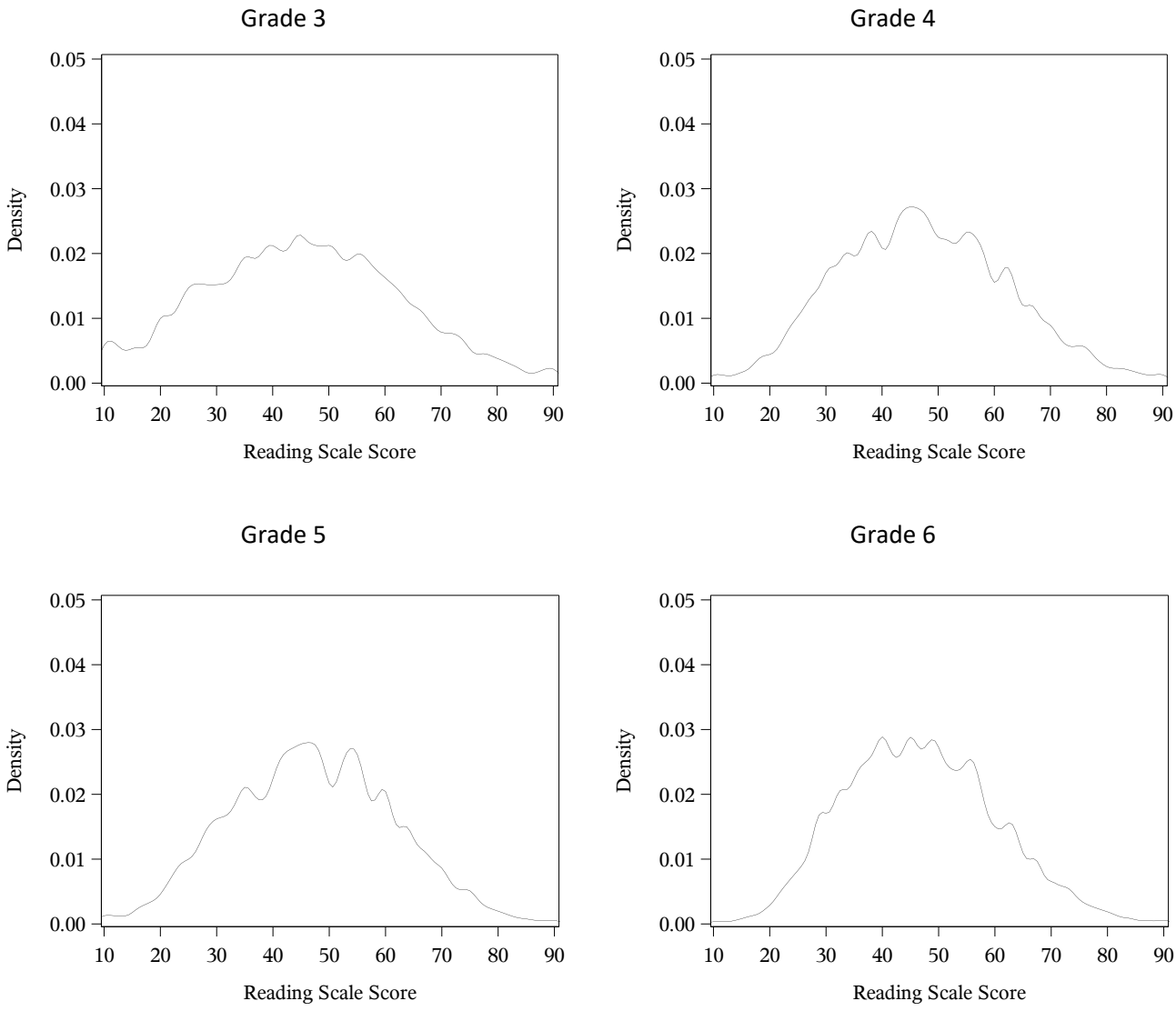

Figure 12.3 Distributions of Reading Scale Scores: Grades 3–11

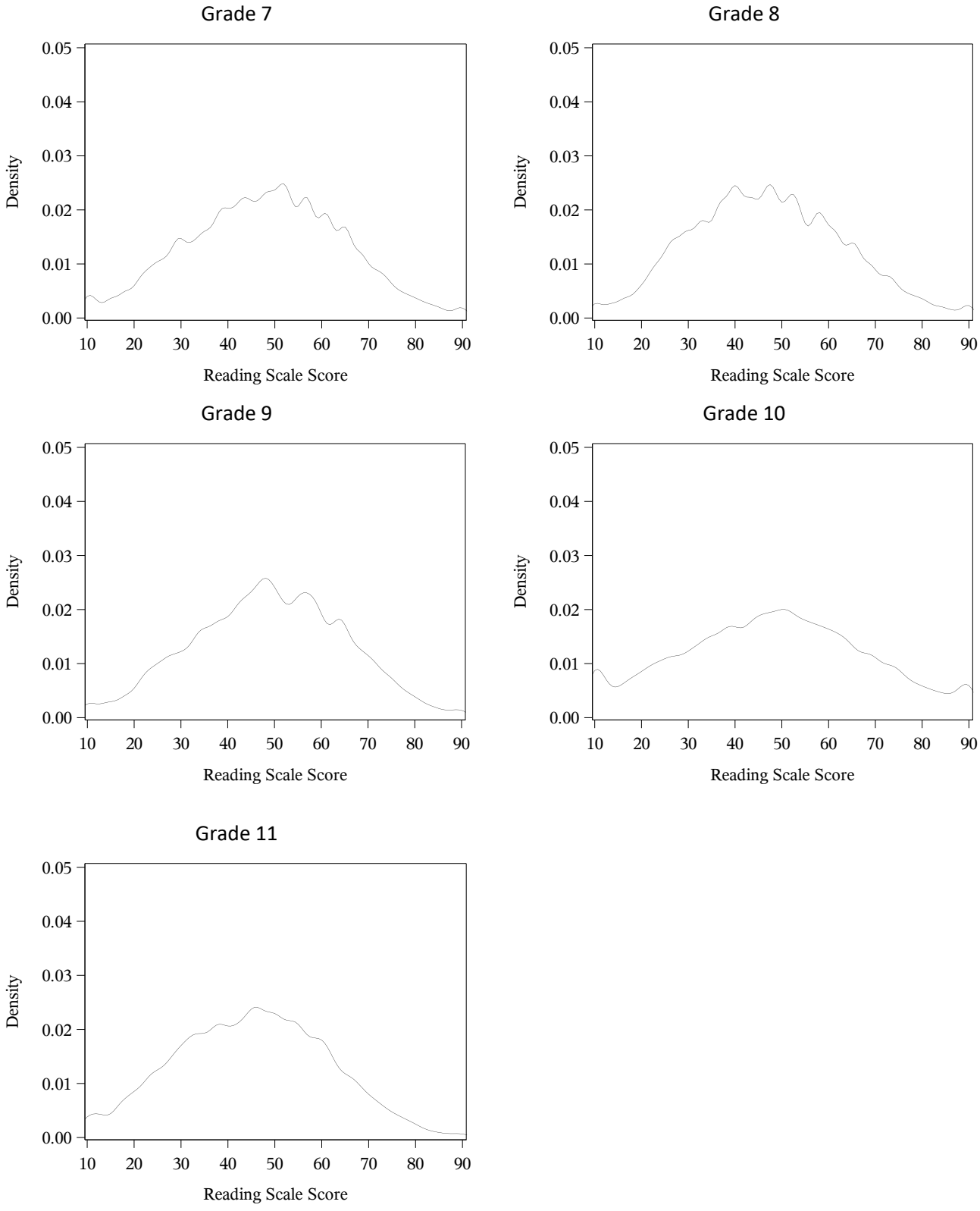

Figure 12.3 (continued) Distributions of Reading Scale Scores: Grades 3–11

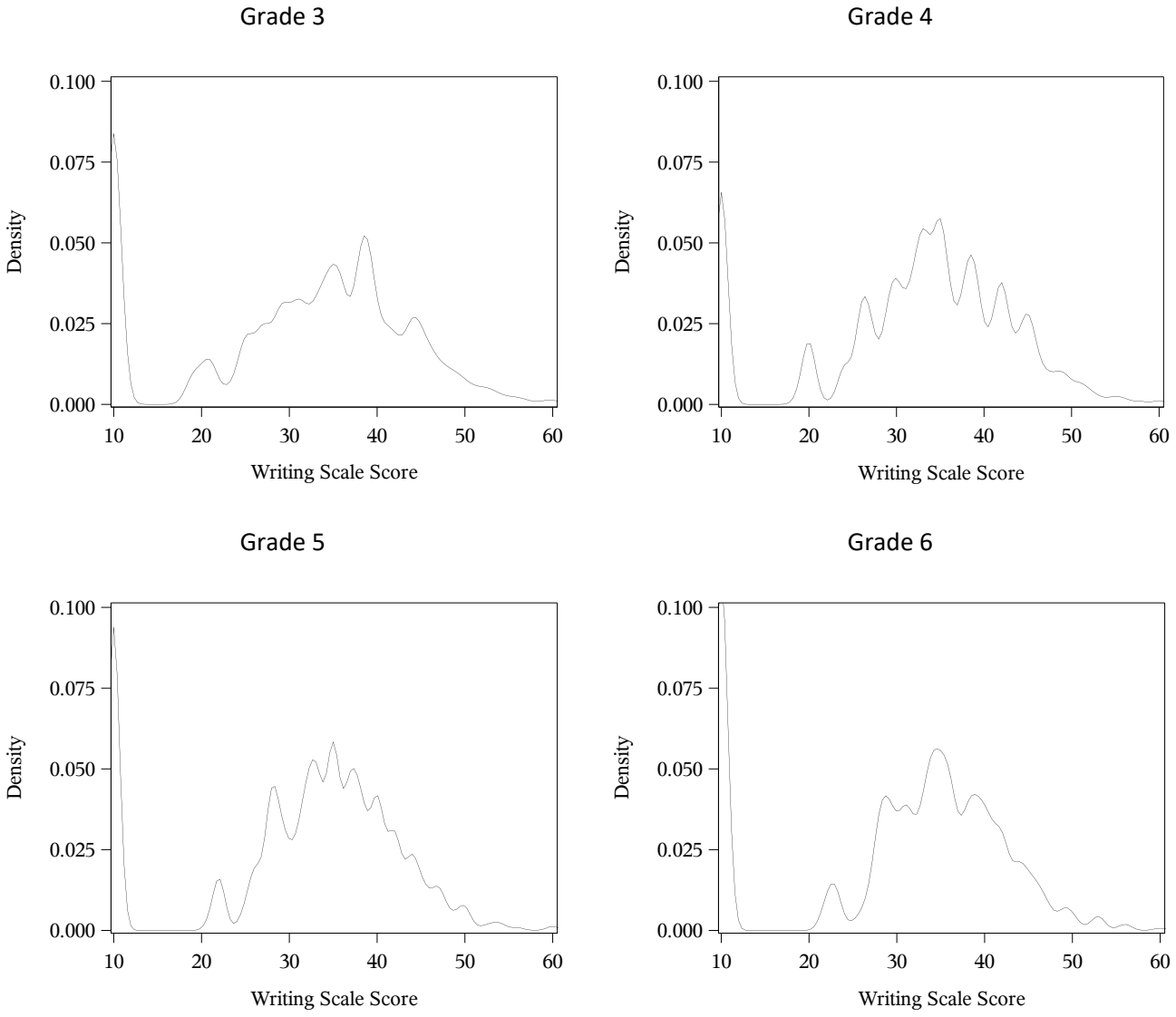

Figure 12.4 Distributions of Writing Scale Scores: Grades 3–11

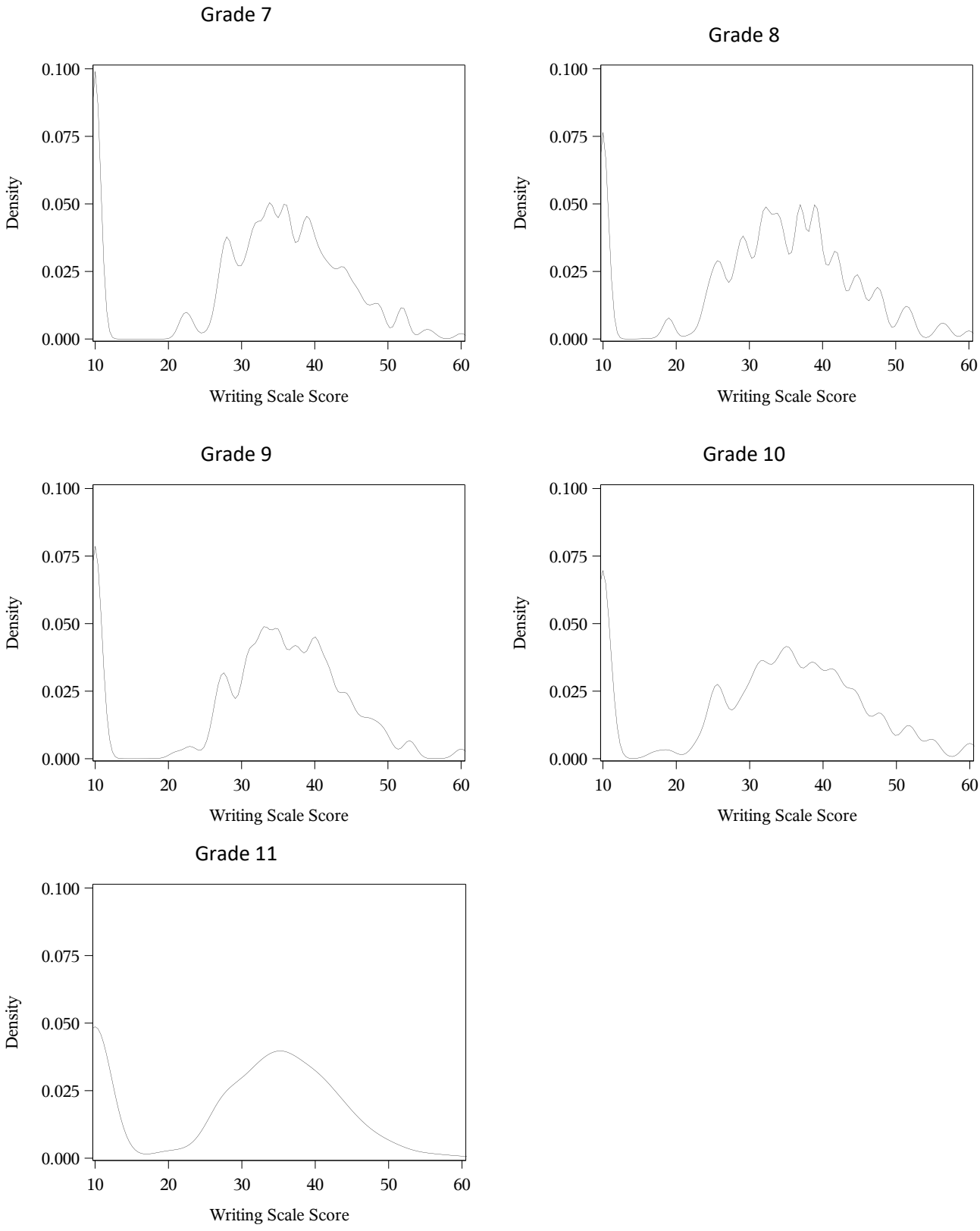

Figure 12.4 (continued) Distributions of Writing Scale Scores: Grades 3–11

Groups

Grade 3 group statistics for ELA/L full summative, Reading, and Writing scale scores are presented in Table 12.7.[15] Mean scores were higher for female students relative to male students. Mean scores were highest for Asian students and were lowest for American Indian/Alaska native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than non-EL students. Students with disabilities (SWD) performed less well than students without disabilities. Patterns of mean scale scores were similar in grades 4 through 8, although the ordering of ethnic groups varied slightly; corresponding tables for all grades are presented in Appendix 12.4.

Table 12.7 Subgroup Performance for ELA/L: Grade 3

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		338,927	738.80	42.68	650	850
Gender	Female	166,189	744.26	43.00	650	850
	Male	172,738	733.55	41.70	650	850
Ethnicity	American Indian/Alaska Native	4,941	718.39	35.86	650	850
	Asian	22,383	769.37	40.35	650	850
	Black or African American	67,641	723.07	40.17	650	850
	Hispanic/Latino	94,348	726.78	40.31	650	850
	Native Hawaiian or Pacific Islander	578	748.27	45.12	650	850
	Multiple Race Selected	11,490	744.74	43.10	650	850
	White	137,498	750.01	39.86	650	850
Economic Status*	Economically Disadvantaged	168,229	723.74	39.45	650	850
	Not Economically Disadvantaged	170,629	753.66	40.47	650	850
English Learner Status	English Learner (EL)	50,459	712.86	35.69	650	850
	Non-English Learner	288,272	743.36	42.17	650	850
Disabilities	Students with Disabilities (SWD)	53,040	709.57	40.25	650	850
	Students without Disabilities	284,826	744.31	40.87	650	850

[15] Table A.12.27 in Appendix 12.4 is identical to Table 12.7.

Group Type	Group	*N*	Mean	*SD*	Min	Max
Reading Score		338,927	45.45	17.31	10	90
Gender	Female	166,189	46.95	17.42	10	90
	Male	172,738	44.00	17.08	10	90
Ethnicity	American Indian/Alaska Native	4,941	36.51	14.37	10	90
	Asian	22,383	56.75	16.54	10	90
	Black or African American	67,641	39.04	15.79	10	90
	Hispanic/Latino	94,348	40.01	15.96	10	90
	Native Hawaiian or Pacific Islander	578	48.48	17.43	10	90
	Multiple Race Selected	11,490	48.49	17.57	10	90
	White	137,498	50.55	16.47	10	90
Economic Status*	Economically Disadvantaged	168,229	39.08	15.59	10	90
	Not Economically Disadvantaged	170,629	51.73	16.61	10	90
English Learner Status	English Learner (EL)	50,459	34.16	13.68	10	90
	Non-English Learner	288,272	47.43	17.12	10	90
Disabilities	Students with Disabilities (SWD)	53,040	34.62	16.38	10	90
	Students without Disabilities	284,826	47.49	16.72	10	90
Writing Score		338,927	31.03	12.28	10	60
Gender	Female	166,189	33.02	11.92	10	60
	Male	172,738	29.12	12.32	10	60
Ethnicity	American Indian/Alaska Native	4,941	27.15	11.41	10	60
	Asian	22,383	38.92	10.39	10	60
	Black or African American	67,641	27.23	12.36	10	60
	Hispanic/Latino	94,348	28.74	12.21	10	60
	Native Hawaiian or Pacific Islander	578	33.48	12.94	10	60
	Multiple Race Selected	11,490	31.82	12.30	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
	White	137,498	33.26	11.49	10	60
Economic Status*	Economically Disadvantaged	168,229	27.68	12.19	10	60
	Not Economically Disadvantaged	170,629	34.34	11.44	10	60
English Learner Status	English Learner (EL)	50,459	25.74	11.81	10	60
	Non-English Learner	288,272	31.97	12.12	10	60
Disabilities	Students with Disabilities (SWD)	53,040	22.62	12.31	10	60
	Students without Disabilities	284,826	32.62	11.61	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP)—receipt of free or reduced-price lunch (FRL).

Grade 9 group statistics for ELA/L, Reading, and Writing scale scores are presented in Table 12.8. Mean scores were very similar to what was observed for grades 3 through 8. Mean scores were higher for female students than for male students. Mean scores were highest for Asian students and were lowest for American Indian/Alaska native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than non-EL students. Students with disabilities (SWD) performed less well than students without disabilities. Similar patterns are observed in other grades, with some small variations in the ordering of the ethnic groups. Corresponding tables for grades 10 and 11 are presented in Appendix 12.4.

Table 12.8 Subgroup Performance for ELA/L: Grade 9

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		128,229	746.64	39.64	650	850
Gender	Female	63,018	754.49	38.30	650	850
	Male	65,211	739.06	39.44	650	850
Ethnicity	American Indian/Alaska Native	3,151	725.14	30.21	650	833
	Asian	10,602	780.68	36.46	650	850
	Black or African American	17,883	731.61	36.13	650	850
	Hispanic/Latino	41,613	731.21	37.00	650	850
	Native Hawaiian or Pacific Islander	309	753.49	37.70	659	850
	Multiple Race Selected	1,753	757.91	39.13	650	850
	White	52,907	757.92	35.44	650	850

Group Type	Group	*N*	Mean	*SD*	Min	Max
Economic Status*	Economically Disadvantaged	53,989	729.41	35.70	650	850
	Not Economically Disadvantaged	74,237	759.18	37.61	650	850
English Learner Status	English Learner (EL)	8,463	695.59	28.04	650	811
	Non-English Learner	119,469	750.30	37.79	650	850
Disabilities	Students with Disabilities (SWD)	22,846	718.94	35.60	650	850
	Students without Disabilities	105,209	752.70	37.86	650	850
Reading Score		128,229	48.87	16.08	10	90
Gender	Female	63,018	50.99	15.69	10	90
	Male	65,211	46.82	16.19	10	90
Ethnicity	American Indian/Alaska Native	3,151	39.63	12.52	10	88
	Asian	10,602	61.52	15.10	10	90
	Black or African American	17,883	42.97	14.72	10	90
	Hispanic/Latino	41,613	42.69	14.99	10	90
	Native Hawaiian or Pacific Islander	309	50.78	15.25	10	90
	Multiple Race Selected	1,753	53.59	15.88	10	90
	White	52,907	53.57	14.50	10	90
Economic Status*	Economically Disadvantaged	53,989	41.96	14.50	10	90
	Not Economically Disadvantaged	74,237	53.89	15.29	10	90
English Learner Status	English Learner (EL)	8,463	28.30	11.01	10	79
	Non-English Learner	119,469	50.34	15.36	10	90
Disabilities	Students with Disabilities (SWD)	22,846	38.36	14.66	10	90
	Students without Disabilities	105,209	51.17	15.45	10	90
Writing Score		128,229	32.19	12.07	10	60
Gender	Female	63,018	35.23	10.83	10	60
	Male	65,211	29.26	12.47	10	60
Ethnicity	American Indian/Alaska Native	3,151	27.59	10.58	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
	Asian	10,602	41.39	9.53	10	60
	Black or African American	17,883	28.06	11.96	10	60
	Hispanic/Latino	41,613	28.15	12.03	10	60
	Native Hawaiian or Pacific Islander	309	34.67	11.10	10	60
	Multiple Race Selected	1,753	34.86	11.73	10	60
	White	52,907	35.09	10.73	10	60
Economic Status*	Economically Disadvantaged	53,989	27.64	11.89	10	60
	Not Economically Disadvantaged	74,237	35.50	11.07	10	60
English Learner Status	English Learner (EL)	8,463	18.34	10.37	10	52
	Non-English Learner	119,469	33.19	11.56	10	60
Disabilities	Students with Disabilities (SWD)	22,846	23.71	12.21	10	60
	Students without Disabilities	105,209	34.05	11.21	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP)—receipt of free or reduced-price lunch (FRL).

12.4.2 Score Distributions for Mathematics

All Students

Figure 12.5 graphically represents the distributions of scale scores for grades 3 through 8 mathematics. For the summative distributions, the y-axis ranges from 0 to 0.02 and the x-axis from 650 to 850. Scale score distributions generally peaked between approximately 700 and the Level 4 performance level cut of 750. Figure 12.6 graphically represents the distributions of scale scores for Algebra I, Geometry, Algebra II, and Integrated Mathematics I, II, and III. Scale score distributions generally peaked between approximately 700 and the 750 Level 4 performance level cut score for Algebra I, Geometry, and Algebra II. Integrated Mathematics I, II, and III distributions are positively skewed and peaked slightly around or below 700.

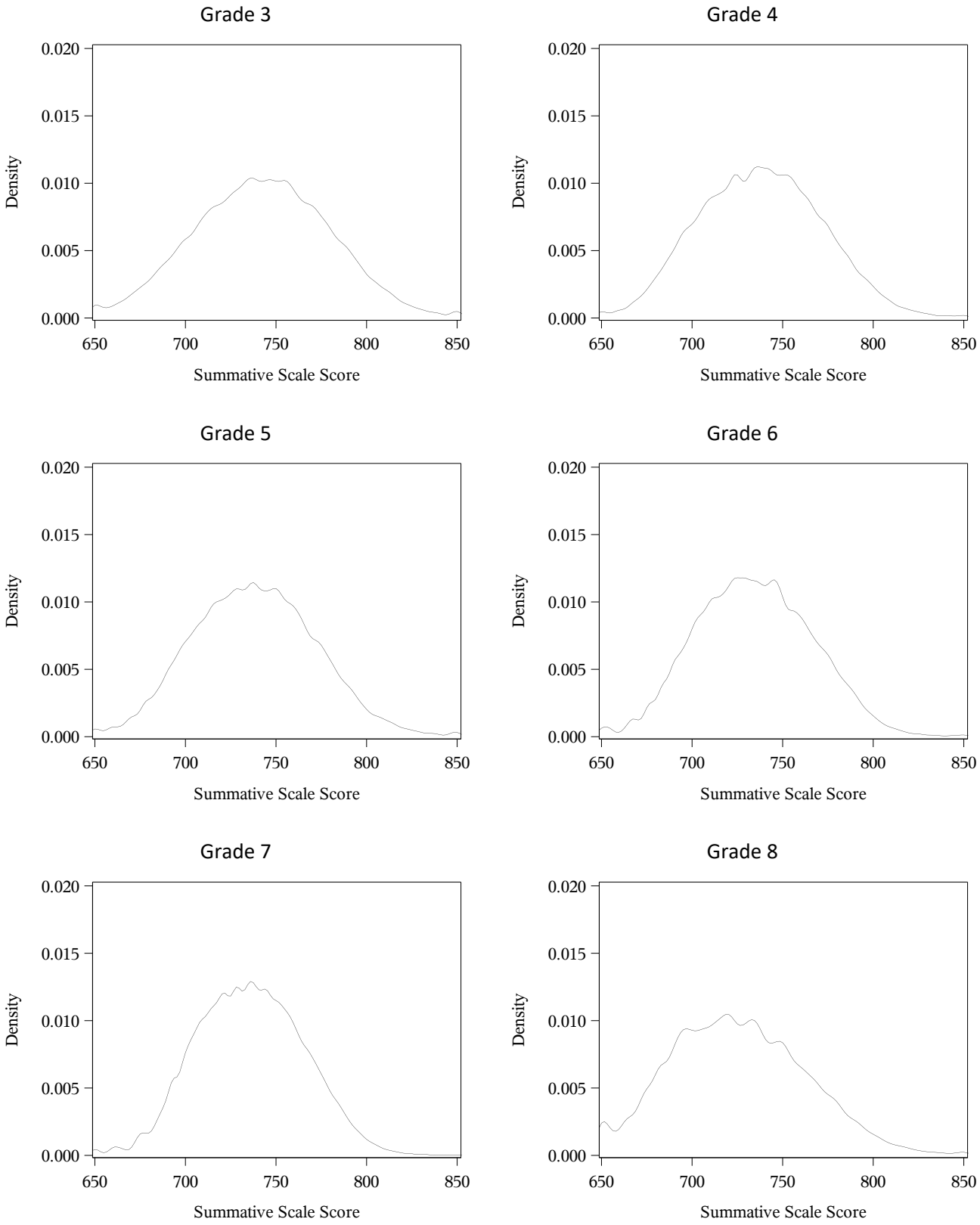

Figure 12.5 Distributions of Mathematics Scale Scores: Grades 3–8

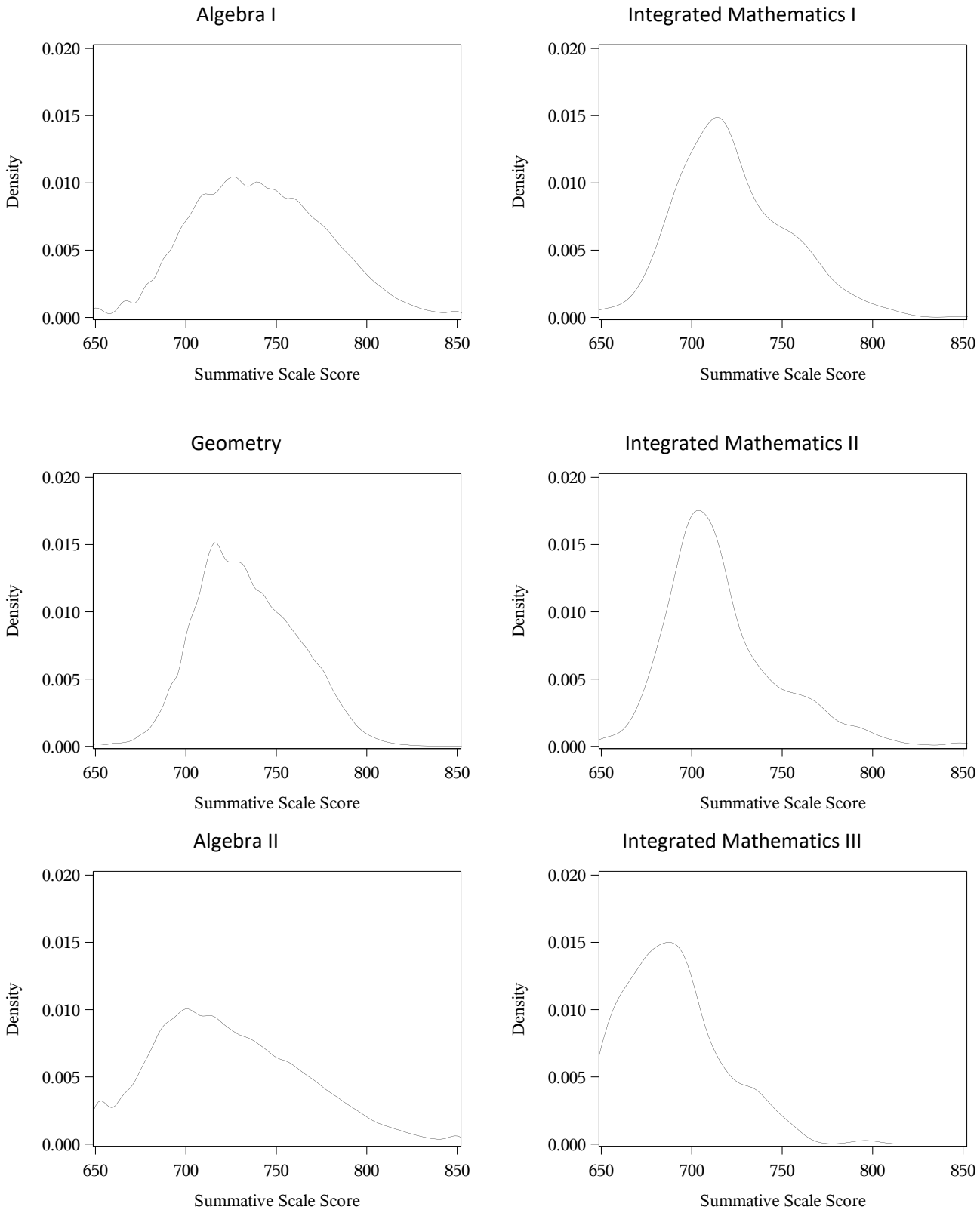

Figure 12.6 Distributions of Mathematics Scale Scores: High School

Groups

Grade 3 group statistics for mathematics scale scores are presented in Table 12.9.[16] Mean scores were slightly higher for female students relative to male students. Mean scores were highest for Asian students and were lowest for American Indian/Alaska native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than non-EL students. Students with disabilities (SWD) performed less well than students without disabilities. Students using the Spanish Language form tended to have lower mean scores. Generally similar patterns were observed in other grades, with some slight variations in the orderings of the ethnic groups. Corresponding tables for all grades are presented in Appendix 12.4.

Table 12.9 Subgroup Performance for Mathematics Scale Scores: Grade 3

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		348,117	742.56	37.24	650	850
Gender	Female	170,632	742.81	36.27	650	850
	Male	177,485	742.31	38.15	650	850
Ethnicity	American Indian/Alaska Native	4,971	724.10	30.77	650	850
	Asian	23,146	774.35	35.70	650	850
	Black or African American	68,498	725.50	34.68	650	850
	Hispanic/Latino	97,908	732.12	33.34	650	850
	Native Hawaiian or Pacific Islander	688	748.82	36.49	650	850
	Multiple Race Selected	12,281	747.03	38.10	650	850
	White	140,467	753.14	34.47	650	850
Economic Status*	Economically Disadvantaged	170,494	728.39	33.53	650	850
	Not Economically Disadvantaged	172,005	756.35	35.63	650	850
English Learner Status	English Learner (EL)	54,563	724.20	31.41	650	850
	Non-English Learner	293,337	745.98	37.25	650	850
Disabilities	Students with Disabilities (SWD)	54,044	720.44	37.06	650	850
	Students without Disabilities	292,963	746.69	35.81	650	850
Form Language	Spanish	4,906	715.52	30.48	650	832

Note: *Economic status was based on participation in National School Lunch Program (NSLP)—receipt of free or reduced-price lunch (FRL).

[16] Table A.12.36 in Appendix 12.5 is identical to Table 12.9.

Algebra I scale score statistics are presented in Table 12.10. Mean scores were higher for female students relative to male students. Mean scores were highest for Asian students and were lowest for American Indian/Alaska native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than non-EL students. Students with disabilities (SWD) performed less well than students without disabilities. Students using the Spanish Language form tended to have lower mean scores. Similar patterns were observed in the other high school tests with some of the previously mentioned exceptions in the ordering of the ethnicities applying to these tests as well. Corresponding tables are presented in Appendix 12.4.

Table 12.10 Subgroup Performance for Mathematics Scale Scores: Algebra I

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		221,242	741.16	36.64	650	850
Gender	Female	107,530	742.58	35.06	650	850
	Male	113,712	739.82	38.02	650	850
Ethnicity	American Indian/Alaska Native	3,677	719.20	26.53	650	827
	Asian	17,417	775.45	37.13	650	850
	Black or African American	46,624	723.61	30.07	650	850
	Hispanic/Latino	61,703	727.11	31.35	650	850
	Native Hawaiian or Pacific Islander	507	747.04	36.34	650	850
	Multiple Race Selected	5,406	749.11	35.36	650	850
	White	85,812	754.25	33.49	650	850
Economic Status*	Economically Disadvantaged	86,722	725.73	30.96	650	850
	Not Economically Disadvantaged	130,987	751.35	36.72	650	850
English Learner Status	English Learner (EL)	17,105	708.36	26.13	650	850
	Non-English Learner	203,950	743.94	36.05	650	850
Disabilities	Students with Disabilities (SWD)	37,607	718.31	32.60	650	850
	Students without Disabilities	183,295	745.86	35.65	650	850
Form Language	Spanish	3,579	704.51	21.04	650	802

Note: *Economic status was based on participation in National School Lunch Program (NSLP)—receipt of free or reduced-price lunch (FRL).

Integrated Mathematics I scale score statistics are presented in Table 12.11. Mean scores were higher for female students relative to male students. Mean scores were highest for White students and were lowest for Black or African American students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than non-EL students. Sample sizes for Integrated Mathematics I subgroups tended to be small, and a number of categories did not have sufficient sample sizes for reporting purposes in this table. Somewhat similar patterns were observed in Integrated Mathematics II and Integrated Mathematics III, but sample sizes for some subgroups are very small, and caution should be used in interpretations. Tables for these tests can be found in Appendix 12.4.

Table 12.11 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics I

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		750	722.43	30.43	650	850
Gender	Female	338	723.38	29.09	650	812
	Male	412	721.65	31.50	650	850
Ethnicity	American Indian/Alaska Native	40	725.25	30.50	677	812
	Asian	n/r	n/r	n/r	n/r	n/r
	Black or African American	35	708.14	25.84	650	766
	Hispanic/Latino	479	718.14	27.02	650	806
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	n/r	n/r	n/r	n/r	n/r
	White	186	735.39	35.21	674	850
Economic Status*	Economically Disadvantaged	505	717.88	28.24	650	814
	Not Economically Disadvantaged	245	731.81	32.62	650	850
English Learner Status	English Learner (EL)	110	709.45	20.69	662	758
	Non-English Learner	640	724.66	31.28	650	850
Disabilities	Students with Disabilities (SWD)	160	709.23	30.00	650	850
	Students without Disabilities	590	726.01	29.57	650	818
Form Language	Spanish	n/r	n/r	n/r	n/r	n/r

Note: *Economic status was based on participation in National School Lunch Program (NSLP)—receipt of free or reduced-price lunch (FRL).

12.5 Interpreting Claim Scores and Subclaim Scores

12.5.1 Interpreting Claim Scores

PARCC ELA/L assessments provide separate claim scale scores for both Reading and Writing. The claim scale scores and the summative scale score are on different scales; therefore, the sum of the scale scores for each claim will not equal the summative scale score. PARCC Reading scale scores range from 10 to 90 and PARCC Writing scale scores range from 10 to 60.

The claim scores can be interpreted by comparing a student's claim scale score to the average performance for the school, district, and state. The PARCC Individual Student Report (ISR) provides the student scale score results and the average scale score results for the school, district, and state.

12.5.2 Interpreting Subclaim Scores

Within each reporting category are specific skill sets (subclaims) students demonstrate on the PARCC assessments. Subclaim categories are not reported using scale scores or performance levels. Subclaim performance for PARCC assessments is reported using graphical representations that indicate how the student performed relative to the Level 3 and Level 4 performance levels for the content area.

Subclaim indicators represent how well students performed in a subclaim category relative to Level 3 and Level 4 thresholds for the items associated with the subclaim category. To determine a student's subclaim performance, the Level 3 and Level 4 thresholds corresponding to the IRT based performance for the items for a given subclaim determined the reference points for *Approached Expectations* and *Did Not Yet Meet Expectations* or *Partially Met Expectations*, respectively.

Student performance for each subclaim is marked with a subclaim performance indicator.

- An 'up' arrow for the specified subclaim for *Meets or Exceeds Expectations* indicates that the student's performance for the subclaim was equal to or better than the threshold for Level 4 (i.e., students whose summative scale score was 750).
- A 'bidirectional' arrow for the specified subclaim for *Nearly Meets Expectations* indicates that the student's performance was below the Level 4 threshold (i.e., summative scale score was 750) but better than or equal to the Level 3 threshold (i.e., summative scale score was 725).
- A 'down' arrow for the specified subclaim for *Below Expectations* indicates that the student's performance for the subclaim was below the Level 3 threshold (i.e., summative scale score was 725).

Section 13: Student Growth Measures

Student growth percentiles (SGPs) are normative measures of annual progress. Normative measures are useful in answering questions like "How does my academic progress compare with the academic progress of my peers?" In contrast to criterion-referenced measures of growth, which describe academic growth toward a particular goal, norm-referenced measures of growth describe students' growth relative to that of students who performed similarly in the past (Betebenner, 2009).

SGPs measure individual student progress by tracking student scores from one year to the next. SGPs compare a student's performance to that of his or her academic peers both within the state and across the consortium. Academic peers are defined as students in the norm group who took the same PARCC assessment as the student in prior years and achieved a similar score.

The SGP describes a student's location in the distribution of current test scores for all students who performed similarly in the past. SGPs indicate the percentage of academic peers above whom the student scored. With a range of 1 to 99, higher numbers represent higher growth and lower numbers represent lower growth. For example, a SGP of 60 on grade 7 ELA/L means that the student scored better than 60 percent of the students in the state or consortium who took grade 7 ELA/L in spring 2018 *and* who had achieved a similar score as this student on the grade 6 ELA/L assessment in spring 2017 and the grade 5 ELA/L assessment in spring 2016.[17] A SGP of 50 represents typical (median) student growth for the state or consortium. Because students are only compared with other students who performed similarly in the past, all students, regardless of starting point, can demonstrate high or low growth.

The 2017–2018 academic year is the fourth year of PARCC test administration. Students in states that participated in the consortium in spring 2016 and spring 2017 generally received SGPs based on two prior scores. Students in states that participated in the consortium in spring 2017 received SGPs based on one prior score. Students who do not have a previous PARCC test score, which include any new testers and all grade 3 students, do not receive an SGP.

13.1 Norm Groups

The norm groups consisted of students with the same prior scores based on grade or content area progressions (academic peers). SGPs were based on up to two years of prior test scores from spring 2016 and spring 2017 PARCC administrations. States administering traditional mathematics PARCC assessments in fall 2016 or fall 2017 may also have SGPs based on these prior scores. Tables 13.1–13.8 list the grade or content area progressions required for SGPs based on one prior or two prior test scores for ELA/L grades 3 through 11, mathematics grades 3 through 8, Algebra I, Geometry, Algebra II,

[17] Note: Because regression modeling is used to establish the relationship between prior and current scores, the SGP is for students with the exact same prior scores. This often leads to confusion among non-technical stakeholders who often ask, "How many students are there with exactly the same prior scores?" To avoid explaining regression to non-technical stakeholders, the "similar scores" is often used to finesse the idea of regression without mentioning it.

Integrated Mathematics I, II, and III, respectively. In general, the progressions of grade levels and content areas are consecutive. The traditional and integrated mathematics courses have progressions that are not consecutive but reflect student progression for high school mathematics courses. SGPs were calculated for all norm groups with at least 1,000 students. Some progressions did not meet the minimum sample size for SGP calculations.

Table 13.1 ELA/L Grade-Level Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
N/A	N/A	Grade 3*
N/A	Grade 3	Grade 4
Grades 3 and 4	Grade 4	Grade 5
Grades 4 and 5	Grade 5	Grade 6
Grades 5 and 6	Grade 6	Grade 7
Grades 6 and 7	Grade 7	Grade 8
Grades 7 and 8	Grade 8	Grade 9
Grades 8 and 9	Grade 9	Grade 10
Grades 9 and 10	Grade 10	Grade 11

*SGP not calculated for grade 3 since there are no prior scores.

Table 13.2 Mathematics Grade-Level Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
N/A	N/A	Grade 3*
N/A	Grade 3	Grade 4
Grades 3 and 4	Grade 4	Grade 5
Grades 4 and 5	Grade 5	Grade 6
Grades 5 and 6	Grade 6	Grade 7
Grades 6 and 7	Grade 7	Grade 8

*SGP not calculated for grade 3 since there are no prior scores.

Table 13.3 Algebra I Grade/Content Area Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
Grades 5 and 6	Grade 6	Algebra I
Grades 6 and 7	Grade 7	Algebra I
Grades 6 or 7 and 8	Grade 8	Algebra I
Grades 6, 7, or 8 and Geometry	Geometry	Algebra I
Grade 8 and Integrated Math I	Integrated Mathematics I	Algebra I
Integrated Mathematics I and Integrated Mathematics II	Integrated Mathematics II	Algebra I

Table 13.4 Geometry Grade/Content Area Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
Grades 5 and 6	Grade 6	Geometry
Grades 6 and 7	Grade 7	Geometry
Grades 6 or 7 and 8	Grade 8	Geometry
Grades 6, 7, or 8 and Algebra I	Algebra I	Geometry
Grade 8 and Integrated Mathematics I	Integrated Mathematics I	Geometry
Integrated Mathematics I and Integrated Mathematics II	Integrated Mathematics II	Geometry

Table 13.5 Algebra II Grade/Content Area Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
Grades 6 and 7	Grade 7	Algebra II
Grades 7 and 8	Grade 8	Algebra II
Grades 7 or 8 and Algebra I	Algebra I	Algebra II
Grade 8 or Algebra I and Geometry	Geometry	Algebra II
Grade 8 and Integrated Mathematics I	Integrated Mathematics I	Algebra II
Integrated Mathematics I and Integrated Mathematics II	Integrated Mathematics II	Algebra II

Table 13.6 Integrated Mathematics I Grade/Content Area Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
Grades 5 and 6	Grade 6	Integrated Mathematics I
Grades 6 and 7	Grade 7	Integrated Mathematics I
Grades 6 or 7 and 8	Grade 8	Integrated Mathematics I
Grades 7 or 8 and Algebra I	Algebra I	Integrated Mathematics I
Grade 8 or Algebra I and Geometry	Geometry	Integrated Mathematics I

Table 13.7 Integrated Mathematics II Grade/Content Area Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
Grades 6 and 7	Grade 7	Integrated Mathematics II
Grades 7 and 8	Grade 8	Integrated Mathematics II
Grades 7 or 8 and Integrated Mathematics I	Algebra I	Integrated Mathematics II

Table 13.8 Integrated Mathematics III Grade/Content Area Progressions for One- and Two-year Prior Test Scores

Two Prior Year Test Scores	One Prior Year Test Score	Current Year Test Score
Grades 6 and 7	Grade 7	Integrated Mathematics III
Grades 7 and 8	Grade 8	Integrated Mathematics III
Grades 7 or 8 and Integrated Mathematics I	Algebra I	Integrated Mathematics III
Integrated Mathematics I and Integrated Mathematics II	Integrated Mathematics II	Integrated Mathematics III

In addition to the above progressions, in 2018 the PARCC State Leads approved a state-specific SGP progression for one state. In this state, grade 9 students are not required to take the PARCC test. Therefore, grade 10 students were not receiving a SGP. For this state, both mathematics and ELA/L progressions were adjusted (see Table 13.9) such that the grade 10 students would receive growth estimates. Other states were not affected by this change.

Table 13.9 State-specific SGP Progressions

Two Prior Test Scores	One Prior Test Score	Current Test Score
ELA/L Grades 7 and 8	ELA/L Grade 8	ELA/L Grade 10
Mathematics Grade 7 and 8	Mathematics Grade 8	Geometry
Mathematics Grade 7 and Algebra I	Algebra I	Geometry

13.2 Student Growth Percentile Estimation

SGPs are calculated using quantile regression, which describes the conditional distribution of the response variable with greater precision than traditional linear regression, which describes only the conditional mean (Betebenner, 2009). This application of quantile regression uses B-spline smoothing to fit a curvilinear relationship between a norm group's prior and current scores. Cubic B-spline basis functions are used when calculating SGPs to better model the heteroscedasticity, nonlinearity, and skewness in assessment data.

For each group, the quantile regression fits 100 relationships (one for each percentile) between students' prior and current scores. The result is a single coefficient matrix that relates students' prior achievement to their current achievement at each percentile. The National Center for the Improvement of Educational Assessment (NCIEA) performed the analyses using Betebenner's (2009) non-linear quantile-regression based SGP. The analysis was done in the SGP package in R (Betebenner, Van Iwaarden, Domingue, & Shang, 2017). For details on student growth percentiles, see Betebenner's *A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth Projections/Trajectories* (2011).

Betebenner's (2009) SGP model uses Koenker's (2005) quantile regression approach to estimate the conditional density associated with a student's score at administration *t* conditioned on the student's prior score(s). Quantile regression functions represent the solution to a loss function much like least squares regression represents the solution to a minimization of squared deviations. The conditional quantile functions are parametrized as a linear combination of B-spline basis functions (Wei and He, 2006) to smooth irregularities found in the data. For PARCC scores from administration *t* (where $t \geq 2$), the τth quantile function for Y_t conditional on prior scores ($Y_{t-1}, \ldots, Y_1$) is

$$Q_{Yt}(\tau \mid Y_{t-1}, \ldots, Y_1) = \sum_{j=1}^{t-1} \sum_{i=1}^{n} \emptyset_{ij}(Y_j)\beta_{ij}(\tau),$$

where $\emptyset_{i,j}$ (i=1,2,…, n students; j=1, …, t-1 administrations) represent the B-spline basis functions. The SGP of each student *i* is the midpoint between the two consecutive τs whose quantile scores capture the student's current score, multiplied by 100. For example, a student with a current score that lies between the fitted value for $\tau = .595$ and $\tau = .605$ would receive a SGP of 60.

SGPs are assumed to be uniformly distributed and uncorrelated with prior achievement. Scale score conditional standard errors of measurement (CSEMs) were incorporated for calculation of SGP standard errors of measurement (SEMs). Goodness of fit results were checked (i.e., uniform distribution of SGPs by prior achievement) for indications of ceiling/floor effects for each SGP norm-group analysis.

13.3 Student Growth Percentile Results/Model Fit for Total Group

The estimation of SGPs was conducted for each student who had at least one prior score. Each analysis is defined by the norm cohort group (grade/sequence). A goodness of fit plot is produced for each analysis run. A ceiling/floor effects test identifies potential problems at the highest obtainable scale scores (HOSS) and lowest obtainable scale scores (LOSS). Other fit plots compare the observed conditional density of SGP estimates with the theoretical uniform density. If there is perfect model fit, 10 percent of the estimated growth percentiles are expected within each decile band. A Q-Q plot compares the observed distribution with the theoretical distribution; ideally the step function lines do not deviate much from the ideal line of perfect fit.

Tables 13.10 and 13.11 summarize SGP estimates for the total testing group for ELA/L and mathematics, respectively. SGPs were calculated at the consortium level and, if sample size was sufficient, the state level. Consortium-based median SGPs ranged from 49 to 54, with most having a median of 50. If the model is a perfect fit, the median is expected to be 50 with norm-referenced data. The minimum SGP is 1 and the maximum SGP is 99. The average standard error for the consortium-based SGPs is within expectations for these models.

In general, SGPs can be divided into three categories: below 30 indicating that a student is not meeting a year's worth of growth, a SGP of 30–70 indicating that a student did achieve a year's worth of growth, and a SGP over 70 indicating that the student surpassed a year's worth of growth. It is important to note that definitions such as these are not inherent to the SGP method, but rather require expert judgment

(Betenbenner, 2009). The observed standard errors, ranging from 11.6–16.0, support these interpretations (Betenbenner et al., 2016).

Table 13.10 Summary of ELA/L SGP Estimates for Total Group

Grade Level	Sample Size	Average SGP	Average Standard Error	Median SGP
4	328,263	49.99	13.37	50
5	331,900	50.07	13.80	50
6	321,669	50.02	13.79	50
7	316,919	50.00	13.48	50
8	318,518	49.98	14.18	50
9	108,168	49.59	13.60	49
10	164,216	52.49	12.34	54
11	92,016	50.11	11.58	50

Table 13.11 Summary of Mathematics SGP Estimates for Total Group

Grade Level	Sample Size	Average SGP	Average Standard Error	Median SGP
4	328,245	50.03	13.01	50
5	331,745	50.10	14.40	50
6	321,400	50.05	14.73	50
7	305,043	50.02	15.28	50
8	246,636	49.98	16.04	50
A1	167,002	50.16	14.63	50
GO	117,656	50.39	15.11	50
A2	96,045	49.84	15.69	50
M1	--	--	--	--
M2	--	--	--	--
M3	--	--	--	--

Note: "--" indicates insufficient sample for SGP calculation for these tests.

13.4 Student Growth Percentile Results for Subgroups of Interest

Median SGPs are provided for subgroups of interest. With norm-referenced data, the median of all SGPs is expected to be close to 50. Median subgroup growth percentiles below 50 represent growth lower than the consortium median, and median growth percentiles above 50 represent growth higher than the consortium median. Table 13.12 summarizes SGPs for groups of interest for ELA/L grade 4. The ELA/L

tables for grades 5–11 are provided in the Appendix (Tables A.13.1 – A.13.8). Table 13.13 summarizes SGPs for groups of interest for mathematics grade 4; the other mathematics subgroup results are provided in the Appendix (Tables A.13.9 – A.13.17). Median SGPs for subgroups of interest fell within the band of 30–70, which is considered to be adequate growth. Integrated Mathematics had insufficient sample size for SGP subgroup results to be reported.

Gender

English Language Arts/Literacy

The median SGPs for females tend to be higher than the median SGPs for males. The median SGP for females ranges from 50 to 56, whereas the median SGP for males ranges from 45 to 51. The standard error for males and females is comparable to the total group.

Mathematics

There was no consistent pattern between median SGPs for females and males. The median SGP for females ranges from 49 to 53, and the median SGP for males ranges from 47 to 51. The standard errors for both are similar to the total group.

Ethnicity

English Language Arts/Literacy

The African American group median SGP ranges between 44 and 49, with students in higher grades at the higher range. Asian/Pacific Islanders tend to have the highest median SGPs, over 60 for all tests but grade 11. American Indian/Alaska Native had a wide range in median SGPs across grades, with the lowest in grade 4 of 42 and the highest in grade 11 of 63.5. The median SGP for Hispanics tend to be within 3 points of 50, except grade 11 where the median is 56. For all ethnicity groups, standard errors are similar to that of the total group.

Mathematics

The median SGP for African Americans ranges from 41 to 47, with the highest growth in Algebra II. Asian/Pacific Islanders have the highest SGPs across all tests, with a minimum of 58 and a maximum of 63. American Indian/Alaska Native have a minimum SGP of 40 and a maximum of 58. The median SGP for Hispanics is between 47–50, except Algebra I where the median SGP is 44. For all ethnicities, the standard errors for all groups are under 19 points.

Special Instructional Needs

English Language Arts/Literacy

Except for grade 11, economically disadvantaged students and English language learners show lower median growth than the general population. The median SGP ranges from 46 to 55 for economically disadvantaged students and from 45 to 55 for English language learners. Also, students with disabilities tend to show lower growth compared to students without disabilities, with the minimum observed median SGP of 40 in grade 3 and the maximum observed median SGP of 46 in grade 10. The standard errors for special instructional needs subgroups are similar to those observed for the total group.

Mathematics

Economically disadvantaged and English language learner students tend to have lower median SGPs than the general population, with median SGP ranging from 45 to 59 for economically disadvantaged students and from 43 to 49 for English language learners. Similarly, students with disabilities median SGP ranges from 41 in grade 7 to 49 in grade 5, whereas for students without disabilities the median SGP ranges from 50 to 52. The standard errors for special education students are similar to the total group.

Students Taking Spanish Forms

Mathematics

There is a wide range of median growth percentiles for students taking Spanish forms. The number of students in this group tended to be smaller, on average. In Algebra II, 975 students received a SGP. Algebra II had the lowest SGP median of 29. The greatest median SGP for Spanish language forms is 59 for grade 7 mathematics. These forms had a slightly higher standard error on average, likely due to lower sample sizes.

Table 13.12 Summary of SGP Estimates for Subgroups: Grade 4 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	166,982	47.84	13.49	47
Female	161,281	52.21	13.25	53
Ethnicity				
White	135,737	50.89	13.14	51
African American	64,219	45.70	13.76	44
Asian/Pacific Islander	21,899	60.41	12.64	64
American Indian/Alaska Native	4,702	44.04	13.97	42
Hispanic	90,846	49.54	13.60	49
Multiple	10,834	49.40	13.28	49
Special Instructional Needs				
Economically Disadvantaged	162,052	47.15	13.70	46
Not-economically Disadvantaged	166,211	52.75	13.06	54
English Learner (EL)	43,019	48.83	14.09	49
Non English Learner	285,244	50.16	13.27	50
Students with Disabilities (SWD)	54,975	42.72	14.23	40
Students without Disabilities	273,288	51.45	13.20	52

Table 13.13 Summary of SGP Estimates for Subgroups: Grade 4 Mathematics

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	166,998	49.72	13.00	50
Female	161,247	50.35	13.01	50
Ethnicity				
White	135,939	52.61	12.87	54
African American	64,202	43.69	13.19	41
Asian/Pacific Islander	22,280	59.54	13.32	63
American Indian/Alaska Native	4,686	45.90	13.30	44
Hispanic	90,286	48.48	12.98	48
Multiple	10,827	50.42	13.09	50
Special Instructional Needs				
Economically Disadvantaged	161,672	46.31	13.05	45
Not-economically Disadvantaged	166,573	53.64	12.96	55
English Learner (EL)	43,449	46.58	13.31	45
Non English Learner	284,796	50.56	12.96	51
Students with Disabilities (SWD)	54,672	45.72	13.38	44
Students without Disabilities	273,573	50.89	12.93	51
Spanish Language Form	2,734	45.64	12.91	44

Section 14: Quality Control Procedures

Quality control in a testing program is a comprehensive and ongoing process. This section describes procedures put into place to monitor the quality of the item bank, test form, and ancillary material development. The quality checks for scanning, image editing, scoring, and data screening during psychometric analyses are also outlined. Additional quality information can be found in the PARCC Program Quality Plan document.

14.1 Quality Control of the Item Bank

The PARCC summative item bank consists of test passages and items, their associated metadata, and status (e.g., operational-ready, field-test ready, released, etc.). The items on the 2017–2018 assessments were developed by Pearson and West Ed and put in the item bank once created.

The ABBI bank houses the passages and items, art, associated metadata, rubrics, alternate text for use on accommodated forms, and text complexity documentation. It provides an item previewer that allows items to be viewed and interacted with in the same way students see and interact with items and tools, and manages versioning of items with a date/time stamp. It allows PARCC reviewers to vote on item acceptance, and to record and retain their review notes for later reconciliation and reference. PARCC item and passage review committee participants conducted their review in the item banking system. The committee members viewed the items as the student would, and could vote to alter the item, accept or reject the item, and record their comments in the system. After each meeting, reports were forwarded to New Meridian. The reports were generated by the item banking system and summarized feedback from the committee reviewers.

All new development for the PARCC assessments is being created within the ABBI system, which employs templates to control the consistency of the underlying scoring logic and QTI creation for each item type. The ABBI system incorporates a previewer that allows the PARCC reviewers to validate the content of the item and validate the expected scoring of tasks. It supports the full range of PARCC review activities, including content review, bias and sensitivity review, expert editorial review, data review, and test construction review. It provides insight into the item edit process through versioning. A series of metadata validations at key points in the development cycle provide support for metadata consistency. The bank can be queried on the full range of PARCC metadata values to support bank analysis.

14.2 Quality Control of Test Form Development

Test forms were built based upon targets and the established blueprints set by PARCC. The construction process started with specification and requirement capture to create the test specification document. From there items were pulled into forms based on the criteria approved in the test specifications document. After forms composition, the forms went through a review process that involved groups from Pearson and the PARCC states. Quality control steps were conducted on the items and forms evaluating several item characteristics (e.g., content accuracy, completeness, style guide conformity,

tools function). Revisions were incorporated into the forms before final review and approval. Section 2.2 provides more details on the form development process.

The forms quality assurance was performed by Pearson's Assessment and Information Quality (AIQ) organization. AIQ completed a comprehensive review of all *online* forms for the PARCC administration cycle. This group is part of Pearson's larger Organizational Quality group and operates exclusively to validate form operability. The group validates that the functionality of every online form is working to specifications. The overall functionality and maneuverability of each form is checked, and the behavior of each item within the form is verified. (Quality processes for paper forms are described in Section 14.3.)

The items within each form were tested to verify that they operated as expected for test takers. As a further aspect of the testing process, AIQ confirmed that forms were loaded correctly and that the audio was correct when compared to text. Sections and overviews were reviewed. Technology-enhanced items also were tested as an additional measure. As enumerated in the *Technology Guidelines for PARCC Assessments*,[18] user interfaces were compatible with a range of common computer devices, operating systems, and browsers.

Pearson also performed QC tests to verify that a standard set of responses was outputted to the XML as expected after PARCC had approved the final version of the form. These responses were based on the keys provided in the test map or a standard open-ended (OE) responses string that contained a valid range of characters. The test maps also were validated against the form layout and item types for correctness as part of these tests.

Pearson conducted a multifaceted validation of all item layout, rendering, and functionality. Reviewers conducted comparisons between the approved item and the item as it appeared in the field-test form or how it previously appeared, validated that tools and functions in the test delivery system, TestNav, were accurately applied, and verified that the style and layout met all requirements. In addition, answer keys were validated through a formal key review process. More details on the test development procedures are provided in Section 2.

14.3 Quality Control of Test Materials

Pearson provided high quality materials in a timely and efficient manner to meet PARCC's test administration needs. Since the majority of printing work was done in-house, it was possible to fully control the production environment, press schedule, and quality process for print materials. Additionally, strict security requirements were employed to protect secure materials production; Section 3 provides details on the secure handling of test materials. Materials were produced according to the PARCC style guide and to the detailed specifications supplied in the materials list.

[18] This document is available online from: http://avocet.pearson.com/PARCC/Home#10429

Pearson Print Service operates within the sanctions of an ISO 9001:2008 Quality Management System, and practices process improvement through Lean principles and employee involvement.

Raw materials (paper and ink) used for scannable forms production were manufactured exclusively for Pearson Print Service using specifications created by Pearson Print Service. Samples of ink and paper were tested by Pearson prior to use in production. Project specialists were the point of contact for incoming production.

Purchase orders and other order information were assessed against manufacturing capabilities and assigned to the optimal production methodology. PARCC expectations, quality requirements, and cost considerations were foremost in these decisions. Prior to release for manufacture, order information was checked against PARCC specifications, technical requirements, and other communication that includes expected outcomes. Records of these checks were maintained.

Files for image creation flow through one of two file preparation functions: digital pre-press (DPP) for digital print methodology, or plateroom for offset print methodology. Both the DPP and plateroom functions verify content, file naming, imposition, pagination, numbering stream, registration of technical components, color mapping, workflow, and file integrity. Records of these checks are created and saved.

Offset production requires printing that uses a lithographic process. Offline finishing activities are required to create books and package offset output. Digital output may flow through an inkjet digital production line (DPL) or a sheet-fed toner application process in the Xpress Center. A battery of quality checks was performed in these areas. The checks included color match, correct file selection, content match to proof, litho-code to serial number synchronization, registration of technical components, ink density controlled by densitometry, inspection for print flaws, perforations, punching, pagination, scanning requirements, and any unique features specified for the order. Records of these checks and samples pulled from planned production points were maintained. Offline finishing included cutting, shrink-wrapping, folding, and collating. The collation process has three robust inline detection systems that inspected each book for:

- Caliper validation that detects too few or too many pages. This detector will stop the collator if an incorrect caliper reading is registered.
- An optical reader that will only accept one sheet. Two or zero sheets will result in a collator stoppage.
- The correct bar code for the signature being assembled. An incorrect or upside down signature will be rejected by the bar code scanner and will result in a collator stoppage.

Pearson's Quality Assurance (QA) department personnel inspected print output prior to collation and shipment. QA also supported process improvement, work area documentation, audited process adherence, and established training programs for employees.

14.4 Quality Control of Scanning

Establishing and maintaining the accuracy of scanning, editing, and imaging processes is a cornerstone of the Pearson scoring process. While the scanners are designed to perform with great precision,

Pearson implements other quality assurance processes to confirm that the data captured from scan processing produce a complete and accurate map to the expected results.

Pearson pioneered optical mark reading (OMR) and image scanning, and continues to improve in-house scanners for this purpose. Software programs drive the capture of student demographic data and student responses from the test materials during scan processing. Routinely scheduled maintenance and adjustments to the scanner components (e.g., camera) maintain scanner calibration. Test sheets inserted into every batch test scanner accuracy and calibration.

Controlled processes for developing and testing software specifications included a series of validation and verification procedures to confirm the captured data can be mapped accurately and completely to the expected results and that editing application rules are properly applied.

14.5 Quality Control of Image Editing

The final step in producing accurate data for scoring is the editing process. Once information from the documents was captured in the scanning process, the scan program file was executed, comparing the data captured from the student documents to the project specifications. The result of the comparison was a report (or edit listing) of documents needing corrections or validation. Image Editing Services performed the tasks necessary to correct and verify the student data prior to scoring.

Using the report, editors verified that all unscanned documents were scanned, or the data were imported into the system through some other method such as flatbed scan or key entry.

Documents with missing or suspect data were pulled, verified, and corrections or additional data were entered. Standard edits included:

- Incorrect or double gridding
- Incorrect dates (including birth year)
- Mismatches between pre-ID label and gridded information
- Incomplete names

When all edits were resolved, corrections were incorporated into the document file containing student records.

Additional quality checks were also performed. These included student n-count checks to make certain:

- students were placed under the correct header,
- all sheets belonged to the appropriate document,
- documents were not scanned twice, and
- no blank documents existed.

Finally, accuracy checks were performed by checking random documents against scanned data to verify the accuracy of the scanning process.

Once all corrections were made, the scan program was tested a second time to verify all data were valid. When the resulting output showed that no fields were flagged as suspect, the file was considered clean

and scoring began. Once all scanning was completed, the right/wrong response data were securely handed off.

14.6 Quality Control of Answer Document Processing and Scoring

Quality control of answer document processing and scoring involves all aspects of the scoring procedures, including key-based and rule-based machine scoring and handscoring for constructed-response items and performance tasks.

For the 2015 PARCC operational administration, Pearson's validation team prepared test plans used throughout the scoring process. Test plan preparation was organized around detailed specifications.

Based on lessons learned from previous administrations, the following quality steps were implemented:

- Raw score validation (e.g., score key validation; evidence statement, field-test non-score; double-grid combinations; possible correct combination, if applicable; out-of-range/negative test cases)
- Matching (e.g., validation of high-confidence criteria, low-confidence criteria, cross document, external or forced matching by customer; prior to and after data updates; extract file of matched and unmatched documents)
- Demographic update tests (e.g., verification of data extract against corresponding layout; valid values for updatable fields; invalid values for updatable/non-updatable fields; negative test for non-existing record or empty file)

The following components were added to the quality control process specifically for the PARCC program. These additional steps were introduced to address issues with item-level scoring that were identified in the 2014 PARCC field-test administration:

- XML Validation: A combination of automated validation against 100 percent of item XMLs and human inspection of XML from selected difficult item types or composite items.
- Administration/End-to-End Data Validation: An automated generation of response data from approved test maps that have known conditions against the operational scoring systems and data generation systems to verify scoring accuracy.
- Psychometric Validation: Verification of data integrity using criteria typically used in psychometric processes (e.g., statistical keychecks) and categorization of identified issues to help inform investigation by other groups.
- Content Validation: An examination, by subject matter experts, of all items using a combination of automated tools to generate response and scoring data.

In addition to the steps described above, the following quality control process for answer keys and scoring that was implemented for the first PARCC operational administration was used:

1. Pearson's psychometrics team conducted empirical analyses based on preliminary data files and flagged items based on statistical criteria;
2. Pearson content team reviewed the flagged items and provided feedback on the accuracy of content, answer keys, and scoring;
3. Items potentially requiring changes were added to the product validation (PV) log for further investigation by other Pearson teams;

4. PARCC staff was notified of items for which keys or scoring changes were recommended;
5. PARCC approved/rejected scoring changes; and
6. All approved scoring changes were implemented and validated prior to the generation of the data files used for psychometric processing.

14.7 Quality Control of Psychometric Processes

High quality psychometric work for the 2017–2018 PARCC operational administrations was necessary to provide accurate and reliable results of student performance. Pearson, HumRRO, and Measured Progress implemented quality control procedures to ensure the quality of the work including:

1. Well-defined psychometric specifications
2. Consistently applied data cleaning rules
3. Clear and frequent communication
4. Test run analyses
5. Quality checks of the analyses
6. Checklists for statistical procedures

14.7.1 Pearson Psychometric Quality Control Process

Pearson was responsible for the psychometric analyses of the 2017–2018 PARCC operational administration and implemented measures to ensure the quality of work. The psychometric analyses were all conducted according to well-defined specifications. Data cleaning rules were clearly articulated and applied consistently throughout the process. Results from all analyses underwent comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists were used by members of the team for each statistical procedure.

Described below is an overview of the quality control steps performed at different stages of the psychometric analyses. Greater detail is provided in Sections 6 (Classical Item Analysis), 7 (Differential Item Functioning), 10 (IRT Calibration and Scaling), and 12 (Scale Scores).

Data Screening

Data screening is an important first step to ensure quality data input for meaningful analysis. The Pearson Customer Data Quality (CDQ) team validated all student data files used in the operational psychometric analyses. The data validation for the student data files (SDF) and item response files (IRF) included the following steps:

1. Validated variables in the data file for values in acceptable ranges.
2. Validated that the test form ID, unique item numbers (UINs), and item sequence on the data file were consistent with the test form values on the corresponding test map.
3. Computed the composite raw score, claim raw scores, and subclaim raw scores, given the item scores in the student data file.
4. Compared computed raw scores to the raw scores in the student data file.
5. Compared the student item response block (SIRB) to the item scores.
6. Flagged student records with inconsistencies for further investigation.

Pearson Psychometrics and HumRRO established predefined valid case criteria, which were implemented consistently throughout the process. Refer to Section 5.3 for rules for inclusion of students in analyses and Section 10.2 for IRT calibration data preparation criteria and procedures.

Classical Item Analysis

Classical item analysis (IA) produces item level statistics (e.g., item difficulty and item-total correlations). The IA results were reviewed by Pearson psychometricians. Items flagged for unusual statistical properties were reviewed by the content team. A subset of items identified as having key issues, scoring issues, or content issues was presented to the PARCC Priority Alert Task Force, which made decisions on whether to exclude them from the IRT calibration process and, consequently, the calculation of reported student scores. Refer to Section 6.4 for classical IA item flagging criteria.

Calibrations

Creation of item response theory (IRT) sparse data matrices is an important step before the calibrations can begin. Using the same scored item response data, Pearson and HumRRO teams filtered the data and generated their own sparse data matrices independently. Processing of all data was done in parallel by two psychometricians and compared for number of students. This verification of the data preparation was important to ensure that student exclusion rules were applied consistently across the analyses.

During the calibration process, checks were made to ensure that the correct options for the analyses were selected. Checks were also made on the number of items, number of test takers with valid scores, IRT item difficulties, standard errors for the item difficulties, and the consistency between selected IRT statistics to the corresponding statistics obtained during item analyses. Psychometricians also performed detailed reviews of statistics to investigate the extent to which the assumptions of the model fit the observed data. Refer to Section 10.4 for IRT model fit evaluation criteria.

Scaling

During the scaling process, checks were made on the number of linking items, the number of items that were excluded from linking during the stability check of the scaling process, and the scaling constants. Linking items that did not meet the anchor criteria were excluded as linking items. For example, C-DIF items flagged in the mode comparability study were dropped. Additionally, items with large weighted root mean square difference (WRMSD) values in Round 1 of scaling were excluded as linking items in Round 2. Finally, reviewers computed the linking constants and then checked that the linking constants were correctly applied. Refer to Section 10.6 for a description of the scaling process.

Conversion Tables

Conversion tables must be accurate because they are used to generate reported scores for test takers. Comprehensive records were meticulously maintained on item-level decisions, and thorough checks were made to ensure that the correct items were included in the final score. All conversion tables were processed in parallel by Pearson and HumRRO and completely matched. A reasonableness check was also conducted by psychometricians for each content and grade level to make sure the results were in alignment with observations during the analyses prior to conversion table creation. Refer to Section 12.3 for the procedure to create conversion tables.

Delivering Item Statistics

Item statistics based on classical item analyses and IRT analyses were obtained during the psychometric analysis process. The statistics were compiled by two data analysts independently to ensure that the correct statistics were delivered for the item bank.

14.7.2 HumRRO Psychometric Quality Control Process

HumRRO served as the psychometric replicator for the 2017–2018 PARCC operational administration. HumRRO replicated the IRT analyses, scaling analyses, and the conversion file creations. The following steps outline the replication process:

1. Calibrated online and paper data separately.
2. Scaled the paper item parameter estimates to the online scale.
3. Sent the item parameter estimates and scaling constants to Measured Progress for comparison.
4. Reconciled differences, if any, in results with Pearson and Measured Progress.
5. Sent data files to Measured Progress for comparison and reconciled differences, if any.
6. Generated the performance levels, summative, claim, and subclaim conversion tables.
7. Sent conversion tables to Measured Progress for comparison and reconciled differences, if any.

14.7.3 Measured Progress Psychometric Quality Control Process

Measured Progress (MP) served as the external evaluator for the 2017–2018 PARCC operational administration. MP's main task was to evaluate the reasonableness of IRT calibration results, and to compare to the IRT calibration results, scaling constants, summative, claim, and subclaim conversion tables created by HumRRO and Pearson.

IRT Calibrations Comparison

MP reviewed and compared the psychometric IRT calibrations performed primarily by Pearson and HumRRO for all grade levels in ELA/L and mathematics administered both online and on paper.

Pearson and HumRRO each provided comparison files containing IRT item parameter estimates, IRT model fit statistics, and classical item statistics (item-level mean score, item-total correlation). Pearson also provided the IRT model fit plots for the items. For each test, the reasonableness of IRT parameters and the comparability of IRT parameter estimates between Pearson and HumRRO were evaluated on the following aspects:

- Number of items and types of interventions in the IRT calibration process
- Descriptive statistics of the IRT *a*-, *b*-, and *d*-parameter estimates
- Scatterplot of IRT *a*-, *b*-, and *d*-parameter estimates
- Absolute differences in IRT *a*-, *b*-, and *d*-parameter estimates
- Mean absolute difference (MAD) and root mean square difference (RMSD) in IRT-model-predicted item mean scores if there were nontrivial absolute differences in IRT parameter estimates

- IRT model fit statistics and plots
- Item parameter linking status for paper forms

IRT Comparison Results

In general, MP observed highly comparable IRT item parameter estimates between Pearson and HumRRO across all grades and subjects and in both online and paper forms in the 2017–2018 PARCC operational analyses. The largest differences in item parameter estimates occurred at the fourth decimal place. In general, model fit ranged from good to reasonable, with a few items illustrating more variability when sample sizes were small. For a very few items across all the tests, MP observed extreme IRT parameter estimates and/or standard errors, and sent the findings to Pearson for further investigation. Those items were sent to the Priority Alert Task Force for review, and they were either spoiled in operational scoring or flagged for re-field-testing with larger sample sizes.

Conversion Files Comparison

MP provided comparison results for the scaling constants, the summative, claim, and subclaim score conversion tables and their performance levels of both regular and accommodated forms in each of the grades 3 through 8 and high school tests.

The conversion tables were evaluated and compared in the following aspects:

- Form ID and the number of total score points
- Minimum and maximum score points
- Raw cut-scores
- Theta and scaled scores associated with each raw score point

Conversion Files Comparison Results

MP observed identical lower and upper limits of scale scores for summative and subclaim performance levels in each test. In the final comparison files and after any observed differences were reconciled, MP observed identical form IDs and number of raw score points for each test. The largest differences in the theta points only showed on the fourth decimals across all the conversion tables. Only very few scale score points differed by one point due to rounding out of the numerous forms across the grades and subjects. None of the differences occurred on the scale cut-scores. Overall, the final conversion files provided by HumRRO and Pearson were highly comparable.

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.

Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012, January). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Austin, TX: Pearson. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/Bulletin21_Evidence_Based_Standard_Setting.pdf

Betebenner, D. W. (2009). *Norm-and criterion-referenced student growth.* Educational Measurement: Issues and Practice, 28(4), 42-51.

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories.* The National Center for the Improvement of Educational Assessment.

Betebenner, D. W., DePascale, C., Marion, S., Domaleski, C., Martineau, J. (2016). *Precision, interpretability, and utility of SGPs: A Response to why we should abandon student growth percentiles by Sireci, Wells, and Keller.* The National Center for the Improvement of Educational Assessment.

Betebenner, D. W., Van Iwaarden, A., Domingue, B. and Shang, Y. (2017). SGP: Student Growth Percentiles & Percentile Growth Trajectories. (R package version 1.7-7.13)

Brandt, R., Bercovitz, E., McNally, S., Zimmerman, L. (2015). Drawing response interaction Usability Study for PARCC. Austin, TX: Pearson.

Brandt, R., Bercovitz, E., & Zimmerman, L. (2015). Drawing response interaction Usability Study for PARCC. Austin, TX: Pearson.

Brennan, R. L. (2004). BB-CLASS [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment (CASMA).

Cai, L., Thissen, D., & du Toit, S. (2011). IRTPRO 2.1 [Computer software]. Lincolnwood, IL: Scientific Software International.

Clark, M., Skorupski, W., Jirka, S., McBride, M., Wang, C., Murphy, S. (2014). *An investigation into statistical methods to identify aberrant response patterns*. Retrieved from https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/efficacy-and-research/schools/021_Data-Forensics-RD-Research-Report-Person-Fit.pdf.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* 78(1), 98-104.

Council of Chief State School Officers. (2015). Common Core State Standards Initiative. Retrieved from http://www.corestandards.org/

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Doorey, N. & Polikoff, M. (2016). *Evaluating the Content and Quality of Next Generation Assessments*. Washington, DC: Thomas B. Fordham Institute.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach. (ETS Research Report No. 91-47).* Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 35-66). Hillsdale, NJ: Erlbaum.

Dorans, N. J. (2013). *ETS Contributions to the Quantitative Assessment of Item, Test and Score Fairness (ETS R&D Science and Policy Contributions Series, ETS SPC-13-04).* Princeton, NJ: Educational Testing Service.

Feldt, L. S., & Brennan, R. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York: Macmillan.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Holland, P. W., & Thayer, D. T. (1988). Differential item performances and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Huynh, H., & Meyer, P. (2010). Use of robust z in detecting unstable items in item response theory models. *Practical Assessment, Research & Evaluation*, 15(2), 1-8.

Kim, S., & Kolen M. J. (2004). *STUIRT: A Computer Program for Scale Transformation under Unidimensional Item Response Theory Models.* University of Iowa. Version 1.0.

Kolen, M.J., & Brennan, R.L. (2004). *Test equating, scaling, and linking: Methods and practices*. New York: Springer.

Kolen, M. J. (2004). POLYCSEM windows console version [Computer software]. Iowa City, IA: The Center for Advanced Studies in Measurement and Assessment (CASMA), The University of Iowa.

Kolen, M. J., Zeng, L., & Hanson, B. A. (1996). Conditional standard errors of measurement for scale scores using IRT. *Journal of Educational Measurement*, 33(2), 129-140.

Landis, J.R.; Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics. 33 (1): 159–174.

Livingston, S. A., & Lewis, C. (1993). *Estimating the consistency and accuracy of classifications based on test scores*. ETS Research Report Series, No. RR-93-48, Princeton, NJ: ETS.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*, 179-197.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lord, F. M., & Wingersky, M.S. (1984). Comparison of IRT true score and equipercentile observed-score "equatings." *Applied Psychological Measurement, 8*, 453-461.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690-700.

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.

McClarty, K. L., Korbin, J. L., Moyer, E, Griffin, S., Huth, K., Carey, S., & Medberry, S. (2015). *PARCC Benchmarking Study*. Pearson Educational Measurement, Iowa City, IA: Pearson.

McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-based standard setting: establishing a validity framework for cut scores. *Educational Researcher*, 42(2), 78-88.

Minchen, N. LaSalle, A., & Boyd, A. (2018). Operational study 4: Accessibility of new items/functionality component 4 report. Austin, TX: Pearson.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, *16(2),* 159-176.

Muraki, E. & Bock, R. D. (1997). *PARSCALE: IRT item analysis and test scoring for rating-Scale data*. Chicago, IL: Scientific Software International.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982), "The Polyserial Correlation Coefficient," *Biometrika*, 47, 337–347.

Pike, C.K. & Hudson, W.W. (1998). Reliability and measurement error in the presence of homogeneity. *Journal of Social Service Research*, 24, 149-163.

Plake, B. S., Ferdous, A. A., Impara, J. C., & Buckendahl, C. W. (2005). Setting Multiple Performance Standards Using the Yes/No Method: An Alternative Item Mapping Method. *Meeting of the National Council on Measurement in Education.* Montreal. Canada.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, (8), 350-353.

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*. (HumRRO Report 2016 No. 001). Alexandria, VA: Human Resources Research Organization.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.

Steedle, J., & LaSalle, A. (2016). Operational study 4: Accessibility of new items/functionality component 3 report. Austin, TX: Pearson.

Steedle, J., McBride, M. Johnson, M., & Keng, L. (2015). *Spring 2015 Digital Devices Comparability Research Study*. Pearson Educational Measurement, Iowa City, IA: Pearson.

Steedle, J., Quesen, S., Boyd, A. (2017). Longitudinal Study of External Validity of the PARCC Performance Levels: Phase I Report. Austin, TX: Pearson.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education,* 2, 53-55. DOI: 10.5116/ijme.4dfb.8dfd

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal Design Applied to Large Scale Assessments. Synthesis Report.

Wainer, H., & Thissen, D. (2001). Test scoring. Mahwah, NJ: Erlbaum.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practices, 31,* 2–13.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5,* 245-262.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*, 125-145.

Zenisky, A. L., Hambleton, R. K., & Sireci, S.C. (2003). *Effects of local dependence on the validity of IRT item test, and ability statistics*. (Technical Report). American College Admissions Test.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–348). Hillsdale, NJ: Erlbaum

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). *Describing and Categorizing DIF in Polytomous Items (ETS Research Report RR-97-05).* Princeton, NJ: Educational Testing Service.

Appendices

Appendix 5: Test Takers by Grade and Mode, for Each State

Table A.5.1 ELA/L Test Takers, by State, and Grade

		English Language Arts/Literacy									
State	**Category**	**Total**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**	**Grade 9**	**Grade 10**	**Grade 11**
PARCC	N of Students	2,475,049	338,927	345,483	348,524	344,520	338,731	339,283	128,229	188,597	102,755
PARCC	N of CBT	2,348,753	296,516	304,903	333,909	333,518	331,387	331,657	127,586	187,426	101,851
PARCC	% of CBT	94.9	87.5	88.3	95.8	96.8	97.8	97.8	99.5	99.4	99.1
PARCC	N of PBT	126,296	42,411	40,580	14,615	11,002	7,344	7,626	643	1,171	904
PARCC	% of PBT	5.1	12.5	11.7	4.2	3.2	2.2	2.2	0.5	0.6	0.9
BIE	N of Students	8,647	1,372	1,461	1,409	1,321	1,092	1,029	224	197	542
BIE	% of PARCC Data	0.3	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0
BIE	N of CBT	2,571	341	383	397	377	386	376	56	58	197
BIE	% of CBT	29.7	24.9	26.2	28.2	28.5	35.3	36.5	25.0	29.4	36.3
BIE	N of PBT	6,076	1,031	1,078	1,012	944	706	653	168	139	345
BIE	% of PBT	70.3	75.1	73.8	71.8	71.5	64.7	63.5	75.0	70.6	63.7
DC	N of Students	40,105	6,372	6,085	5,921	5,038	4,697	4,285	3,396	4,258	53
DC	% of PARCC Data	1.6	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.0
DC	N of CBT	39,955	6,343	6,035	5,892	5,023	4,687	4,280	3,393	4,249	53
DC	% of CBT	99.6	99.5	99.2	99.5	99.7	99.8	99.9	99.9	99.8	100
DC	N of PBT	150	29	50	29	15	10	5	3	9	0
DC	% of PBT	0.4	0.5	0.8	0.5	0.3	0.2	0.1	0.1	0.2	0.0
DD	N of Students	17,146	n/a	n/a	n/a	5,177	4,638	4,244	n/a	3,087	n/a
DD	% of PARCC Data	0.7	n/a	n/a	n/a	0.2	0.2	0.2	n/a	0.1	n/a
DD	N of CBT	17,011	n/a	n/a	n/a	5,134	4,614	4,207	n/a	3,056	n/a
DD	% of CBT	99.2	n/a	n/a	n/a	99.2	99.5	99.1	n/a	99.0	n/a
DD	N of PBT	135	n/a	n/a	n/a	43	24	37	n/a	31	n/a
DD	% of PBT	0.8	n/a	n/a	n/a	0.8	0.5	0.9	n/a	1.0	n/a

		English Language Arts/Literacy									
State	**Category**	**Total**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**	**Grade 9**	**Grade 10**	**Grade 11**
IL	N of Students	859,725	141,226	144,266	146,243	143,620	142,125	142,245	n/a	n/a	n/a
IL	% of PARCC Data	34.7	5.7	5.8	5.9	5.8	5.7	5.7	n/a	n/a	n/a
IL	N of CBT	743,479	100,103	105,081	132,950	133,881	135,831	135,633	n/a	n/a	n/a
IL	% of CBT	86.5	70.9	72.8	90.9	93.2	95.6	95.4	n/a	n/a	n/a
IL	N of PBT	116,246	41,123	39,185	13,293	9,739	6,294	6,612	n/a	n/a	n/a
IL	% of PBT	13.5	29.1	27.2	9.1	6.8	4.4	4.6	n/a	n/a	n/a
MD	N of Students	477,509	67,676	69,398	68,636	65,610	63,905	63,835	2,427	63,484	12,538
MD	% of PARCC Data	19.3	2.7	2.8	2.8	2.7	2.6	2.6	0.1	2.6	0.5
MD	N of CBT	476,491	67,620	69,318	68,555	65,548	63,846	63,778	2,423	62,974	12,429
MD	% of CBT	99.8	99.9	99.9	99.9	99.9	99.9	99.9	99.8	99.2	99.1
MD	N of PBT	1,018	56	80	81	62	59	57	4	510	109
MD	% of PBT	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.8	0.9
NJ	N of Students	857,603	98,223	99,376	100,992	99,318	98,159	99,994	98,754	94,567	68,220
NJ	% of PARCC Data	34.6	4.0	4.0	4.1	4.0	4.0	4.0	4.0	3.8	2.8
NJ	N of CBT	856,007	98,124	99,267	100,880	99,215	98,053	99,866	98,427	94,246	67,929
NJ	% of CBT	99.8	99.9	99.9	99.9	99.9	99.9	99.9	99.7	99.7	99.6
NJ	N of PBT	1,596	99	109	112	103	106	128	327	321	291
NJ	% of PBT	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.3	0.4
NM	N of Students	214,093	24,050	24,891	25,309	24,423	24,104	23,636	23,408	22,871	21,401
NM	% of PARCC Data	8.7	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.9
NM	N of CBT	213,019	23,977	24,813	25,221	24,327	23,959	23,503	23,267	22,710	21,242
NM	% of CBT	99.5	99.7	99.7	99.7	99.6	99.4	99.4	99.4	99.3	99.3
NM	N of PBT	1,074	73	78	88	96	145	133	141	161	159
NM	% of PBT	0.5	0.3	0.3	0.3	0.4	0.6	0.6	0.6	0.7	0.7

Note: BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico; CBT = computer-based test; PBT = paper-based test; n/a = not applicable.

Table A.5.2 Mathematics Test Takers, by State, and Grade

		Mathematics												
State	**Category**	**Total**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**	**A1**	**GO**	**A2**	**M1**	**M2**	**M3**
PARCC	N of Students	2,468,267	348,117	354,080	355,854	345,712	323,440	263,809	221,242	130,412	123,787	750	873	191
PARCC	N of CBT	2,341,341	305,323	313,273	341,143	334,675	316,116	256,332	220,012	129,723	122,954	750	854	186
PARCC	% of CBT	94.9	87.7	88.5	95.9	96.8	97.7	97.2	99.4	99.5	99.3	100	97.8	97.4
PARCC	N of PBT	126,926	42,794	40,807	14,711	11,037	7,324	7,477	1,230	689	833	n/a	19	5
PARCC	% of PBT	5.1	12.3	11.5	4.1	3.2	2.3	2.8	0.6	0.5	0.7	n/a	2.2	2.6
BIE	N of Students	8,627	1,371	1,456	1,407	1,318	1,089	1,019	227	240	483	n/a	17	n/a
BIE	% of PARCC Data	0.3	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	n/a	0.0	n/a
BIE	N of CBT	2,569	340	384	396	373	381	364	79	63	189	n/a	0	n/a
BIE	% of CBT	29.8	24.8	26.4	28.1	28.3	35.0	35.7	34.8	26.3	39.1	n/a	0.0	n/a
BIE	N of PBT	6,058	1,031	1,072	1,011	945	708	655	148	177	294	n/a	17	n/a
BIE	% of PBT	70.2	75.2	73.6	71.9	71.7	65.0	64.3	65.2	73.8	60.9	n/a	100	n/a
DC	N of Students	39,189	6,376	6,083	5,914	5,029	4,591	3,480	3,323	3,903	345	n/a	145	n/a
DC	% of PARCC Data	1.6	0.3	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.0	n/a	0.0	n/a
DC	N of CBT	39,041	6,348	6,033	5,885	5,014	4,580	3,476	3,318	3,897	345	n/a	145	n/a
DC	% of CBT	99.6	99.6	99.2	99.5	99.7	99.8	99.9	99.8	99.8	100	n/a	100	n/a
DC	N of PBT	148	28	50	29	15	11	4	5	6	0	n/a	0	n/a
DC	% of PBT	0.4	0.4	0.8	0.5	0.3	0.2	0.1	0.2	0.2	0.0	n/a	0.0	n/a
DD	N of Students	32,249	6,085	5,997	5,915	5,163	n/a	n/a	3,541	3,085	2,463	n/a	n/a	n/a
DD	% of PARCC Data	1.3	0.2	0.2	0.2	0.2	n/a	n/a	0.1	0.1	0.1	n/a	n/a	n/a
DD	N of CBT	31,910	6,002	5,932	5,856	5,123	n/a	n/a	3,514	3,048	2,435	n/a	n/a	n/a
DD	% of CBT	98.9	98.6	98.9	99.0	99.2	n/a	n/a	99.2	98.8	98.9	n/a	n/a	n/a
DD	N of PBT	339	83	65	59	40	n/a	n/a	27	37	28	n/a	n/a	n/a
DD	% of PBT	1.1	1.4	1.1	1.0	0.8	n/a	n/a	0.8	1.2	1.1	n/a	n/a	n/a
IL	N of Students	862,339	141,946	144,832	146,506	144,009	141,727	138,510	4,344	335	111	17	2	n/a
IL	% of PARCC Data	34.9	5.8	5.9	5.9	5.8	5.7	5.6	0.2	0.0	0.0	0.0	0.0	n/a
IL	N of CBT	745,552	100,526	105,479	133,175	134,240	135,442	131,946	4,280	334	111	17	2	n/a
IL	% of CBT	86.5	70.8	72.8	90.9	93.2	95.6	95.3	99	100	100	100	100	n/a
IL	N of PBT	116,787	41,420	39,353	13,331	9,769	6,285	6,564	64	1	0	0	0	n/a
IL	% of PBT	13.5	29.2	27.2	9.1	6.8	4.4	4.7	1.5	0.3	0.0	0.0	0.0	n/a

			Mathematics											
State	**Category**	**Total**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**	**A1**	**GO**	**A2**	**M1**	**M2**	**M3**
MD	N of Students	475,484	68,024	69,733	68,908	65,577	57,104	37,298	74,767	11,190	22,883	n/a	n/a	n/a
MD	% of PARCC Data	19.3	2.8	2.8	2.8	2.7	2.3	1.5	3.0	0.5	0.9	n/a	n/a	n/a
MD	N of CBT	474,458	67,969	69,654	68,829	65,514	57,048	37,253	74,258	11,181	22,752	n/a	n/a	n/a
MD	% of CBT	99.8	99.9	99.9	99.9	99.9	99.9	99.9	99.3	99.9	99.4	n/a	n/a	n/a
MD	N of PBT	1,026	55	79	79	63	56	45	509	9	131	n/a	n/a	n/a
MD	% of PBT	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.1	0.6	n/a	n/a	n/a
NJ	N of Students	837,080	99,159	100,295	101,634	100,043	94,766	64,305	110,151	89,306	77,421	n/a	n/a	n/a
NJ	% of PARCC Data	33.9	4.0	4.1	4.1	4.1	3.8	2.6	4.5	3.6	3.1	n/a	n/a	n/a
NJ	N of CBT	835,476	99,057	100,184	101,519	99,934	94,648	64,196	109,779	88,996	77,163	n/a	n/a	n/a
NJ	% of CBT	99.8	99.9	99.9	99.9	99.9	99.9	99.8	99.7	99.7	99.7	n/a	n/a	n/a
NJ	N of PBT	1,604	102	111	115	109	118	109	372	310	258	n/a	n/a	n/a
NJ	% of PBT	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.3	0.3	0.3	n/a	n/a	n/a
NM	N of Students	213,086	25,148	25,678	25,556	24,560	24,152	19,183	24,777	22,350	20,081	716	694	191
NM	% of PARCC Data	8.6	1.0	1.0	1.0	1.0	1.0	0.8	1.0	0.9	0.8	0.0	0.0	0.0
NM	N of CBT	212,123	25,073	25,601	25,469	24,464	24,006	19,084	24,672	22,201	19,959	716	692	186
NM	% of CBT	99.5	99.7	99.7	99.7	99.6	99.4	99.5	99.6	99.3	99.4	100	99.7	97.4
NM	N of PBT	963	75	77	87	96	146	99	105	149	122	0	2	5
NM	% of PBT	0.5	0.3	0.3	0.3	0.4	0.6	0.5	0.4	0.7	0.6	0.0	0.3	2.6

Note: BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico; A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; n/a = not applicable.

Table A.5.3 Spanish-Language Mathematics Test Takers, by State, and Grade

Mathematics (Spanish-Language forms)														
State*	**Category**	**Total**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**	**A1**	**GO**	**A2**	**M1**	**M2**	**M3**
PARCC	N of Students	25,601	4,906	3,647	2,960	2,536	2,586	2,108	3,579	1,961	1,300	6	11	1
PARCC	N of CBT	23,980	4,223	3,178	2,916	2,516	2,566	2,103	3,203	1,958	1,300	6	11	n/a
PARCC	% of CBT	93.7	86.1	87.1	98.5	99.2	99.2	99.8	89.5	99.8	100	100	100	n/a
PARCC	N of PBT	1,621	683	469	44	20	20	5	376	3	n/a	n/a	n/a	1
PARCC	% of PBT	6.3	13.9	12.9	1.5	0.8	0.8	0.2	10.5	0.2	n/a	n/a	n/a	100
BIE	N of Students	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	N of CBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	% of CBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	N of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	% of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
DC	N of Students	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
DC	% of PARCC Data	300	53	50	55	35	45	20	24	18	n/a	n/a	n/a	n/a
DC	N of CBT	1.2	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.1	n/a	n/a	n/a	n/a
DC	% of CBT	300	53	50	55	35	45	20	24	18	n/a	n/a	n/a	n/a
DC	N of PBT	100	100	100	100	100	100	100	100	100	n/a	n/a	n/a	n/a
DC	% of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
DD	N of Students	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
DD	% of PARCC Data	30	11	5	3	5	n/a	n/a	2	4	n/a	n/a	n/a	n/a
DD	N of CBT	0.1	0.0	0.0	0.0	0.0	n/a	n/a	0.0	0.0	n/a	n/a	n/a	n/a
DD	% of CBT	28	9	5	3	5	n/a	n/a	2	4	n/a	n/a	n/a	n/a
DD	N of PBT	93.3	81.8	100	100	100	n/a	n/a	100	100	n/a	n/a	n/a	n/a
DD	% of PBT	2	2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
IL	N of Students	6.7	18.2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
IL	% of PARCC Data	7,214	2,242	1,483	1,379	950	732	428	n/a	n/a	n/a	n/a	n/a	n/a
IL	N of CBT	28.2	8.8	5.8	5.4	3.7	2.9	1.7	n/a	n/a	n/a	n/a	n/a	n/a
IL	% of CBT	6,034	1,571	1,027	1,346	941	724	425	n/a	n/a	n/a	n/a	n/a	n/a
IL	N of PBT	83.6	70.1	69.3	97.6	99.1	98.9	99.3	n/a	n/a	n/a	n/a	n/a	n/a
IL	% of PBT	1,180	671	456	33	9	8	3	n/a	n/a	n/a	n/a	n/a	n/a

Mathematics (Spanish-Language forms)														
State*	**Category**	**Total**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**	**A1**	**GO**	**A2**	**M1**	**M2**	**M3**
MD	N of Students	914	31	42	75	44	61	59	566	4	32	n/a	n/a	n/a
MD	% of PARCC Data	3.6	0.1	0.2	0.3	0.2	0.2	0.2	2.2	0.0	0.1	n/a	n/a	n/a
MD	N of CBT	565	31	42	75	43	61	59	218	4	32	n/a	n/a	n/a
MD	% of CBT	61.8	100	100	100	97.7	100	100	38.5	100	100	n/a	n/a	n/a
MD	N of PBT	349	n/a	n/a	n/a	1	n/a	n/a	348	n/a	n/a	n/a	n/a	n/a
MD	% of PBT	38.2	n/a	n/a	n/a	2.3	n/a	n/a	61.5	n/a	n/a	n/a	n/a	n/a
NJ	N of Students	13,668	1,481	1,269	1,170	1,275	1,517	1,368	2,738	1,704	1,146	n/a	n/a	n/a
NJ	% of PARCC Data	53.4	5.8	5.0	4.6	5.0	5.9	5.3	10.7	6.7	4.5	n/a	n/a	n/a
NJ	N of CBT	13,582	1,472	1,257	1,159	1,265	1,506	1,366	2,710	1,701	1,146	n/a	n/a	n/a
NJ	% of CBT	99.4	99.4	99.1	99.1	99.2	99.3	99.9	99.0	99.8	100	n/a	n/a	n/a
NJ	N of PBT	86	9	12	11	10	11	2	28	3	n/a	n/a	n/a	n/a
NJ	% of PBT	0.6	0.6	0.9	0.9	0.8	0.7	0.1	1.0	0.2	n/a	n/a	n/a	n/a
NM	N of Students	3,475	1,088	798	278	227	231	233	249	231	122	6	11	1
NM	% of PARCC Data	13.6	4.2	3.1	1.1	0.9	0.9	0.9	1.0	0.9	0.5	0.0	0.0	0.0
NM	N of CBT	3,471	1,087	797	278	227	230	233	249	231	122	6	11	n/a
NM	% of CBT	99.9	99.9	99.9	100	100	99.6	100	100	100	100	100	100	n/a
NM	N of PBT	4	1	1	n/a	n/a	1	n/a	n/a	n/a	n/a	n/a	n/a	1
NM	% of PBT	0.1	0.1	0.1	n/a	n/a	0.4	n/a	n/a	n/a	n/a	n/a	n/a	100

Note: BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico; A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; n/a = not applicable.

* No students in BIE tested in mathematics using Spanish-language forms.

Table A.5.4 All States Combined: ELA/L Test Takers by Grade, Mode, and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
3	**All**	338,927	166,189	49.0	172,738	51.0
3	CBT	296,516	145,188	49.0	151,328	51.0
3	PBT	42,411	21,001	49.5	21,410	50.5
4	**All**	345,483	169,612	49.1	175,871	50.9
4	CBT	304,903	149,432	49.0	155,471	51.0
4	PBT	40,580	20,180	49.7	20,400	50.3
5	**All**	348,524	170,846	49.0	177,678	51.0
5	CBT	333,909	163,849	49.1	170,060	50.9
5	PBT	14,615	6,997	47.9	7,618	52.1
6	**All**	344,520	169,028	49.1	175,492	50.9
6	CBT	333,518	163,666	49.1	169,852	50.9
6	PBT	11,002	5,362	48.7	5,640	51.3
7	**All**	338,731	165,429	48.8	173,302	51.2
7	CBT	331,387	162,005	48.9	169,382	51.1
7	PBT	7,344	3,424	46.6	3,920	53.4
8	**All**	339,283	165,771	48.9	173,512	51.1
8	CBT	331,657	162,165	48.9	169,492	51.1
8	PBT	7,626	3,606	47.3	4,020	52.7
9	**All**	128,229	63,018	49.1	65,211	50.9
9	CBT	127,586	62,739	49.2	64,847	50.8
9	PBT	643	279	43.4	364	56.6
10	**All**	188,597	92,340	49.0	96,257	51.0
10	CBT	187,426	91,800	49.0	95,626	51.0
10	PBT	1,171	540	46.1	631	53.9
11	**All**	102,755	48,814	47.5	53,941	52.5
11	CBT	101,851	48,406	47.5	53,445	52.5
11	PBT	904	408	45.1	496	54.9

Note: CBT = computer-based tests; PBT = paper-based tests.

Table A.5.5 All States Combined: All Mathematics Test Takers by Grade, Mode, and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
3	**All**	348,117	170,632	49.0	177,485	51.0
3	CBT	305,323	149,437	48.9	155,886	51.1
3	PBT	42,794	21,195	49.5	21,599	50.5
4	**All**	354,080	173,910	49.1	180,170	50.9
4	CBT	313,273	153,611	49.0	159,662	51.0
4	PBT	40,807	20,299	49.7	20,508	50.3
5	**All**	355,854	174,514	49.0	181,340	51.0
5	CBT	341,143	167,474	49.1	173,669	50.9
5	PBT	14,711	7,040	47.9	7,671	52.1
6	**All**	345,712	169,707	49.1	176,005	50.9
6	CBT	334,675	164,323	49.1	170,352	50.9
6	PBT	11,037	5,384	48.8	5,653	51.2
7	**All**	323,440	158,086	48.9	165,354	51.1
7	CBT	316,116	154,669	48.9	161,447	51.1
7	PBT	7,324	3,417	46.7	3,907	53.3
8	**All**	263,809	127,232	48.2	136,577	51.8
8	CBT	256,332	123,710	48.3	132,622	51.7
8	PBT	7,477	3,522	47.1	3,955	52.9
A1	**All**	221,242	107,530	48.6	113,712	51.4
A1	CBT	220,012	107,018	48.6	112,994	51.4
A1	PBT	1,230	512	41.6	718	58.4
GO	**All**	130,412	64,117	49.2	66,295	50.8
GO	CBT	129,723	63,814	49.2	65,909	50.8
GO	PBT	689	303	44.0	386	56.0
A2	**All**	123,787	62,071	50.1	61,716	49.9
A2	CBT	122,954	61,661	50.1	61,293	49.9
A2	PBT	833	410	49.2	423	50.8
M1	**All**	750	338	45.1	412	54.9
M1	CBT	750	338	45.1	412	54.9
M1	PBT	n/a	n/a	n/a	n/a	n/a
M2	**All**	873	450	51.5	423	48.5
M2	CBT	854	437	51.2	417	48.8
M2	PBT	n/r	n/r	n/r	n/r	n/r
M3	**All**	191	94	49.2	97	50.8
M3	CBT	186	92	49.5	94	50.5
M3	PBT	n/r	n/r	n/r	n/r	n/r

Note: Includes students taking English-language mathematics tests, students taking Spanish-language mathematics tests, and students taking accommodated forms. A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.6 All States Combined: Spanish-Language Mathematics Test Takers, by Grade, Mode, and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
3	**All**	4,906	2,394	48.8	2,512	51.2
3	CBT	4,223	2,042	48.4	2,181	51.6
3	PBT	683	352	51.5	331	48.5
4	**All**	3,647	1,825	50.0	1,822	50.0
4	CBT	3,178	1,564	49.2	1,614	50.8
4	PBT	469	261	55.7	208	44.3
5	**All**	2,960	1,403	47.4	1,557	52.6
5	CBT	2,916	1,381	47.4	1,535	52.6
5	PBT	44	22	50.0	22	50.0
6	**All**	2,536	1,195	47.1	1,341	52.9
6	CBT	2,516	1,185	47.1	1,331	52.9
6	PBT	20	10	50.0	10	50.0
7	**All**	2,586	1,219	47.1	1,367	52.9
7	CBT	2,566	1,209	47.1	1,357	52.9
7	PBT	20	10	50.0	10	50.0
8	**All**	2,108	936	44.4	1,172	55.6
8	CBT	2,103	934	44.4	1,169	55.6
8	PBT	n/r	n/r	n/r	n/r	n/r
A1	**All**	3,579	1,654	46.2	1,925	53.8
A1	CBT	3,203	1,492	46.6	1,711	53.4
A1	PBT	376	162	43.1	214	56.9
GO	**All**	1,961	954	48.6	1,007	51.4
GO	CBT	1,958	951	48.6	1,007	51.4
GO	PBT	n/r	n/r	n/r	n/r	n/r
A2	**All**	1,300	644	49.5	656	50.5
A2	CBT	1,300	644	49.5	656	50.5
A2	PBT	n/a	n/a	n/a	n/a	n/a
M1	**All**	n/r	n/r	n/r	n/r	n/r
M1	CBT	n/r	n/r	n/r	n/r	n/r
M1	PBT	n/a	n/a	n/a	n/a	n/a
M2	**All**	n/r	n/r	n/r	n/r	n/r
M2	CBT	n/r	n/r	n/r	n/r	n/r
M2	PBT	n/a	n/a	n/a	n/a	n/a
M3	**All**	n/r	n/r	n/r	n/r	n/r
M3	CBT	n/a	n/a	n/a	n/a	n/a
M3	PBT	n/r	n/r	n/r	n/r	n/r

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.7 Demographic Information for Grade 3 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	49.6	98.5	79.0	n/a	53.8	40.4	40.7	76.9
SWD (%)	15.6	12.6	16.6	n/a	14.4	14.1	18.6	15.2
EL (%)	14.9	26.4	14.9	n/a	19.6	13.6	8.1	18.2
Male (%)	51.0	50.6	51.3	n/a	51.0	50.8	51.2	50.5
Female (%)	49.0	49.4	48.7	n/a	49.0	49.2	48.8	49.5
AmInd/ANat (%)	1.5	98.3	n/r	n/a	0.3	0.2	0.1	12.0
Asian (%)	6.6	n/a	1.5	n/a	5.1	6.4	10.6	1.5
Black/AA (%)	20.0	n/r	66.5	n/a	17.4	33.4	15.6	3.7
Hisp/Lat (%)	27.8	n/a	17.0	n/a	26.9	18.6	29.4	57.7
Wh/Caus (%)	40.6	n/a	12.4	n/a	46.5	36.2	41.5	24.5
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/a	0.1	0.2	0.2	0.4
Two or More (%)	3.4	n/r	2.5	n/a	3.8	5.0	2.6	0.1
Unknown (%)	0.0	n/r	n/r	n/a	0.0	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.8 Demographic Information for Grade 4 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	49.5	98.2	81.1	n/a	53.9	40.4	40.0	77.2
SWD (%)	16.7	13.6	18.4	n/a	15.3	15.1	19.8	16.4
EL (%)	13.7	32.2	14.3	n/a	18.9	11.4	6.1	18.6
Male (%)	50.9	48.1	50.3	n/a	50.8	51.2	50.9	51.0
Female (%)	49.1	51.9	49.7	n/a	49.2	48.8	49.1	49.0
AmInd/ANat (%)	1.5	98.2	n/r	n/a	0.4	0.3	0.1	11.6
Asian (%)	6.6	n/a	1.4	n/a	5.0	6.5	10.7	1.5
Black/AA (%)	19.8	n/r	68.1	n/a	17.0	34.1	15.5	3.6
Hisp/Lat (%)	27.7	n/a	17.8	n/a	27.1	17.8	28.8	58.9
Wh/Caus (%)	40.8	n/r	10.3	n/a	46.7	36.2	42.3	23.8
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/a	0.1	0.2	0.2	0.4
Two or More (%)	3.3	n/r	2.1	n/a	3.8	4.9	2.4	0.2
Unknown (%)	0.0	n/r	n/r	n/a	0.0	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.9 Demographic Information for Grade 5 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	49.1	98.4	79.8	n/a	53.5	39.8	39.4	77.5
SWD (%)	17.1	13.6	19.1	n/a	15.5	15.8	20.2	16.8
EL (%)	8.5	36.7	11.6	n/a	10.8	7.0	3.9	16.0
Male (%)	51.0	50.5	50.0	n/a	51.0	51.1	50.9	51.3
Female (%)	49.0	49.5	50.0	n/a	49.0	48.9	49.1	48.7
AmInd/ANat (%)	1.4	97.8	n/r	n/a	0.3	0.3	0.1	11.6
Asian (%)	6.4	n/r	1.2	n/a	4.9	6.5	10.3	1.5
Black/AA (%)	19.5	n/a	68.2	n/a	16.9	33.4	15.4	3.1
Hisp/Lat (%)	28.0	n/r	17.2	n/a	27.4	18.0	28.8	59.8
Wh/Caus (%)	41.3	n/r	10.9	n/a	46.9	37.0	43.0	23.5
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/a	0.1	0.2	0.2	0.4
Two or More (%)	3.1	n/r	2.3	n/a	3.6	4.6	2.2	0.1
Unknown (%)	0.0	n/r	n/r	n/a	n/r	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.10 Demographic Information for Grade 6 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	47.2	98.6	77.9	n/r	52.5	39.2	38.3	75.3
SWD (%)	16.9	14.5	20.6	14.4	15.5	15.8	19.6	16.4
EL (%)	6.1	44.5	7.3	7.5	7.0	5.1	3.0	12.8
Male (%)	50.9	47.4	51.0	49.7	50.9	51.2	51.0	50.6
Female (%)	49.1	52.6	49.0	50.3	49.1	48.8	49.0	49.4
AmInd/ANat (%)	1.4	98.6	n/r	0.5	0.3	0.3	0.1	11.7
Asian (%)	6.5	n/a	1.5	6.6	4.9	6.6	10.3	1.5
Black/AA (%)	19.1	n/a	70.0	11.8	16.6	33.6	15.1	3.1
Hisp/Lat (%)	27.4	n/r	17.1	20.7	27.1	17.4	27.9	59.2
Wh/Caus (%)	42.2	n/a	9.0	44.0	47.6	37.3	44.3	24.0
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.8	0.1	0.2	0.2	0.4
Two or More (%)	3.1	n/r	2.1	12.8	3.4	4.6	2.0	0.1
Unknown (%)	0.0	n/r	n/a	1.8	n/r	n/a	n/a	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.11 Demographic Information for Grade 7 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	45.7	98.2	77.5	n/r	50.9	37.5	36.7	73.6
SWD (%)	16.9	15.7	20.2	13.9	15.5	15.7	19.9	16.3
EL (%)	5.6	37.5	7.3	7.1	6.3	5.1	3.1	11.0
Male (%)	51.2	49.3	50.1	49.2	51.4	50.9	51.2	50.6
Female (%)	48.8	50.7	49.9	50.8	48.6	49.1	48.8	49.4
AmInd/ANat (%)	1.3	97.0	n/r	n/r	0.3	0.3	0.1	11.4
Asian (%)	6.6	n/r	1.4	6.7	5.1	6.9	10.3	1.6
Black/AA (%)	18.7	n/r	70.3	12.0	15.9	33.6	14.8	3.0
Hisp/Lat (%)	27.1	n/r	16.6	21.6	26.8	16.7	27.3	59.5
Wh/Caus (%)	43.1	n/a	9.6	42.9	48.5	38.2	45.4	24.1
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.9	0.1	0.2	0.2	0.3
Two or More (%)	2.9	n/r	1.8	12.3	3.3	4.3	1.8	n/r
Unknown (%)	0.0	n/r	n/a	2.3	n/r	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.12 Demographic Information for Grade 8 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	44.3	99.2	78.2	n/r	49.8	35.8	35.0	73.1
SWD (%)	16.7	14.7	20.9	13.1	15.3	15.8	19.7	15.5
EL (%)	5.3	37.9	7.1	6.7	5.6	4.9	3.1	12.0
Male (%)	51.1	49.8	49.5	51.9	51.3	50.7	51.3	50.9
Female (%)	48.9	50.2	50.5	48.1	48.7	49.3	48.7	49.1
AmInd/ANat (%)	1.3	98.0	n/r	n/r	0.2	0.2	0.1	11.7
Asian (%)	6.7	n/a	1.6	6.9	5.1	6.8	10.4	1.5
Black/AA (%)	18.6	n/a	70.8	12.0	15.9	33.6	14.7	2.9
Hisp/Lat (%)	26.3	n/a	16.4	20.2	26.3	15.6	26.2	59.6
Wh/Caus (%)	44.1	n/r	9.5	43.1	49.0	39.3	46.8	23.9
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.9	0.1	0.1	0.2	0.3
Two or More (%)	2.9	n/r	1.6	12.9	3.3	4.3	1.6	0.1
Unknown (%)	0.0	n/r	n/a	2.6	n/r	n/a	n/a	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.13 Demographic Information for Grade 9 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	42.1	98.7	81.9	n/a	n/a	48.7	34.4	67.7
SWD (%)	17.8	9.4	18.3	n/a	n/a	15.8	18.7	14.5
EL (%)	6.6	34.8	7.9	n/a	n/a	2.1	5.4	11.7
Male (%)	50.9	49.1	50.3	n/a	n/a	51.1	51.0	50.5
Female (%)	49.1	50.9	49.7	n/a	n/a	48.9	49.0	49.5
AmInd/ANat (%)	2.5	98.2	n/r	n/a	n/a	n/r	0.1	12.0
Asian (%)	8.3	n/a	1.9	n/a	n/a	1.9	10.2	1.7
Black/AA (%)	13.9	n/a	70.6	n/a	n/a	19.2	14.5	3.0
Hisp/Lat (%)	32.5	n/a	16.8	n/a	n/a	9.2	27.3	59.4
Wh/Caus (%)	41.3	n/r	9.2	n/a	n/a	63.7	46.1	23.4
NtvHawaii/Pacific (%)	0.2	n/a	n/r	n/a	n/a	n/r	0.2	0.3
Two or More (%)	1.4	n/a	1.3	n/a	n/a	5.8	1.6	0.1
Unknown (%)	n/r	n/r	n/a	n/a	n/a	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.14 Demographic Information for Grade 10 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	37.0	99.5	83.4	n/a	n/a	31.3	33.1	64.7
SWD (%)	17.0	16.8	20.1	9.9	n/a	16.2	18.5	12.8
EL (%)	6.6	29.9	11.3	3.9	n/a	7.2	4.8	11.9
Male (%)	51.0	52.8	50.2	49.9	n/a	51.5	51.0	50.4
Female (%)	49.0	47.2	49.8	50.1	n/a	48.5	49.0	49.6
AmInd/ANat (%)	1.7	98.5	n/r	n/r	n/a	0.3	0.1	12.3
Asian (%)	7.8	n/a	1.4	9.2	n/a	6.8	10.1	1.8
Black/AA (%)	21.5	n/a	69.0	11.6	n/a	35.8	14.6	3.0
Hisp/Lat (%)	26.7	n/a	20.2	20.8	n/a	16.3	26.4	58.9
Wh/Caus (%)	39.8	n/r	7.7	42.0	n/a	36.9	47.3	23.6
NtvHawaii/Pacific (%)	0.2	n/a	n/r	2.4	n/a	0.1	0.2	0.3
Two or More (%)	2.2	n/r	1.5	12.3	n/a	3.9	1.3	0.1
Unknown (%)	0.0	n/r	n/r	1.2	n/a	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.15 Demographic Information for Grade 11 ELA/L, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	44.2	98.2	79.2	n/a	n/a	46.4	37.7	62.2
SWD (%)	18.1	13.8	39.6	n/a	n/a	16.0	20.6	11.7
EL (%)	6.3	21.0	n/r	n/a	n/a	3.9	5.6	9.5
Male (%)	52.5	46.9	49.1	n/a	n/a	52.2	53.4	49.9
Female (%)	47.5	53.1	50.9	n/a	n/a	47.8	46.6	50.1
AmInd/ANat (%)	3.0	98.9	n/a	n/a	n/a	0.2	0.1	11.6
Asian (%)	6.1	n/a	n/r	n/a	n/a	1.8	8.3	1.9
Black/AA (%)	19.4	n/a	47.2	n/a	n/a	59.8	17.3	2.7
Hisp/Lat (%)	33.6	n/a	47.2	n/a	n/a	12.1	29.9	58.7
Wh/Caus (%)	36.6	n/r	n/r	n/a	n/a	23.4	43.2	24.6
NtvHawaii/Pacific (%)	0.2	n/a	n/a	n/a	n/a	0.2	0.2	0.4
Two or More (%)	1.0	n/a	n/r	n/a	n/a	2.5	1.0	0.1
Unknown (%)	n/r	n/r	n/a	n/a	n/a	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.16 Demographic Information for Grade 3 Mathematics, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	49.0	98.5	79.0	n/r	53.9	40.5	40.8	77.9
SWD (%)	15.5	12.6	16.6	15.3	14.3	14.0	18.5	14.8
EL (%)	15.7	26.4	14.8	12.2	20.1	14.1	9.0	21.9
Male (%)	51.0	50.6	51.2	51.7	50.9	50.8	51.2	50.6
Female (%)	49.0	49.4	48.8	48.3	49.1	49.2	48.8	49.4
AmInd/ANat (%)	1.4	98.3	n/r	0.5	0.3	0.2	0.1	11.5
Asian (%)	6.6	n/a	1.5	5.5	5.2	6.5	10.7	1.4
Black/AA (%)	19.7	n/r	66.5	11.5	17.3	33.3	15.5	3.6
Hisp/Lat (%)	28.1	n/a	16.9	21.0	27.0	18.8	29.7	59.5
Wh/Caus (%)	40.4	n/a	12.4	45.0	46.3	36.0	41.2	23.5
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.7	0.1	0.2	0.2	0.4
Two or More (%)	3.5	n/r	2.5	13.0	3.8	5.0	2.5	0.1
Unknown (%)	0.0	n/r	n/r	1.8	0.0	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.17 Demographic Information for Grade 4 Mathematics, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	48.8	98.1	81.1	n/r	53.9	40.5	40.1	77.8
SWD (%)	16.5	13.5	18.3	14.6	15.3	15.1	19.6	16.1
EL (%)	14.3	32.3	14.4	10.1	19.4	11.9	7.0	21.2
Male (%)	50.9	48.0	50.2	49.8	50.8	51.2	50.9	51.1
Female (%)	49.1	52.0	49.8	50.2	49.2	48.8	49.1	48.9
AmInd/ANat (%)	1.5	98.2	n/r	0.5	0.4	0.3	0.1	11.3
Asian (%)	6.7	n/a	1.4	5.8	5.2	6.6	10.7	1.4
Black/AA (%)	19.6	n/r	68.1	12.3	16.8	34.0	15.4	3.5
Hisp/Lat (%)	27.9	n/a	17.8	21.7	27.2	18.0	29.1	60.2
Wh/Caus (%)	40.7	n/r	10.3	43.9	46.6	36.1	42.1	23.0
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.8	0.1	0.2	0.2	0.4
Two or More (%)	3.4	n/r	2.1	12.5	3.8	4.9	2.4	0.1
Unknown (%)	0.0	n/r	n/r	1.5	0.0	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.18 Demographic Information for Grade 5 Mathematics, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	48.3	98.3	79.8	n/a	53.5	39.8	39.5	77.7
SWD (%)	16.9	13.6	19.2	14.9	15.4	15.7	20.0	16.7
EL (%)	9.1	36.9	11.7	9.0	11.2	7.5	4.7	16.9
Male (%)	51.0	50.6	50.1	51.1	50.9	51.1	50.9	51.2
Female (%)	49.0	49.4	49.9	48.9	49.1	48.9	49.1	48.8
AmInd/ANat (%)	1.4	97.7	n/r	0.5	0.3	0.3	0.1	11.5
Asian (%)	6.5	n/r	1.2	5.9	5.0	6.5	10.3	1.5
Black/AA (%)	19.3	n/a	68.2	11.8	16.8	33.3	15.3	3.1
Hisp/Lat (%)	28.1	n/r	17.2	21.5	27.4	18.2	29.1	60.1
Wh/Caus (%)	41.2	n/r	10.9	43.6	46.8	36.9	42.7	23.2
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.8	0.1	0.2	0.2	0.4
Two or More (%)	3.3	n/r	2.3	13.2	3.6	4.6	2.2	0.1
Unknown (%)	0.0	n/r	n/r	1.6	n/r	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.19 Demographic Information for Grade 6 Mathematics, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	47.4	98.6	77.9	n/r	52.5	39.3	38.5	75.6
SWD (%)	16.8	14.5	20.5	14.4	15.5	15.7	19.5	16.3
EL (%)	6.6	44.5	7.4	7.6	7.4	5.6	3.9	13.7
Male (%)	50.9	47.2	51.1	49.8	50.9	51.2	50.9	50.6
Female (%)	49.1	52.8	48.9	50.2	49.1	48.8	49.1	49.4
AmInd/ANat (%)	1.4	98.6	n/r	0.5	0.3	0.3	0.1	11.6
Asian (%)	6.5	n/a	1.5	6.6	5.0	6.5	10.3	1.5
Black/AA (%)	19.0	n/a	69.9	11.8	16.6	33.6	15.0	3.1
Hisp/Lat (%)	27.6	n/r	17.3	20.8	27.1	17.6	28.3	59.6
Wh/Caus (%)	42.1	n/a	9.0	43.9	47.5	37.3	44.0	23.7
NtvHawaii/Pacific (%)	0.2	n/r	n/r	1.8	0.1	0.2	0.2	0.4
Two or More (%)	3.1	n/r	2.1	12.8	3.4	4.6	2.0	0.1
Unknown (%)	0.0	n/r	n/a	1.7	n/r	n/a	n/a	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.20 Demographic Information for Grade 7 Mathematics, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	47.8	98.3	78.4	n/a	51.1	41.2	38.0	74.1
SWD (%)	17.3	15.8	20.5	n/a	15.5	16.9	20.4	16.1
EL (%)	6.4	37.8	7.5	n/a	6.7	6.2	4.1	12.0
Male (%)	51.1	49.3	50.1	n/a	51.4	50.7	51.1	50.5
Female (%)	48.9	50.7	49.9	n/a	48.6	49.3	48.9	49.5
AmInd/ANat (%)	1.4	97.2	n/r	n/a	0.3	0.3	0.1	11.4
Asian (%)	5.8	n/r	1.4	n/a	5.1	4.7	9.0	1.5
Black/AA (%)	18.9	n/r	70.8	n/a	15.9	35.7	15.2	3.1
Hisp/Lat (%)	28.1	n/r	16.9	n/a	27.1	18.1	28.4	60.1
Wh/Caus (%)	42.8	n/a	9.0	n/a	48.3	36.9	45.3	23.6
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/a	0.1	0.2	0.2	0.3
Two or More (%)	2.7	n/r	1.7	n/a	3.3	4.1	1.8	n/r
Unknown (%)	0.0	n/r	n/a	n/a	n/r	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.21 Demographic Information for Grade 8 Mathematics, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	50.5	99.2	83.9	n/a	50.4	45.4	42.9	78.2
SWD (%)	19.4	14.9	23.7	n/a	15.6	21.4	26.5	18.4
EL (%)	6.9	38.5	7.2	n/a	6.1	7.4	5.4	15.4
Male (%)	51.8	50.0	50.0	n/a	51.4	52.2	52.6	51.5
Female (%)	48.2	50.0	50.0	n/a	48.6	47.8	47.4	48.5
AmInd/ANat (%)	1.5	98.0	n/r	n/a	0.2	0.3	0.1	12.2
Asian (%)	4.4	n/a	1.3	n/a	5.0	2.9	5.0	0.9
Black/AA (%)	20.0	n/a	77.0	n/a	16.0	42.1	18.2	3.1
Hisp/Lat (%)	28.8	n/a	15.3	n/a	26.6	16.9	31.7	62.8
Wh/Caus (%)	42.5	n/r	5.1	n/a	48.8	33.7	43.3	20.5
NtvHawaii/Pacific (%)	0.1	n/r	n/r	n/a	0.1	0.2	0.2	0.3
Two or More (%)	2.6	n/r	1.1	n/a	3.2	4.0	1.5	0.1
Unknown (%)	0.0	n/r	n/a	n/a	n/r	n/a	n/a	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.22 Demographic Information for Algebra I, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	39.2	98.2	83.9	n/a	32.0	34.8	35.7	68.5
SWD (%)	17.0	14.1	19.3	11.8	3.6	16.3	18.5	14.7
EL (%)	7.7	38.3	8.4	6.0	0.6	9.4	5.7	12.6
Male (%)	51.4	49.8	50.4	50.8	50.3	51.6	51.5	50.9
Female (%)	48.6	50.2	49.6	49.2	49.7	48.4	48.5	49.1
AmInd/ANat (%)	1.7	97.8	n/r	n/r	n/r	0.2	0.1	12.7
Asian (%)	7.9	n/a	1.7	7.5	11.7	6.8	10.1	1.7
Black/AA (%)	21.1	n/a	71.7	11.6	10.7	34.6	15.1	3.1
Hisp/Lat (%)	27.9	n/a	17.7	20.2	20.4	19.2	27.8	58.4
Wh/Caus (%)	38.8	n/a	7.2	43.4	53.0	35.0	45.1	23.6
NtvHawaii/Pacific (%)	0.2	n/a	n/r	1.8	n/r	0.1	0.2	0.3
Two or More (%)	2.4	n/r	1.5	12.8	3.8	4.0	1.5	0.1
Unknown (%)	0.0	n/r	n/a	2.3	n/a	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.23 Demographic Information for Geometry, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	38.6	99.6	86.8	n/a	n/r	19.8	33.3	65.9
SWD (%)	16.5	15.0	21.2	10.2	n/r	7.9	18.6	12.7
EL (%)	6.3	28.8	11.0	4.8	n/a	3.5	4.8	12.9
Male (%)	50.8	52.9	50.9	49.3	59.7	50.2	51.0	50.4
Female (%)	49.2	47.1	49.1	50.7	40.3	49.8	49.0	49.6
AmInd/ANat (%)	2.3	97.9	n/r	n/r	n/r	n/r	0.1	12.0
Asian (%)	8.6	n/a	1.2	8.6	21.8	14.3	9.9	1.9
Black/AA (%)	14.9	n/a	72.8	12.2	n/r	21.0	14.9	2.8
Hisp/Lat (%)	30.5	n/a	18.7	21.0	n/r	9.7	26.8	59.9
Wh/Caus (%)	41.5	n/r	6.1	41.9	65.7	49.6	46.7	23.1
NtvHawaii/Pacific (%)	0.3	n/a	n/r	2.5	n/a	n/r	0.2	0.4
Two or More (%)	1.7	n/r	1.1	12.0	6.0	5.0	1.4	0.1
Unknown (%)	0.0	n/r	n/r	1.2	n/a	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.24 Demographic Information for Algebra II, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	35.7	98.3	38.6	n/a	n/r	29.3	31.9	60.7
SWD (%)	13.0	12.4	10.7	7.5	n/a	10.0	15.1	9.6
EL (%)	4.5	19.5	11.6	3.6	n/a	2.4	3.9	9.0
Male (%)	49.9	46.6	48.4	51.8	52.3	49.3	50.2	49.1
Female (%)	50.1	53.4	51.6	48.2	47.7	50.7	49.8	50.9
AmInd/ANat (%)	2.4	99.4	n/a	n/r	n/a	0.2	0.1	11.8
Asian (%)	9.2	n/a	n/r	9.4	n/r	7.7	11.6	2.1
Black/AA (%)	17.0	n/a	29.6	11.2	n/r	39.3	14.3	2.7
Hisp/Lat (%)	27.9	n/a	21.2	16.8	n/r	9.7	26.2	57.7
Wh/Caus (%)	41.4	n/r	39.7	45.4	65.8	39.1	46.4	25.3
NtvHawaii/Pacific (%)	0.3	n/a	n/r	1.7	n/a	0.1	0.2	0.3
Two or More (%)	1.8	n/a	n/r	13.3	n/r	3.9	1.2	n/r
Unknown (%)	0.0	n/r	n/a	1.8	n/a	n/a	n/r	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.25 Demographic Information for Integrated Mathematics I, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	67.3	n/a	n/a	n/a	n/r	n/a	n/a	69.4
SWD (%)	21.3	n/a	n/a	n/a	n/a	n/a	n/a	21.8
EL (%)	14.7	n/a	n/a	n/a	n/a	n/a	n/a	15.4
Male (%)	54.9	n/a	n/a	n/a	n/r	n/a	n/a	55.9
Female (%)	45.1	n/a	n/a	n/a	n/r	n/a	n/a	44.1
AmInd/ANat (%)	5.3	n/a	n/a	n/a	n/a	n/a	n/a	5.6
Asian (%)	n/r	n/a	n/a	n/a	n/r	n/a	n/a	n/r
Black/AA (%)	4.7	n/a	n/a	n/a	n/a	n/a	n/a	4.9
Hisp/Lat (%)	63.9	n/a	n/a	n/a	n/r	n/a	n/a	66.8
Wh/Caus (%)	24.8	n/a	n/a	n/a	n/r	n/a	n/a	21.6
NtvHawaii/Pacific (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
Two or More (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
Unknown (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.26 Demographic Information for Integrated Mathematics II, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	61.2	n/r	42.1	n/a	n/a	n/a	n/a	65.3
SWD (%)	14.0	n/r	17.9	n/a	n/a	n/a	n/a	13.3
EL (%)	15.3	n/a	n/r	n/a	n/a	n/a	n/a	18.4
Male (%)	48.5	n/r	49.7	n/a	n/r	n/a	n/a	47.8
Female (%)	51.5	n/r	50.3	n/a	n/a	n/a	n/a	52.2
AmInd/ANat (%)	3.9	n/r	n/r	n/a	n/a	n/a	n/a	n/r
Asian (%)	n/r	n/a	n/r	n/a	n/a	n/a	n/a	n/r
Black/AA (%)	11.2	n/a	44.8	n/a	n/a	n/a	n/a	4.6
Hisp/Lat (%)	61.6	n/a	24.1	n/a	n/a	n/a	n/a	72.5
Wh/Caus (%)	21.0	n/a	23.4	n/a	n/r	n/a	n/a	19.2
NtvHawaii/Pacific (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
Two or More (%)	n/r	n/a	n/r	n/a	n/a	n/a	n/a	n/a
Unknown (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.5.27 Demographic Information for Integrated Mathematics III, Overall and by State

Demographic	PARCC	BIE	DC	DD	IL	MD	NJ	NM
Econ Dis (%)	96.9	n/a	n/a	n/a	n/a	n/a	n/a	96.9
SWD (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
EL (%)	13.1	n/a	n/a	n/a	n/a	n/a	n/a	13.1
Male (%)	50.8	n/a	n/a	n/a	n/a	n/a	n/a	50.8
Female (%)	49.2	n/a	n/a	n/a	n/a	n/a	n/a	49.2
AmInd/ANat (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
Asian (%)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Black/AA (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
Hisp/Lat (%)	80.1	n/a	n/a	n/a	n/a	n/a	n/a	80.1
Wh/Caus (%)	12.0	n/a	n/a	n/a	n/a	n/a	n/a	12.0
NtvHawaii/Pacific (%)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/r
Two or More (%)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Unknown (%)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note: PARCC = data from all participating states combined; BIE=Bureau of Indian Education, DD=Department of Defence Education Activity, DC=District of Columbia, IL=Illinois, MD=Maryland, NJ=New Jersey, and NM=New Mexico. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/a = not applicable; and n/r = not reported due to n<20.

Table A.7.1 Differential Item Functioning for ELA/L Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs AmerIndian	CBT	42	46	46					46	100				
	PBT	27	27	26			5	19	17	65	4	15		
White vs Asian	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs Black	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs Hispanic	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				
White vs Pacific Islander	CBT	42	46	46			1	2	44	96	1	2		
	PBT	27	27	0										
White vs Multiracial	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				
NoEcnDis vs EcnDis	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				
ELN vs ELY	CBT	42	46	46			1	2	45	98				
	PBT	27	27	26			1	4	25	96				
SWDN vs SWDY	CBT	42	46	46					46	100				
	PBT	27	27	26					26	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.2 Differential Item Functioning for ELA/L Grade 4

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	54	62	62			4	6	57	92	1	2		
	PBT	36	36	35	1	3	2	6	31	89	1	3		
White vs AmerIndian	CBT	54	62	62			2	3	60	97				
	PBT	36	36	35	1	3	1	3	29	83	3	9	1	3
White vs Asian	CBT	54	62	62					62	100				
	PBT	36	36	35					35	100				
White vs Black	CBT	54	62	62			1	2	61	98				
	PBT	36	36	35					34	97	1	3		
White vs Hispanic	CBT	54	62	62					62	100				
	PBT	36	36	35					35	100				
White vs Pacific Islander	CBT	54	62	62			3	5	59	95				
	PBT	36	36	0										
White vs Multiracial	CBT	54	62	62					62	100				
	PBT	36	36	35					35	100				
NoEcnDis vs EcnDis	CBT	54	62	62					62	100				
	PBT	36	36	35					35	100				
ELN vs ELY	CBT	54	62	62					62	100				
	PBT	36	36	35					35	100				
SWDN vs SWDY	CBT	54	62	62					62	100				
	PBT	36	36	35					35	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.3 Differential Item Functioning for ELA/L Grade 5

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	51	62	62	1	2	3	5	57	92	1	2		
	PBT	36	36	35	1	3	1	3	32	91	1	3		
White vs AmerIndian	CBT	51	62	62	1	2	1	2	59	95	1	2		
	PBT	36	36	35	2	6	6	17	22	63	1	3	4	11
White vs Asian	CBT	51	62	62					62	100				
	PBT	36	36	35			1	3	34	97				
White vs Black	CBT	51	62	62					62	100				
	PBT	36	36	35					35	100				
White vs Hispanic	CBT	51	62	62			2	3	60	97				
	PBT	36	36	35					35	100				
White vs Pacific Islander	CBT	51	62	62			5	8	54	87	2	3	1	2
	PBT	36	36	0										
White vs Multiracial	CBT	51	62	62					62	100				
	PBT	36	36	35					35	100				
NoEcnDis vs EcnDis	CBT	51	62	62					62	100				
	PBT	36	36	35					35	100				
ELN vs ELY	CBT	51	62	62	1	2	2	3	59	95				
	PBT	36	36	35			2	6	33	94				
SWDN vs SWDY	CBT	51	62	62					62	100				
	PBT	36	36	35					35	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.4 Differential Item Functioning for ELA/L Grade 6

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	55	62	62	2	3	6	10	53	85	1	2		
	PBT	36	36	35	1	3	1	3	29	83	4	11		
White vs AmerIndian	CBT	55	62	62			5	8	56	90	1	2		
	PBT	36	36	35	3	9	1	3	28	80			3	9
White vs Asian	CBT	55	62	62			2	3	60	97				
	PBT	36	36	35	1	3			34	97				
White vs Black	CBT	55	62	62	2	3			60	97				
	PBT	36	36	35	1	3			34	97				
White vs Hispanic	CBT	55	62	62	2	3			60	97				
	PBT	36	36	35	1	3			34	97				
White vs Pacific Islander	CBT	55	62	62			1	2	60	97	1	2		
	PBT	36	36	0										
White vs Multiracial	CBT	55	62	62					62	100				
	PBT	36	36	35					35	100				
NoEcnDis vs EcnDis	CBT	55	62	62			1	2	61	98				
	PBT	36	36	35					35	100				
ELN vs ELY	CBT	55	62	62	1	2	2	3	59	95				
	PBT	36	36	35			2	6	30	86			3	9
SWDN vs SWDY	CBT	55	62	62					62	100				
	PBT	36	36	35					35	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.5 Differential Item Functioning for ELA/L Grade 7

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	56	62	62			3	5	59	95				
	PBT	36	36	35	1	3	7	20	21	60	6	17		
White vs AmerIndian	CBT	56	62	62			2	3	60	97				
	PBT	36	36	35	5	14	2	6	22	63			6	17
White vs Asian	CBT	56	62	62					62	100				
	PBT	36	36	35			2	6	31	89	2	6		
White vs Black	CBT	56	62	62					62	100				
	PBT	36	36	35			2	6	33	94				
White vs Hispanic	CBT	56	62	62					62	100				
	PBT	36	36	35			1	3	34	97				
White vs Pacific Islander	CBT	56	62	62	1	2	4	6	57	92				
	PBT	36	36	0										
White vs Multiracial	CBT	56	62	62					62	100				
	PBT	36	36	35					34	97	1	3		
NoEcnDis vs EcnDis	CBT	56	62	62					62	100				
	PBT	36	36	35					35	100				
ELN vs ELY	CBT	56	62	62	1	2	1	2	60	97				
	PBT	36	36	35	2	6	1	3	26	74	6	17		
SWDN vs SWDY	CBT	56	62	62					62	100				
	PBT	36	36	35					35	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.6 Differential Item Functioning for ELA/L Grade 8

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	53	62	62	2	3	4	6	54	87	2	3		
	PBT	36	36	35	1	3	4	11	24	69	6	17		
White vs AmerIndian	CBT	53	62	62	2	3	5	8	53	85	2	3		
	PBT	36	36	35	2	6	7	20	20	57	5	14	1	3
White vs Asian	CBT	53	62	62					62	100				
	PBT	36	36	35					34	97	1	3		
White vs Black	CBT	53	62	62			1	2	61	98				
	PBT	36	36	35					35	100				
White vs Hispanic	CBT	53	62	62					62	100				
	PBT	36	36	35					35	100				
White vs Pacific Islander	CBT	53	62	62			5	8	57	92				
	PBT	36	36	0										
White vs Multiracial	CBT	53	62	62					62	100				
	PBT	36	36	35					35	100				
NoEcnDis vs EcnDis	CBT	53	62	62					62	100				
	PBT	36	36	35					35	100				
ELN vs ELY	CBT	53	62	62			1	2	61	98				
	PBT	36	36	35			7	20	23	66	5	14		
SWDN vs SWDY	CBT	53	62	62					62	100				
	PBT	36	36	35					35	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.7 Differential Item Functioning for ELA/L Grade 9

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	58	62	62	1	2	6	10	53	85	2	3		
	PBT	36	36	35	4	11			26	74	3	9	2	6
White vs AmerIndian	CBT	58	62	62	4	6	3	5	54	87	1	2		
	PBT	36	36	0										
White vs Asian	CBT	58	62	62					62	100				
	PBT	36	36	0										
White vs Black	CBT	58	62	62			1	2	61	98				
	PBT	36	36	0										
White vs Hispanic	CBT	58	62	62					62	100				
	PBT	36	36	0										
White vs Pacific Islander	CBT	58	62	31			3	10	28	90				
	PBT	36	36	0										
White vs Multiracial	CBT	58	62	62					62	100				
	PBT	36	36	0										
NoEcnDis vs EcnDis	CBT	58	62	62					62	100				
	PBT	36	36	35	1	3	1	3	32	91	1	3		
ELN vs ELY	CBT	58	62	62			2	3	59	95	1	2		
	PBT	36	36	35	2	6			33	94				
SWDN vs SWDY	CBT	58	62	62					62	100				
	PBT	36	36	35					33	94	1	3	1	3

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.8 Differential Item Functioning for ELA/L Grade 10

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	55	62	62	1	2	6	10	54	87	1	2		
	PBT	36	36	35	1	3	1	3	31	89	2	6		
White vs AmerIndian	CBT	55	62	62	3	5	2	3	55	89	2	3		
	PBT	36	36	35	5	14	4	11	20	57	3	9	3	9
White vs Asian	CBT	55	62	62					62	100				
	PBT	36	36	0										
White vs Black	CBT	55	62	62					62	100				
	PBT	36	36	35	1	3			34	97				
White vs Hispanic	CBT	55	62	62			2	3	60	97				
	PBT	36	36	35			2	6	33	94				
White vs Pacific Islander	CBT	55	62	62	1	2			61	98				
	PBT	36	36	0										
White vs Multiracial	CBT	55	62	62					62	100				
	PBT	36	36	0										
NoEcnDis vs EcnDis	CBT	55	62	62					62	100				
	PBT	36	36	35			1	3	32	91	2	6		
ELN vs ELY	CBT	55	62	62	1	2	1	2	59	95	1	2		
	PBT	36	36	35	2	6	2	6	29	83			2	6
SWDN vs SWDY	CBT	55	62	62					62	100				
	PBT	36	36	35					35	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.9 Differential Item Functioning for ELA/L Grade 11

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	53	62	62	2	3	3	5	56	90	1	2		
	PBT	36	36	35	2	6	1	3	30	86	1	3	1	3
White vs AmerIndian	CBT	53	62	62	10	16	6	10	43	69	1	2	2	3
	PBT	36	36	35	2	6	3	9	27	77	1	3	2	6
White vs Asian	CBT	53	62	62					62	100				
	PBT	36	36	0										
White vs Black	CBT	53	62	62	1	2	4	6	57	92				
	PBT	36	36	0										
White vs Hispanic	CBT	53	62	62	2	3	4	6	56	90				
	PBT	36	36	0										
White vs Pacific Islander	CBT	53	62	0										
	PBT	36	36	0										
White vs Multiracial	CBT	53	62	62					62	100				
	PBT	36	36	0										
NoEcnDis vs EcnDis	CBT	53	62	62			2	3	60	97				
	PBT	36	36	35	1	3	4	11	28	80	2	6		
ELN vs ELY	CBT	53	62	62	4	6	1	2	55	89	2	3		
	PBT	36	36	35	1	3	1	3	31	89	1	3	1	3
SWDN vs SWDY	CBT	53	62	62					62	100				
	PBT	36	36	35			3	9	29	83	3	9		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.10 Differential Item Functioning for Mathematics Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	86	86	86	1	1	5	6	79	92	1	1		
	PBT	42	42	42	1	2			41	98				
White vs AmerIndian	CBT	86	86	86			3	3	83	97				
	PBT	42	42	42	1	2	3	7	38	90				
White vs Asian	CBT	86	86	86					83	97	3	3		
	PBT	42	42	42					39	93	3	7		
White vs Black	CBT	86	86	86			3	3	78	91	4	5	1	1
	PBT	42	42	42					41	98	1	2		
White vs Hispanic	CBT	86	86	86			1	1	85	99				
	PBT	42	42	42			1	2	41	98				
White vs Pacific Islander	CBT	86	86	86			2	2	82	95	2	2		
	PBT	42	42	0										
White vs Multiracial	CBT	86	86	86					86	100				
	PBT	42	42	42					42	100				
NoEcnDis vs EcnDis	CBT	86	86	86					86	100				
	PBT	42	42	42			1	2	41	98				
ELN vs ELY	CBT	86	86	86			1	1	85	99				
	PBT	42	42	42					42	100				
SWDN vs SWDY	CBT	86	86	86			1	1	85	99				
	PBT	42	42	42					42	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.11 Differential Item Functioning for Mathematics Grade 4

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	80	80	80			3	4	76	95	1	1		
	PBT	40	40	40			2	5	38	95				
White vs AmerIndian	CBT	80	80	80			3	4	76	95	1	1		
	PBT	40	40	40	1	3	2	5	35	88	2	5		
White vs Asian	CBT	80	80	80					80	100				
	PBT	40	40	40			1	3	39	98				
White vs Black	CBT	80	80	80			1	1	79	99				
	PBT	40	40	40					40	100				
White vs Hispanic	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
White vs Pacific Islander	CBT	80	80	80			2	3	77	96	1	1		
	PBT	40	40	0										
White vs Multiracial	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
NoEcnDis vs EcnDis	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
ELN vs ELY	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
SWDN vs SWDY	CBT	80	80	80					80	100				
	PBT	40	40	40			1	3	39	98				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.12 Differential Item Functioning for Mathematics Grade 5

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	80	80	80			3	4	77	96				
	PBT	40	40	40			3	8	37	93				
White vs AmerIndian	CBT	80	80	80			1	1	79	99				
	PBT	40	40	40	1	3	2	5	36	90	1	3		
White vs Asian	CBT	80	80	80					79	99	1	1		
	PBT	40	40	40					39	98	1	3		
White vs Black	CBT	80	80	80					80	100				
	PBT	40	40	40	1	3	1	3	38	95				
White vs Hispanic	CBT	80	80	80					80	100				
	PBT	40	40	40			1	3	39	98				
White vs Pacific Islander	CBT	80	80	80			1	1	79	99				
	PBT	40	40	0										
White vs Multiracial	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
NoEcnDis vs EcnDis	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
ELN vs ELY	CBT	80	80	80					80	100				
	PBT	40	40	40					40	100				
SWDN vs SWDY	CBT	80	80	80					78	98	1	1	1	1
	PBT	40	40	40					40	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.13 Differential Item Functioning for Mathematics Grade 6

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	76	76	76	3	4	2	3	70	92	1	1		
	PBT	38	38	38			2	5	35	92	1	3		
White vs AmerIndian	CBT	76	76	76	1	1	2	3	73	96				
	PBT	38	38	38	1	3	3	8	32	84	2	5		
White vs Asian	CBT	76	76	76					76	100				
	PBT	38	38	38	1	3	1	3	33	87	3	8		
White vs Black	CBT	76	76	76			2	3	74	97				
	PBT	38	38	38			1	3	37	97				
White vs Hispanic	CBT	76	76	76					76	100				
	PBT	38	38	38					38	100				
White vs Pacific Islander	CBT	76	76	76			4	5	71	93	1	1		
	PBT	38	38	0										
White vs Multiracial	CBT	76	76	76					76	100				
	PBT	38	38	38					38	100				
NoEcnDis vs EcnDis	CBT	76	76	76					76	100				
	PBT	38	38	38					38	100				
ELN vs ELY	CBT	76	76	76	1	1	1	1	74	97				
	PBT	38	38	38			3	8	33	87	2	5		
SWDN vs SWDY	CBT	76	76	76					73	96	2	3	1	1
	PBT	38	38	38					36	95	2	5		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.14 Differential Item Functioning for Mathematics Grade 7

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	76	76	76	1	1	3	4	71	93	1	1		
	PBT	38	38	38					38	100				
White vs AmerIndian	CBT	76	76	76	1	1	1	1	74	97				
	PBT	38	38	38			3	8	30	79	3	8	2	5
White vs Asian	CBT	76	76	76			3	4	70	92	3	4		
	PBT	38	38	38			1	3	34	89	3	8		
White vs Black	CBT	76	76	76			1	1	74	97	1	1		
	PBT	38	38	38					36	95	1	3	1	3
White vs Hispanic	CBT	76	76	76			1	1	75	99				
	PBT	38	38	38			1	3	37	97				
White vs Pacific Islander	CBT	76	76	76			4	5	70	92	2	3		
	PBT	38	38	0										
White vs Multiracial	CBT	76	76	76					76	100				
	PBT	38	38	38					38	100				
NoEcnDis vs EcnDis	CBT	76	76	76					76	100				
	PBT	38	38	38					38	100				
ELN vs ELY	CBT	76	76	76	1	1	2	3	73	96				
	PBT	38	38	38			1	3	35	92	2	5		
SWDN vs SWDY	CBT	76	76	76					75	99	1	1		
	PBT	38	38	38					37	97	1	3		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.15 Differential Item Functioning for Mathematics Grade 8

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	72	72	72			1	1	71	99				
	PBT	36	36	36			1	3	35	97				
White vs AmerIndian	CBT	72	72	72	1	1	1	1	70	97				
	PBT	36	36	36			6	17	27	75	3	8		
White vs Asian	CBT	72	72	72	1	1			71	99				
	PBT	36	36	36			3	8	32	89	1	3		
White vs Black	CBT	72	72	72			1	1	71	99				
	PBT	36	36	36	1	3	1	3	32	89			2	6
White vs Hispanic	CBT	72	72	72	1	1			71	99				
	PBT	36	36	36					36	100				
White vs Pacific Islander	CBT	72	72	36			4	11	32	89				
	PBT	36	36	0										
White vs Multiracial	CBT	72	72	72					72	100				
	PBT	36	36	36					36	100				
NoEcnDis vs EcnDis	CBT	72	72	72					72	100				
	PBT	36	36	36			1	3	35	97				
ELN vs ELY	CBT	72	72	72	1	1	1	1	70	97				
	PBT	36	36	36			3	8	33	92				
SWDN vs SWDY	CBT	72	72	72					71	99	1	1		
	PBT	36	36	36					36	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.16 Differential Item Functioning for Mathematics Algebra I

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	82	84	84	1	1	1	1	82	98				
	PBT	42	42	42	1	2	1	2	38	90	2	5		
White vs AmerIndian	CBT	82	84	84			1	1	83	99				
	PBT	42	42	42					40	95	2	5		
White vs Asian	CBT	82	84	84			2	2	79	94	2	2	1	1
	PBT	42	42	0										
White vs Black	CBT	82	84	84					84	100				
	PBT	42	42	0										
White vs Hispanic	CBT	82	84	84					84	100				
	PBT	42	42	0										
White vs Pacific Islander	CBT	82	84	84			1	1	81	96	2	2		
	PBT	42	42	0										
White vs Multiracial	CBT	82	84	84					84	100				
	PBT	42	42	0										
NoEcnDis vs EcnDis	CBT	82	84	84					84	100				
	PBT	42	42	42					41	98	1	2		
ELN vs ELY	CBT	82	84	84			3	4	76	90	3	4	2	2
	PBT	42	42	42			1	2	40	95	1	2		
SWDN vs SWDY	CBT	82	84	84					84	100				
	PBT	42	42	42	1	2	2	5	39	93				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.17 Differential Item Functioning for Mathematics Geometry

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	84	86	86			2	2	84	98				
	PBT	43	43	43					43	100				
White vs AmerIndian	CBT	84	86	86			3	3	81	94	2	2		
	PBT	43	43	0										
White vs Asian	CBT	84	86	86					85	99	1	1		
	PBT	43	43	0										
White vs Black	CBT	84	86	86					86	100				
	PBT	43	43	0										
White vs Hispanic	CBT	84	86	86					86	100				
	PBT	43	43	0										
White vs Pacific Islander	CBT	84	86	86			1	1	85	99				
	PBT	43	43	0										
White vs Multiracial	CBT	84	86	86					86	100				
	PBT	43	43	0										
NoEcnDis vs EcnDis	CBT	84	86	86			1	1	85	99				
	PBT	43	43	43			1	2	42	98				
ELN vs ELY	CBT	84	86	86	2	2	3	3	78	91	3	3		
	PBT	43	43	43					43	100				
SWDN vs SWDY	CBT	84	86	86					86	100				
	PBT	43	43	43			1	2	37	86	5	12		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Table A.7.18 Differential Item Functioning for Mathematics Algebra II

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	81	82	82			1	1	81	99				
	PBT	41	41	41			1	2	39	95	1	2		
White vs AmerIndian	CBT	81	82	82			3	4	78	95	1	1		
	PBT	41	41	41			3	7	34	83	1	2	3	7
White vs Asian	CBT	81	82	82			1	1	79	96	2	2		
	PBT	41	41	0										
White vs Black	CBT	81	82	82					82	100				
	PBT	41	41	0										
White vs Hispanic	CBT	81	82	82					82	100				
	PBT	41	41	0										
White vs Pacific Islander	CBT	81	82	82					82	100				
	PBT	41	41	0										
White vs Multiracial	CBT	81	82	82					82	100				
	PBT	41	41	0										
NoEcnDis vs EcnDis	CBT	81	82	82					82	100				
	PBT	41	41	41					39	95	2	5		
ELN vs ELY	CBT	81	82	82					79	96	3	4		
	PBT	41	41	0										
SWDN vs SWDY	CBT	81	82	82					82	100				
	PBT	41	41	41					39	95	2	5		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability, SWDY = student with disability.

Appendix 8: Reliability of Classification by Content and Grade Level

Table A.8.1 Reliability of Classification: Grade 3 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 - 699	0.16	0.03	0.00	0.00	0.00	0.19
		700 - 724	0.04	0.10	0.05	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.13	0.04	0.00	0.22
		750 - 809	0.00	0.00	0.05	0.30	0.03	0.38
		810 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 – 699	0.15	0.04	0.01	0.00	0.00	0.20
		700 - 724	0.04	0.08	0.06	0.01	0.00	0.18
		725 - 749	0.01	0.04	0.10	0.05	0.00	0.20
		750 - 809	0.00	0.01	0.06	0.27	0.02	0.37
		810 - 850	0.00	0.00	0.00	0.02	0.02	0.05
PBT	Decision Accuracy	650 – 699	0.15	0.03	0.00	0.00	0.00	0.18
		700 - 724	0.04	0.09	0.05	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.13	0.05	0.00	0.22
		750 - 809	0.00	0.00	0.05	0.31	0.03	0.40
		810 - 850	0.00	0.00	0.00	0.01	0.01	0.02
	Decision Consistency	650 – 699	0.14	0.04	0.01	0.00	0.00	0.19
		700 - 724	0.04	0.07	0.06	0.01	0.00	0.18
		725 - 749	0.01	0.04	0.09	0.06	0.00	0.21
		750 - 809	0.00	0.01	0.07	0.27	0.03	0.38
		810 - 850	0.00	0.00	0.00	0.02	0.02	0.04

Table A.8.2 Reliability of Classification: Grade 4 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.12	0.04	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.16	0.05	0.00	0.25
		750 - 789	0.00	0.00	0.04	0.27	0.03	0.34
		790 - 850	0.00	0.00	0.00	0.02	0.09	0.11
	Decision Consistency	650 – 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.10	0.05	0.00	0.00	0.19
		725 - 749	0.00	0.05	0.13	0.06	0.00	0.24
		750 - 789	0.00	0.00	0.06	0.23	0.03	0.33
		790 - 850	0.00	0.00	0.00	0.04	0.09	0.12
PBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.13	0.05	0.00	0.00	0.21
		725 - 749	0.00	0.05	0.19	0.05	0.00	0.28
		750 - 789	0.00	0.00	0.05	0.23	0.03	0.32
		790 - 850	0.00	0.00	0.00	0.01	0.05	0.06
	Decision Consistency	650 - 699	0.10	0.04	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.11	0.07	0.01	0.00	0.21
		725 - 749	0.00	0.05	0.14	0.06	0.00	0.26
		750 - 789	0.00	0.01	0.07	0.20	0.03	0.31
		790 - 850	0.00	0.00	0.00	0.03	0.05	0.08

Table A.8.3 Reliability of Classification: Grade 5 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.09	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.13	0.04	0.00	0.00	0.20
		725 - 749	0.00	0.04	0.18	0.04	0.00	0.26
		750 - 798	0.00	0.00	0.04	0.32	0.02	0.38
		799 - 850	0.00	0.00	0.00	0.01	0.04	0.05
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.11	0.05	0.00	0.00	0.19
		725 - 749	0.00	0.05	0.15	0.06	0.00	0.25
		750 - 798	0.00	0.00	0.06	0.29	0.02	0.38
		799 - 850	0.00	0.00	0.00	0.02	0.03	0.06
PBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.11	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.19	0.05	0.00	0.28
		750 - 798	0.00	0.00	0.06	0.34	0.03	0.43
		799 - 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.09	0.06	0.00	0.00	0.18
		725 - 749	0.00	0.05	0.15	0.06	0.00	0.26
		750 - 798	0.00	0.00	0.07	0.31	0.02	0.41
		799 - 850	0.00	0.00	0.00	0.02	0.01	0.03

Table A.8.4 Reliability of Classification: Grade 6 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.08	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.15	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.20	0.04	0.00	0.29
		750 - 789	0.00	0.00	0.04	0.27	0.02	0.33
		790 - 850	0.00	0.00	0.00	0.01	0.06	0.07
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.13	0.05	0.00	0.00	0.21
		725 - 749	0.00	0.05	0.17	0.06	0.00	0.27
		750 - 789	0.00	0.00	0.06	0.25	0.02	0.33
		790 - 850	0.00	0.00	0.00	0.02	0.06	0.08
PBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.15	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.20	0.05	0.00	0.29
		750 - 789	0.00	0.00	0.04	0.28	0.02	0.34
		790 - 850	0.00	0.00	0.00	0.01	0.04	0.05
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.12	0.05	0.00	0.00	0.21
		725 - 749	0.00	0.05	0.17	0.06	0.00	0.28
		750 - 789	0.00	0.00	0.06	0.25	0.02	0.33
		790 - 850	0.00	0.00	0.00	0.02	0.03	0.06

Table A.8.5 Reliability of Classification: Grade 7 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.12	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.10	0.04	0.00	0.00	0.17
		725 - 749	0.00	0.03	0.15	0.04	0.00	0.23
		750 - 784	0.00	0.00	0.04	0.23	0.03	0.31
		785 - 850	0.00	0.00	0.00	0.03	0.14	0.16
	Decision Consistency	650 - 699	0.11	0.03	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.08	0.05	0.00	0.00	0.17
		725 - 749	0.00	0.04	0.12	0.06	0.00	0.22
		750 - 784	0.00	0.00	0.06	0.20	0.04	0.30
		785 - 850	0.00	0.00	0.00	0.05	0.13	0.18
PBT	Decision Accuracy	650 – 699	0.14	0.02	0.00	0.00	0.00	0.16
		700 - 724	0.03	0.13	0.04	0.00	0.00	0.20
		725 - 749	0.00	0.04	0.17	0.05	0.00	0.26
		750 - 784	0.00	0.00	0.05	0.23	0.03	0.30
		785 - 850	0.00	0.00	0.00	0.02	0.06	0.08
	Decision Consistency	650 - 699	0.13	0.03	0.00	0.00	0.00	0.17
		700 - 724	0.04	0.10	0.06	0.00	0.00	0.20
		725 - 749	0.00	0.05	0.13	0.06	0.00	0.25
		750 - 784	0.00	0.00	0.06	0.19	0.03	0.29
		785 - 850	0.00	0.00	0.00	0.04	0.06	0.09

Table A.8.6 Reliability of Classification: Grade 8 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.13	0.02	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.11	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.15	0.04	0.00	0.23
		750 - 793	0.00	0.00	0.04	0.27	0.02	0.34
		794 - 850	0.00	0.00	0.00	0.02	0.08	0.10
	Decision Consistency	650 - 699	0.12	0.03	0.00	0.00	0.00	0.15
		700 - 724	0.03	0.09	0.05	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.12	0.05	0.00	0.22
		750 - 793	0.00	0.00	0.06	0.24	0.03	0.33
		794 - 850	0.00	0.00	0.00	0.03	0.08	0.11
PBT	Decision Accuracy	650 - 699	0.14	0.02	0.00	0.00	0.00	0.16
		700 - 724	0.03	0.13	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.17	0.04	0.00	0.26
		750 - 793	0.00	0.00	0.05	0.25	0.02	0.31
		794 - 850	0.00	0.00	0.00	0.01	0.05	0.06
	Decision Consistency	650 - 699	0.13	0.04	0.00	0.00	0.00	0.17
		700 - 724	0.04	0.11	0.06	0.00	0.00	0.21
		725 - 749	0.00	0.05	0.14	0.05	0.00	0.24
		750 - 793	0.00	0.00	0.06	0.22	0.02	0.31
		794 - 850	0.00	0.00	0.00	0.03	0.05	0.07

Table A.8.7 Reliability of Classification: Grade 9 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.11	0.01	0.00	0.00	0.00	0.13
		700 - 724	0.02	0.10	0.03	0.00	0.00	0.16
		725 - 749	0.00	0.03	0.16	0.04	0.00	0.22
		750 - 790	0.00	0.00	0.04	0.29	0.03	0.36
		791 - 850	0.00	0.00	0.00	0.02	0.11	0.13
	Decision Consistency	650 - 699	0.11	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.09	0.05	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.13	0.05	0.00	0.22
		750 - 790	0.00	0.00	0.05	0.26	0.03	0.35
		791 - 850	0.00	0.00	0.00	0.04	0.10	0.14
PBT	Decision Accuracy	650 – 699	0.23	0.04	0.00	0.00	0.00	0.27
		700 - 724	0.04	0.17	0.05	0.00	0.00	0.26
		725 - 749	0.00	0.04	0.15	0.04	0.00	0.23
		750 - 790	0.00	0.00	0.03	0.15	0.01	0.20
		791 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.22	0.05	0.00	0.00	0.00	0.28
		700 - 724	0.05	0.14	0.06	0.00	0.00	0.25
		725 - 749	0.00	0.06	0.12	0.04	0.00	0.22
		750 - 790	0.00	0.00	0.05	0.13	0.01	0.20
		791 - 850	0.00	0.00	0.00	0.02	0.03	0.05

Table A.8.8 Reliability of Classification: Grade 10 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.16	0.02	0.00	0.00	0.00	0.19
		700 - 724	0.03	0.08	0.04	0.00	0.00	0.15
		725 - 749	0.00	0.03	0.11	0.04	0.00	0.19
		750 - 793	0.00	0.00	0.04	0.24	0.03	0.31
		794 - 850	0.00	0.00	0.00	0.03	0.13	0.16
	Decision Consistency	650 - 699	0.16	0.03	0.01	0.00	0.00	0.20
		700 - 724	0.03	0.06	0.05	0.01	0.00	0.15
		725 - 749	0.00	0.04	0.09	0.05	0.00	0.18
		750 - 793	0.00	0.01	0.06	0.20	0.04	0.30
		794 - 850	0.00	0.00	0.00	0.05	0.13	0.18
PBT	Decision Accuracy	650 – 699	0.20	0.03	0.00	0.00	0.00	0.23
		700 - 724	0.03	0.09	0.04	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.10	0.04	0.00	0.18
		750 - 793	0.00	0.00	0.04	0.21	0.03	0.28
		794 - 850	0.00	0.00	0.00	0.03	0.12	0.15
	Decision Consistency	650 - 699	0.19	0.04	0.01	0.00	0.00	0.24
		700 - 724	0.03	0.07	0.04	0.01	0.00	0.16
		725 - 749	0.00	0.04	0.08	0.05	0.00	0.17
		750 – 793	0.00	0.01	0.05	0.18	0.04	0.27
		794 - 850	0.00	0.00	0.00	0.05	0.11	0.16

Table A.8.9 Reliability of Classification: Grade 11 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.17	0.02	0.00	0.00	0.00	0.20
		700 – 724	0.03	0.11	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.14	0.04	0.00	0.22
		750 - 791	0.00	0.00	0.04	0.25	0.03	0.32
		792 - 850	0.00	0.00	0.00	0.02	0.06	0.08
	Decision Consistency	650 - 699	0.17	0.03	0.00	0.00	0.00	0.20
		700 - 724	0.04	0.08	0.05	0.01	0.00	0.18
		725 - 749	0.00	0.04	0.11	0.05	0.00	0.21
		750 - 791	0.00	0.00	0.06	0.22	0.03	0.31
		792 - 850	0.00	0.00	0.00	0.04	0.05	0.09
PBT	Decision Accuracy	650 – 699	0.19	0.03	0.00	0.00	0.00	0.22
		700 - 724	0.04	0.12	0.05	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.15	0.04	0.00	0.24
		750 - 791	0.00	0.00	0.05	0.18	0.03	0.26
		792 - 850	0.00	0.00	0.00	0.01	0.06	0.08
	Decision Consistency	650 – 699	0.18	0.04	0.01	0.00	0.00	0.23
		700 - 724	0.05	0.09	0.06	0.01	0.00	0.21
		725 - 749	0.01	0.05	0.11	0.05	0.00	0.22
		750 - 791	0.00	0.01	0.06	0.15	0.03	0.26
		792 - 850	0.00	0.00	0.00	0.03	0.06	0.09

Table A.8.10 Reliability of Classification: Grade 3 Mathematics

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.02	0.14	0.03	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.17	0.04	0.00	0.25
		750 - 789	0.00	0.00	0.04	0.27	0.02	0.33
		790 - 850	0.00	0.00	0.00	0.02	0.09	0.11
	Decision Consistency	650 – 699	0.10	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.11	0.05	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.14	0.05	0.00	0.24
		750 - 789	0.00	0.00	0.05	0.24	0.03	0.33
		790 - 850	0.00	0.00	0.00	0.03	0.09	0.12
PBT	Decision Accuracy	650 – 699	0.14	0.02	0.00	0.00	0.00	0.16
		700 - 724	0.03	0.19	0.04	0.00	0.00	0.26
		725 - 749	0.00	0.04	0.19	0.04	0.00	0.27
		750 - 789	0.00	0.00	0.04	0.21	0.02	0.26
		790 - 850	0.00	0.00	0.00	0.01	0.04	0.05
	Decision Consistency	650 - 699	0.13	0.04	0.00	0.00	0.00	0.17
		700 - 724	0.03	0.16	0.06	0.00	0.00	0.25
		725 - 749	0.00	0.06	0.15	0.05	0.00	0.26
		750 - 789	0.00	0.00	0.05	0.19	0.02	0.26
		790 - 850	0.00	0.00	0.00	0.02	0.04	0.06

Table A.8.11 Reliability of Classification: Grade 4 Mathematics

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.16	0.04	0.00	0.00	0.22
		725 - 749	0.00	0.04	0.20	0.04	0.00	0.28
		750 - 795	0.00	0.00	0.04	0.29	0.02	0.34
		796 - 850	0.00	0.00	0.00	0.01	0.04	0.04
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.13	0.05	0.00	0.00	0.22
		725 - 749	0.00	0.05	0.16	0.06	0.00	0.27
		750 - 795	0.00	0.00	0.05	0.27	0.02	0.34
		796 - 850	0.00	0.00	0.00	0.02	0.03	0.05
PBT	Decision Accuracy	650 – 699	0.16	0.03	0.00	0.00	0.00	0.19
		700 - 724	0.03	0.23	0.04	0.00	0.00	0.30
		725 - 749	0.00	0.05	0.17	0.04	0.00	0.25
		750 - 795	0.00	0.00	0.03	0.19	0.01	0.23
		796 - 850	0.00	0.00	0.00	0.00	0.02	0.02
	Decision Consistency	650 - 699	0.15	0.05	0.00	0.00	0.00	0.20
		700 - 724	0.04	0.19	0.05	0.00	0.00	0.29
		725 - 749	0.00	0.06	0.14	0.04	0.00	0.25
		750 - 795	0.00	0.00	0.04	0.18	0.01	0.23
		796 - 850	0.00	0.00	0.00	0.01	0.02	0.03

Table A.8.12 Reliability of Classification: Grade 5 Mathematics

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.11	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.16	0.04	0.00	0.00	0.22
		725 - 749	0.00	0.04	0.20	0.04	0.00	0.28
		750 - 789	0.00	0.00	0.04	0.26	0.02	0.31
		790 - 850	0.00	0.00	0.00	0.01	0.05	0.06
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.13	0.05	0.00	0.00	0.22
		725 - 749	0.00	0.05	0.16	0.06	0.00	0.27
		750 - 789	0.00	0.00	0.05	0.23	0.02	0.31
		790 - 850	0.00	0.00	0.00	0.02	0.05	0.07
PBT	Decision Accuracy	650 – 699	0.14	0.03	0.00	0.00	0.00	0.17
		700 - 724	0.04	0.20	0.04	0.00	0.00	0.28
		725 - 749	0.00	0.05	0.19	0.04	0.00	0.27
		750 - 789	0.00	0.00	0.03	0.19	0.01	0.24
		790 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.13	0.05	0.00	0.00	0.00	0.18
		700 - 724	0.04	0.17	0.06	0.00	0.00	0.27
		725 - 749	0.00	0.06	0.15	0.05	0.00	0.26
		750 - 789	0.00	0.00	0.05	0.17	0.01	0.24
		790 - 850	0.00	0.00	0.00	0.02	0.03	0.05

Table A.8.13 Reliability of Classification: Grade 6 Mathematics

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.11	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.19	0.04	0.00	0.00	0.26
		725 - 749	0.00	0.04	0.21	0.04	0.00	0.29
		750 - 787	0.00	0.00	0.03	0.23	0.01	0.27
		788 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 – 699	0.10	0.04	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.17	0.05	0.00	0.00	0.25
		725 - 749	0.00	0.06	0.17	0.05	0.00	0.28
		750 - 787	0.00	0.00	0.05	0.20	0.02	0.27
		788 - 850	0.00	0.00	0.00	0.02	0.03	0.05
PBT	Decision Accuracy	650 – 699	0.14	0.03	0.00	0.00	0.00	0.17
		700 - 724	0.03	0.20	0.04	0.00	0.00	0.28
		725 - 749	0.00	0.05	0.18	0.04	0.00	0.27
		750 - 787	0.00	0.00	0.03	0.19	0.02	0.24
		788 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.14	0.05	0.00	0.00	0.00	0.18
		700 - 724	0.04	0.17	0.06	0.00	0.00	0.27
		725 - 749	0.00	0.06	0.15	0.05	0.00	0.27
		750 - 787	0.00	0.00	0.05	0.17	0.02	0.24
		788 - 850	0.00	0.00	0.00	0.02	0.03	0.04

Table A.8.14 Reliability of Classification: Grade 7 Mathematics

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.20	0.03	0.00	0.00	0.26
		725 - 749	0.00	0.04	0.24	0.04	0.00	0.31
		750 - 785	0.00	0.00	0.03	0.24	0.01	0.29
		786 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.02	0.18	0.05	0.00	0.00	0.25
		725 - 749	0.00	0.05	0.20	0.05	0.00	0.31
		750 - 785	0.00	0.00	0.05	0.22	0.01	0.28
		786 - 850	0.00	0.00	0.00	0.02	0.03	0.05
PBT	Decision Accuracy	650 – 699	0.11	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.02	0.22	0.04	0.00	0.00	0.29
		725 - 749	0.00	0.05	0.22	0.04	0.00	0.31
		750 - 785	0.00	0.00	0.04	0.20	0.01	0.25
		786 - 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 - 699	0.10	0.04	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.19	0.06	0.00	0.00	0.28
		725 - 749	0.00	0.06	0.19	0.05	0.00	0.30
		750 - 785	0.00	0.00	0.05	0.18	0.01	0.25
		786 - 850	0.00	0.00	0.00	0.01	0.01	0.02

Table A.8.15 Reliability of Classification: Grade 8 Mathematics

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.22	0.03	0.00	0.00	0.00	0.25
		700 - 724	0.04	0.16	0.05	0.00	0.00	0.25
		725 - 749	0.00	0.05	0.15	0.04	0.00	0.24
		750 - 800	0.00	0.00	0.04	0.19	0.01	0.23
		801 - 850	0.00	0.00	0.00	0.00	0.02	0.02
	Decision Consistency	650 - 699	0.21	0.05	0.00	0.00	0.00	0.27
		700 - 724	0.05	0.13	0.06	0.01	0.00	0.24
		725 - 749	0.00	0.06	0.12	0.05	0.00	0.23
		750 - 800	0.00	0.01	0.05	0.17	0.01	0.24
		801 - 850	0.00	0.00	0.00	0.01	0.02	0.03
PBT	Decision Accuracy	650 - 699	0.25	0.04	0.00	0.00	0.00	0.29
		700 - 724	0.04	0.14	0.04	0.00	0.00	0.22
		725 - 749	0.00	0.05	0.13	0.04	0.00	0.21
		750 - 800	0.00	0.00	0.04	0.21	0.02	0.27
		801 - 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 - 699	0.24	0.05	0.01	0.00	0.00	0.30
		700 - 724	0.05	0.11	0.05	0.01	0.00	0.21
		725 - 749	0.00	0.05	0.10	0.05	0.00	0.20
		750 - 800	0.00	0.01	0.05	0.19	0.02	0.26
		801 – 850	0.00	0.00	0.00	0.01	0.01	0.02

Table A.8.16 Reliability of Classification: Algebra I

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.16	0.04	0.00	0.00	0.23
		725 - 749	0.00	0.04	0.18	0.04	0.00	0.26
		750 - 804	0.00	0.00	0.04	0.30	0.01	0.35
		805 - 850	0.00	0.00	0.00	0.01	0.04	0.04
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.14	0.05	0.00	0.00	0.22
		725 - 749	0.00	0.05	0.15	0.05	0.00	0.25
		750 - 804	0.00	0.00	0.05	0.28	0.01	0.35
		805 - 850	0.00	0.00	0.00	0.01	0.04	0.05
PBT	Decision Accuracy	650 - 699	0.25	0.07	0.00	0.00	0.00	0.31
		700 - 724	0.05	0.24	0.05	0.00	0.00	0.34
		725 - 749	0.00	0.05	0.13	0.03	0.00	0.21
		750 - 804	0.00	0.00	0.02	0.10	0.01	0.13
		805 - 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 - 699	0.23	0.09	0.00	0.00	0.00	0.33
		700 - 724	0.06	0.19	0.06	0.00	0.00	0.32
		725 - 749	0.00	0.07	0.10	0.03	0.00	0.21
		750 - 804	0.00	0.00	0.04	0.10	0.00	0.14
		805 - 850	0.00	0.00	0.00	0.00	0.00	0.01

Table A.8.17 Reliability of Classification: Geometry

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.07	0.02	0.00	0.00	0.00	0.09
		700 - 724	0.02	0.26	0.03	0.00	0.00	0.32
		725 - 749	0.00	0.04	0.24	0.04	0.00	0.32
		750 - 782	0.00	0.00	0.03	0.20	0.01	0.24
		783 - 850	0.00	0.00	0.00	0.01	0.04	0.04
	Decision Consistency	650 – 699	0.07	0.03	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.23	0.05	0.00	0.00	0.31
		725 - 749	0.00	0.06	0.21	0.05	0.00	0.31
		750 - 782	0.00	0.00	0.04	0.18	0.01	0.24
		783 - 850	0.00	0.00	0.00	0.01	0.03	0.05
PBT	Decision Accuracy	650 – 699	0.12	0.06	0.00	0.00	0.00	0.18
		700 - 724	0.05	0.40	0.05	0.00	0.00	0.50
		725 - 749	0.00	0.05	0.18	0.02	0.00	0.25
		750 - 782	0.00	0.00	0.01	0.05	0.01	0.06
		783 - 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 - 699	0.11	0.09	0.00	0.00	0.00	0.21
		700 - 724	0.06	0.33	0.07	0.00	0.00	0.46
		725 - 749	0.00	0.08	0.15	0.02	0.00	0.26
		750 - 782	0.00	0.00	0.03	0.04	0.01	0.08
		783 - 850	0.00	0.00	0.00	0.00	0.00	0.00

Table A.8.18 Reliability of Classification: Algebra II

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.26	0.04	0.00	0.00	0.00	0.30
		700 - 724	0.04	0.16	0.04	0.00	0.00	0.24
		725 - 749	0.00	0.04	0.12	0.03	0.00	0.19
		750 - 807	0.00	0.00	0.03	0.19	0.01	0.24
		808 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 – 699	0.25	0.06	0.00	0.00	0.00	0.31
		700 - 724	0.05	0.13	0.05	0.00	0.00	0.23
		725 - 749	0.00	0.05	0.09	0.04	0.00	0.19
		750 - 807	0.00	0.00	0.04	0.18	0.01	0.24
		808 - 850	0.00	0.00	0.00	0.01	0.02	0.03
PBT	Decision Accuracy	650 - 699	0.32	0.06	0.00	0.00	0.00	0.38
		700 - 724	0.05	0.19	0.05	0.00	0.00	0.29
		725 - 749	0.00	0.05	0.11	0.03	0.00	0.19
		750 - 807	0.00	0.00	0.02	0.11	0.00	0.14
		808 - 850	0.00	0.00	0.00	0.00	0.00	0.01
	Decision Consistency	650 - 699	0.31	0.08	0.01	0.00	0.00	0.39
		700 - 724	0.06	0.15	0.05	0.00	0.00	0.27
		725 - 749	0.01	0.06	0.09	0.03	0.00	0.18
		750 - 807	0.00	0.01	0.04	0.10	0.00	0.15
		808 - 850	0.00	0.00	0.00	0.00	0.00	0.01

Table A.8.19 Reliability of Classification: Integrated Mathematics I

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.17	0.06	0.00	0.00	0.00	0.23
		700 - 724	0.06	0.24	0.06	0.00	0.00	0.36
		725 - 749	0.00	0.06	0.13	0.05	0.00	0.24
		750 - 798	0.00	0.00	0.03	0.12	0.01	0.16
		799 - 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 - 699	0.16	0.09	0.01	0.00	0.00	0.25
		700 - 724	0.07	0.18	0.07	0.01	0.00	0.33
		725 - 749	0.01	0.08	0.10	0.05	0.00	0.24
		750 - 798	0.00	0.01	0.04	0.11	0.01	0.17
		799 - 850	0.00	0.00	0.00	0.01	0.01	0.01

Table A.8.20 Reliability of Classification: Integrated Mathematics II

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.27	0.10	0.00	0.00	0.00	0.37
		700 - 724	0.05	0.24	0.04	0.00	0.00	0.34
		725 - 749	0.00	0.04	0.10	0.02	0.00	0.16
		750 - 784	0.00	0.00	0.02	0.07	0.01	0.10
		785 - 850	0.00	0.00	0.00	0.01	0.03	0.03
	Decision Consistency	650 - 699	0.23	0.12	0.00	0.00	0.00	0.36
		700 - 724	0.08	0.20	0.05	0.00	0.00	0.34
		725 - 749	0.00	0.06	0.08	0.03	0.00	0.17
		750 - 784	0.00	0.00	0.03	0.06	0.01	0.10
		785 - 850	0.00	0.00	0.00	0.01	0.03	0.04

Table A.8.21 Reliability of Classification: Integrated Mathematics III

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.71	0.04	0.00	0.02	0.00	0.78
		700 - 724	0.05	0.06	0.04	0.01	0.00	0.16
		725 - 749	0.00	0.01	0.04	0.00	0.00	0.05
		750 - 803	0.00	0.00	0.00	0.00	0.00	0.00
		804 - 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 - 699	0.68	0.05	0.01	0.02	0.00	0.76
		700 - 724	0.08	0.05	0.03	0.01	0.00	0.17
		725 - 749	0.01	0.02	0.03	0.00	0.00	0.06
		750 - 803	-0.54	-0.08	-0.05	-0.02	0.00	-0.69
		804 - 850	0.54	0.08	0.06	0.02	0.00	0.71

Appendix 10.1: IRT Results for Spring 2018 English Language Arts/Literacy (ELA/L)

Table A.10.1 CBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	E03	All Items	102	46	0.56	0.99	-1.35	3.13	0.55	0.27	0.16	1.07
CBT	E03	Reading	72	36	0.28	0.93	-1.35	3.13	0.44	0.18	0.16	0.84
CBT	E03	Writing	30	10	1.59	0.32	1.26	2.20	0.94	0.11	0.72	1.07
CBT	E04	All Items	131	59	0.39	1.19	-5.92	2.01	0.46	0.23	0.13	0.95
CBT	E04	Reading	98	49	0.21	1.23	-5.92	2.01	0.37	0.14	0.13	0.83
CBT	E04	Writing	33	10	1.25	0.41	0.84	2.00	0.87	0.06	0.74	0.95
CBT	E05	All Items	125	56	0.59	1.10	-1.70	3.59	0.46	0.27	0.13	1.02
CBT	E05	Reading	92	46	0.48	1.17	-1.70	3.59	0.35	0.15	0.13	0.81
CBT	E05	Writing	33	10	1.07	0.51	0.63	2.03	0.95	0.07	0.82	1.02
CBT	E06	All Items	135	60	0.59	0.91	-1.54	2.67	0.49	0.22	0.20	1.03
CBT	E06	Reading	100	50	0.45	0.92	-1.54	2.67	0.41	0.13	0.20	0.82
CBT	E06	Writing	35	10	1.26	0.41	0.67	1.86	0.88	0.12	0.69	1.03
CBT	E07	All Items	137	61	0.46	1.02	-2.29	4.13	0.48	0.31	0.14	1.27
CBT	E07	Reading	102	51	0.39	1.09	-2.29	4.13	0.36	0.15	0.14	0.82
CBT	E07	Writing	35	10	0.80	0.43	0.19	1.46	1.09	0.16	0.77	1.27
CBT	E08	All Items	129	57	0.39	0.98	-1.42	4.56	0.47	0.28	0.15	1.24
CBT	E08	Reading	94	47	0.35	1.05	-1.42	4.56	0.36	0.11	0.15	0.62
CBT	E08	Writing	35	10	0.60	0.47	-0.11	1.55	1.00	0.20	0.68	1.24
CBT	E09	All Items	141	63	0.52	0.74	-0.99	2.39	0.53	0.33	0.17	1.44
CBT	E09	Reading	106	53	0.47	0.79	-0.99	2.39	0.40	0.13	0.17	0.73
CBT	E09	Writing	35	10	0.78	0.30	0.27	1.21	1.21	0.17	0.99	1.44
CBT	E10	All Items	135	60	0.90	0.97	-0.90	4.03	0.45	0.31	0.13	1.24
CBT	E10	Reading	100	50	0.92	1.05	-0.90	4.03	0.33	0.12	0.13	0.59
CBT	E10	Writing	35	10	0.79	0.31	0.40	1.25	1.09	0.13	0.85	1.24
CBT	E11	All Items	131	58	0.95	0.77	-0.67	2.80	0.47	0.23	0.17	1.10
CBT	E11	Reading	96	48	0.93	0.83	-0.67	2.80	0.39	0.13	0.17	0.64
CBT	E11	Writing	35	10	1.08	0.32	0.61	1.51	0.85	0.19	0.56	1.10

Table A.10.2 PBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
PBT	E03	All Items	62	27	0.83	1.06	-1.08	3.44	0.59	0.29	0.11	1.02
PBT	E03	Reading	44	21	0.58	1.08	-1.08	3.44	0.48	0.24	0.11	1.02
PBT	E03	Writing	18	6	1.70	0.31	1.35	2.10	0.95	0.07	0.83	1.02
PBT	E04	All Items	84	36	0.60	0.84	-1.08	1.77	0.47	0.25	0.13	0.95
PBT	E04	Reading	64	30	0.49	0.87	-1.08	1.77	0.40	0.19	0.13	0.93
PBT	E04	Writing	20	6	1.18	0.36	0.75	1.77	0.86	0.07	0.77	0.95
PBT	E05	All Items	84	36	1.00	0.98	-1.34	3.59	0.45	0.27	0.13	0.99
PBT	E05	Reading	64	30	0.97	1.05	-1.34	3.59	0.37	0.20	0.13	0.97
PBT	E05	Writing	20	6	1.14	0.53	0.54	1.88	0.88	0.11	0.75	0.99
PBT	E06	All Items	85	36	0.44	0.85	-1.09	1.86	0.52	0.22	0.21	1.02
PBT	E06	Reading	64	30	0.29	0.84	-1.09	1.86	0.45	0.17	0.21	0.87
PBT	E06	Writing	21	6	1.20	0.43	0.67	1.86	0.86	0.11	0.73	1.02
PBT	E07	All Items	85	36	0.46	0.77	-1.70	1.60	0.52	0.32	0.17	1.23
PBT	E07	Reading	64	30	0.42	0.82	-1.70	1.60	0.42	0.23	0.17	1.23
PBT	E07	Writing	21	6	0.70	0.45	0.19	1.46	1.04	0.18	0.77	1.23
PBT	E08	All Items	85	36	0.22	0.95	-1.42	2.38	0.52	0.26	0.17	1.15
PBT	E08	Reading	64	30	0.19	1.02	-1.42	2.38	0.45	0.20	0.17	1.09
PBT	E08	Writing	21	6	0.39	0.44	-0.32	0.86	0.90	0.18	0.68	1.15
PBT	E09	All Items	85	36	0.54	0.72	-1.04	2.06	0.51	0.25	0.18	1.22
PBT	E09	Reading	64	30	0.43	0.73	-1.04	2.06	0.43	0.16	0.18	0.94
PBT	E09	Writing	21	6	1.08	0.39	0.59	1.62	0.93	0.19	0.70	1.22
PBT	E10	All Items	85	36	1.06	0.91	-0.43	4.03	0.49	0.35	0.13	1.19
PBT	E10	Reading	64	30	1.11	0.98	-0.43	4.03	0.37	0.23	0.13	1.14
PBT	E10	Writing	21	6	0.82	0.34	0.40	1.25	1.09	0.10	0.90	1.19

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
PBT	E11	All Items	85	36	1.32	0.98	-0.67	4.21	0.46	0.26	0.17	1.10
PBT	E11	Reading	64	30	1.33	1.06	-0.67	4.21	0.38	0.20	0.17	1.07
PBT	E11	Writing	21	6	1.22	0.29	0.92	1.53	0.83	0.20	0.65	1.10

Table A.10.3 CBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*SE* of *b* Estimates				*SE* of *a* Estimates			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	E03	All Items	102	46	0.010	0.008	0.004	0.052	0.005	0.002	0.002	0.011
CBT	E03	Reading	72	36	0.010	0.008	0.004	0.052	0.004	0.001	0.002	0.006
CBT	E03	Writing	30	10	0.012	0.003	0.010	0.020	0.009	0.001	0.007	0.011
CBT	E04	All Items	131	59	0.010	0.015	0.004	0.117	0.003	0.002	0.001	0.008
CBT	E04	Reading	98	49	0.010	0.016	0.004	0.117	0.003	0.001	0.001	0.005
CBT	E04	Writing	33	10	0.008	0.004	0.005	0.018	0.007	0.001	0.005	0.008
CBT	E05	All Items	125	56	0.011	0.008	0.004	0.058	0.004	0.002	0.002	0.009
CBT	E05	Reading	92	46	0.011	0.009	0.004	0.058	0.003	0.001	0.002	0.006
CBT	E05	Writing	33	10	0.007	0.004	0.005	0.018	0.008	0.001	0.006	0.009
CBT	E06	All Items	135	60	0.009	0.005	0.005	0.032	0.004	0.002	0.002	0.010
CBT	E06	Reading	100	50	0.009	0.006	0.005	0.032	0.003	0.001	0.002	0.007
CBT	E06	Writing	35	10	0.008	0.003	0.005	0.013	0.008	0.002	0.005	0.010
CBT	E07	All Items	137	61	0.010	0.010	0.003	0.065	0.004	0.003	0.002	0.012
CBT	E07	Reading	102	51	0.011	0.010	0.005	0.065	0.003	0.001	0.002	0.006
CBT	E07	Writing	35	10	0.005	0.002	0.003	0.009	0.009	0.002	0.007	0.012
CBT	E08	All Items	129	57	0.010	0.011	0.004	0.070	0.004	0.003	0.002	0.013
CBT	E08	Reading	94	47	0.011	0.012	0.005	0.070	0.003	0.001	0.002	0.005
CBT	E08	Writing	35	10	0.007	0.003	0.004	0.014	0.009	0.002	0.006	0.013
CBT	E09	All Items	141	63	0.012	0.006	0.005	0.038	0.006	0.004	0.003	0.019
CBT	E09	Reading	106	53	0.013	0.006	0.005	0.038	0.005	0.001	0.003	0.007
CBT	E09	Writing	35	10	0.007	0.001	0.005	0.009	0.015	0.003	0.009	0.019
CBT	E10	All Items	135	60	0.016	0.016	0.004	0.095	0.005	0.003	0.002	0.015
CBT	E10	Reading	100	50	0.017	0.017	0.006	0.095	0.004	0.001	0.002	0.006

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*SE* of *b* Estimates				*SE* of *a* Estimates			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	E10	Writing	35	10	0.006	0.001	0.004	0.008	0.012	0.002	0.009	0.015
CBT	E11	All Items	131	58	0.017	0.013	0.006	0.064	0.006	0.003	0.003	0.017
CBT	E11	Reading	96	48	0.018	0.014	0.006	0.064	0.005	0.001	0.003	0.009
CBT	E11	Writing	35	10	0.011	0.003	0.006	0.015	0.012	0.003	0.008	0.017

Table A.10.4 CBT IRT Model Fit for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	E03	All Items	102	46	2705.6	2157.7	523.0	8368.7	2575.7	2092.2	513.3	8253.3
CBT	E03	Reading	72	36	3034.8	2323.0	523.0	8368.7	2872.4	2262.5	513.3	8253.3
CBT	E03	Writing	30	10	1520.6	546.3	901.7	2590.3	1507.6	599.1	917.1	2784.3
CBT	E04	All Items	131	59	3760.9	3306.8	587.0	14921.7	3731.2	3514.4	584.9	16584.2
CBT	E04	Reading	98	49	4063.3	3534.3	587.0	14921.7	4042.2	3761.3	584.9	16584.2
CBT	E04	Writing	33	10	2279.2	955.7	1539.2	4388.9	2207.1	1018.6	1374.2	4555.7
CBT	E05	All Items	125	56	2916.0	2198.0	713.2	9495.7	2833.6	2155.9	686.0	9588.9
CBT	E05	Reading	92	46	3086.0	2367.7	713.2	9495.7	2997.2	2322.9	686.0	9588.9
CBT	E05	Writing	33	10	2133.7	815.5	1334.8	3864.4	2081.2	810.9	1230.4	3715.0
CBT	E06	All Items	135	60	2938.5	2513.7	598.2	13431.5	2777.8	2360.1	556.1	11539.3
CBT	E06	Reading	100	50	3146.1	2697.5	598.2	13431.5	2987.9	2525.3	556.1	11539.3
CBT	E06	Writing	35	10	1900.2	606.3	1376.2	2944.9	1727.1	568.7	1221.4	2709.6
CBT	E07	All Items	137	61	2105.7	1670.9	251.8	8072.7	2030.2	1648.0	241.0	7673.3
CBT	E07	Reading	102	51	2010.7	1787.3	251.8	8072.7	1937.7	1761.2	241.0	7673.3
CBT	E07	Writing	35	10	2590.2	744.0	1927.7	4086.1	2502.2	759.0	1735.6	3755.4
CBT	E08	All Items	129	57	2677.9	2206.6	150.1	9687.4	2574.3	2249.4	146.7	11232.3
CBT	E08	Reading	94	47	2620.9	2364.7	150.1	9687.4	2571.8	2433.7	146.7	11232.3
CBT	E08	Writing	35	10	2945.6	1272.8	1788.5	5643.7	2586.0	1100.5	1575.5	4944.6
CBT	E09	All Items	141	63	1359.6	1127.8	232.4	5524.6	1309.0	1165.9	221.9	5277.8
CBT	E09	Reading	106	53	1384.9	1222.1	232.4	5524.6	1358.0	1261.8	221.9	5277.8
CBT	E09	Writing	35	10	1225.5	329.3	952.0	1910.9	1049.3	275.3	842.4	1600.5
CBT	E10	All Items	135	60	1933.3	1617.9	185.1	9157.8	1837.2	1581.0	186.5	8347.6
CBT	E10	Reading	100	50	1883.3	1756.8	185.1	9157.8	1826.9	1723.5	186.5	8347.6

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	E10	Writing	35	10	2183.3	521.8	1668.9	2948.7	1888.7	458.6	1486.5	2592.8
CBT	E11	All Items	131	58	1021.0	658.5	169.2	3545.0	933.8	605.5	151.1	3488.6
CBT	E11	Reading	96	48	965.9	677.0	169.2	3545.0	896.9	630.2	151.1	3488.6
CBT	E11	Writing	35	10	1285.5	509.2	871.6	2214.0	1110.9	453.9	772.5	1934.6

Appendix 10.2: IRT Results for Spring 2018 Mathematics

Table A.10.5 CBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	M03	All Items	132	86	-0.20	1.09	-2.91	2.09	0.76	0.22	0.31	1.29
CBT	M03	SSMC	18	18	-0.86	0.99	-2.91	1.22	0.80	0.24	0.46	1.29
CBT	M03	CR	114	68	-0.02	1.06	-2.76	2.09	0.75	0.22	0.31	1.22
CBT	M03	Type I	80	72	-0.38	1.08	-2.91	1.74	0.80	0.21	0.38	1.29
CBT	M03	Type II	28	8	0.79	0.63	0.15	2.09	0.56	0.20	0.31	0.95
CBT	M03	Type III	24	6	0.72	0.22	0.31	0.90	0.65	0.16	0.43	0.89
CBT	M04	All Items	132	80	-0.12	1.04	-2.61	2.81	0.69	0.20	0.22	1.36
CBT	M04	SSMC	20	20	-0.58	1.10	-2.61	1.29	0.76	0.20	0.48	1.36
CBT	M04	CR	112	60	0.03	0.99	-2.36	2.81	0.66	0.19	0.22	1.09
CBT	M04	Type I	80	66	-0.35	0.99	-2.61	2.81	0.71	0.20	0.22	1.36
CBT	M04	Type II	28	8	0.86	0.28	0.52	1.32	0.58	0.09	0.40	0.69
CBT	M04	Type III	24	6	1.09	0.50	0.12	1.45	0.56	0.17	0.34	0.80
CBT	M05	All Items	132	80	0.10	1.22	-3.06	4.53	0.64	0.23	0.23	1.31
CBT	M05	SSMC	20	20	-0.27	1.33	-1.74	4.53	0.62	0.27	0.23	1.31
CBT	M05	CR	112	60	0.22	1.16	-3.06	2.65	0.64	0.21	0.27	1.07
CBT	M05	Type I	80	66	-0.09	1.22	-3.06	4.53	0.65	0.24	0.23	1.31
CBT	M05	Type II	28	8	0.94	0.53	0.11	1.63	0.59	0.08	0.51	0.75
CBT	M05	Type III	24	6	1.11	0.83	-0.13	2.03	0.51	0.15	0.29	0.75
CBT	M06	All Items	132	76	0.42	1.01	-2.55	2.75	0.70	0.24	0.24	1.35
CBT	M06	SSMC	22	22	0.06	1.19	-2.25	2.75	0.61	0.19	0.26	0.96
CBT	M06	CR	110	54	0.57	0.90	-2.55	2.26	0.74	0.25	0.24	1.35
CBT	M06	Type I	80	62	0.27	1.04	-2.55	2.75	0.73	0.25	0.24	1.35
CBT	M06	Type II	28	8	1.02	0.53	0.39	1.73	0.55	0.08	0.43	0.66
CBT	M06	Type III	24	6	1.20	0.38	0.75	1.69	0.60	0.15	0.44	0.78
CBT	M07	All Items	132	76	0.73	1.00	-1.39	2.93	0.74	0.26	0.25	1.39
CBT	M07	SSMC	16	16	-0.04	0.95	-1.39	2.14	0.68	0.29	0.25	1.24
CBT	M07	CR	116	60	0.93	0.92	-0.84	2.93	0.76	0.24	0.28	1.39

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	M07	Type I	80	62	0.62	1.04	-1.39	2.93	0.78	0.26	0.25	1.39
CBT	M07	Type II	28	8	1.09	0.45	0.13	1.55	0.61	0.10	0.42	0.78
CBT	M07	Type III	24	6	1.37	0.70	0.14	2.09	0.55	0.16	0.33	0.74
CBT	M08	All Items	132	72	0.95	1.19	-1.83	3.97	0.60	0.22	0.24	1.18
CBT	M08	SSMC	17	17	0.03	1.17	-1.83	2.35	0.50	0.17	0.27	0.85
CBT	M08	CR	115	55	1.24	1.06	-1.26	3.97	0.64	0.22	0.24	1.18
CBT	M08	Type I	80	58	0.76	1.20	-1.83	3.97	0.60	0.23	0.24	1.18
CBT	M08	Type II	28	8	1.78	0.64	1.03	2.74	0.63	0.16	0.32	0.87
CBT	M08	Type III	24	6	1.78	0.97	0.12	2.62	0.58	0.12	0.43	0.74
CBT	A1	All Items	157	82	1.38	1.08	-1.16	4.16	0.66	0.26	0.11	1.29
CBT	A1	SSMC	20	20	0.84	1.17	-1.16	3.12	0.47	0.16	0.11	0.71
CBT	A1	CR	137	62	1.55	1.00	-0.77	4.16	0.72	0.26	0.20	1.29
CBT	A1	Type I	97	67	1.28	1.13	-1.16	4.16	0.66	0.27	0.11	1.29
CBT	A1	Type II	24	7	1.93	0.90	0.05	2.69	0.63	0.22	0.37	0.96
CBT	A1	Type III	36	8	1.74	0.56	0.47	2.15	0.64	0.14	0.45	0.83
CBT	GO	All Items	160	84	1.15	0.93	-1.49	3.53	0.70	0.26	0.23	1.68
CBT	GO	SSMC	15	15	0.35	1.01	-1.49	1.97	0.56	0.15	0.26	0.78
CBT	GO	CR	145	69	1.32	0.83	-0.78	3.53	0.73	0.27	0.23	1.68
CBT	GO	Type I	96	68	0.98	0.95	-1.49	3.53	0.69	0.28	0.23	1.68
CBT	GO	Type II	28	8	1.82	0.30	1.40	2.15	0.78	0.12	0.58	0.91
CBT	GO	Type III	36	8	1.89	0.35	1.36	2.35	0.70	0.16	0.48	0.94
CBT	A2	All Items	161	81	1.43	0.96	-1.41	3.42	0.62	0.27	0.20	1.26
CBT	A2	SSMC	21	21	0.94	0.91	-0.49	2.71	0.54	0.25	0.21	1.12
CBT	A2	CR	140	60	1.60	0.92	-1.41	3.42	0.64	0.27	0.20	1.26
CBT	A2	Type I	97	65	1.28	0.96	-1.41	3.42	0.62	0.28	0.20	1.26
CBT	A2	Type II	28	8	2.03	0.50	1.36	2.82	0.64	0.22	0.36	1.07
CBT	A2	Type III	36	8	2.05	0.84	0.87	3.28	0.56	0.24	0.29	0.92
CBT	M1	All Items	81	42	1.22	0.91	-0.50	2.80	0.59	0.19	0.20	1.07
CBT	M1	SSMC	13	13	0.79	0.87	-0.24	2.34	0.52	0.15	0.20	0.71
CBT	M1	CR	68	29	1.41	0.87	-0.50	2.80	0.63	0.19	0.24	1.07

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
CBT	M1	Type I	49	34	1.05	0.87	-0.50	2.80	0.58	0.20	0.20	1.07
CBT	M1	Type II	14	4	1.48	0.88	0.52	2.32	0.67	0.07	0.57	0.73
CBT	M1	Type III	18	4	2.37	0.28	2.02	2.64	0.60	0.19	0.42	0.83
CBT	M2	All Items	80	41	1.90	1.45	-0.74	5.99	0.53	0.27	0.12	1.18
CBT	M2	SSMC	14	14	1.01	1.21	-0.74	3.24	0.37	0.13	0.12	0.56
CBT	M2	CR	66	27	2.36	1.37	0.04	5.99	0.62	0.29	0.17	1.18
CBT	M2	Type I	48	33	1.81	1.53	-0.74	5.99	0.52	0.29	0.12	1.18
CBT	M2	Type II	14	4	2.87	0.82	2.04	3.76	0.56	0.11	0.43	0.68
CBT	M2	Type III	18	4	1.65	0.98	0.20	2.27	0.64	0.16	0.47	0.79
CBT	M3	All Items	81	40	1.27	1.08	-2.27	4.28	0.59	0.27	0.16	1.27
CBT	M3	SSMC	12	12	0.84	1.71	-2.27	4.28	0.44	0.15	0.23	0.66
CBT	M3	CR	69	28	1.46	0.61	-0.04	2.79	0.66	0.29	0.16	1.27
CBT	M3	Type I	49	32	1.20	1.17	-2.27	4.28	0.58	0.29	0.16	1.27
CBT	M3	Type II	14	4	1.58	0.61	0.96	2.20	0.74	0.08	0.63	0.83
CBT	M3	Type III	18	4	1.57	0.57	0.90	2.11	0.53	0.20	0.30	0.76

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table A.10.6 PBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
PBT	M03	All Items	65	42	-0.30	1.17	-3.03	1.58	0.77	0.23	0.28	1.29
PBT	M03	SSMC	14	14	-1.38	1.04	-3.03	0.03	0.76	0.26	0.28	1.29
PBT	M03	CR	51	28	0.24	0.80	-1.57	1.58	0.77	0.22	0.32	1.19
PBT	M03	Type I	39	35	-0.48	1.17	-3.03	1.58	0.80	0.23	0.28	1.29
PBT	M03	Type II	14	4	0.41	0.66	-0.06	1.39	0.62	0.23	0.37	0.89
PBT	M03	Type III	12	3	0.88	0.55	0.38	1.46	0.61	0.23	0.43	0.87
PBT	M04	All Items	66	40	-0.12	1.13	-2.69	1.93	0.73	0.18	0.33	1.09
PBT	M04	SSMC	14	14	-1.08	0.96	-2.69	0.65	0.75	0.15	0.48	0.95
PBT	M04	CR	52	26	0.39	0.86	-1.63	1.93	0.72	0.20	0.33	1.09
PBT	M04	Type I	40	33	-0.39	1.05	-2.69	1.93	0.75	0.19	0.33	1.09
PBT	M04	Type II	14	4	1.03	0.35	0.57	1.41	0.59	0.10	0.45	0.66
PBT	M04	Type III	12	3	1.33	0.22	1.09	1.53	0.69	0.15	0.52	0.81
PBT	M05	All Items	66	40	0.19	0.93	-1.43	2.65	0.66	0.24	0.22	1.31
PBT	M05	SSMC	14	14	-0.20	0.88	-1.43	1.61	0.64	0.28	0.22	1.31
PBT	M05	CR	52	26	0.40	0.91	-1.21	2.65	0.67	0.23	0.26	1.27
PBT	M05	Type I	40	33	0.08	0.95	-1.43	2.65	0.68	0.25	0.22	1.31
PBT	M05	Type II	14	4	0.55	0.82	-0.59	1.37	0.48	0.15	0.26	0.59
PBT	M05	Type III	12	3	0.96	0.44	0.51	1.39	0.66	0.21	0.42	0.80
PBT	M06	All Items	66	38	0.38	0.92	-2.25	1.73	0.70	0.24	0.26	1.35
PBT	M06	SSMC	13	13	-0.27	1.02	-2.25	1.64	0.65	0.25	0.26	1.09
PBT	M06	CR	53	25	0.72	0.66	-0.70	1.73	0.72	0.23	0.29	1.35
PBT	M06	Type I	40	31	0.24	0.92	-2.25	1.73	0.73	0.24	0.26	1.35
PBT	M06	Type II	14	4	0.69	0.75	-0.15	1.63	0.44	0.05	0.38	0.50
PBT	M06	Type III	12	3	1.40	0.27	1.14	1.69	0.67	0.16	0.49	0.78
PBT	M07	All Items	66	38	0.65	0.99	-1.66	2.24	0.73	0.27	0.28	1.39
PBT	M07	SSMC	12	12	0.05	1.16	-1.66	1.99	0.62	0.28	0.32	1.24
PBT	M07	CR	54	26	0.92	0.79	-0.69	2.24	0.79	0.26	0.28	1.39

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
PBT	M07	Type I	40	31	0.50	1.00	-1.66	1.99	0.76	0.29	0.28	1.39
PBT	M07	Type II	14	4	1.15	0.87	0.13	2.24	0.63	0.13	0.46	0.77
PBT	M07	Type III	12	3	1.47	0.44	1.01	1.87	0.57	0.21	0.33	0.74
PBT	M08	All Items	66	36	0.80	1.13	-2.29	2.35	0.64	0.22	0.24	1.08
PBT	M08	SSMC	11	11	-0.17	1.18	-2.29	1.85	0.49	0.18	0.24	0.82
PBT	M08	CR	55	25	1.23	0.82	-0.39	2.35	0.71	0.21	0.35	1.08
PBT	M08	Type I	40	29	0.57	1.15	-2.29	2.35	0.64	0.23	0.24	1.08
PBT	M08	Type II	14	4	1.74	0.18	1.52	1.94	0.65	0.22	0.35	0.87
PBT	M08	Type III	12	3	1.74	0.34	1.35	2.01	0.67	0.18	0.48	0.84
PBT	A1	All Items	81	42	1.27	1.10	-1.16	3.13	0.64	0.24	0.21	1.16
PBT	A1	SSMC	15	15	0.61	1.02	-1.16	2.11	0.52	0.18	0.21	0.83
PBT	A1	CR	66	27	1.63	0.97	-0.60	3.13	0.70	0.25	0.21	1.16
PBT	A1	Type I	49	34	1.16	1.11	-1.16	3.13	0.64	0.26	0.21	1.16
PBT	A1	Type II	14	4	1.77	1.17	0.05	2.66	0.58	0.12	0.47	0.72
PBT	A1	Type III	18	4	1.68	0.92	0.31	2.29	0.67	0.16	0.45	0.85
PBT	GO	All Items	81	43	1.03	1.07	-1.98	2.68	0.65	0.29	0.23	1.68
PBT	GO	SSMC	16	16	0.38	1.17	-1.98	2.38	0.54	0.20	0.24	0.82
PBT	GO	CR	65	27	1.42	0.80	-0.14	2.68	0.71	0.32	0.23	1.68
PBT	GO	Type I	49	35	0.86	1.10	-1.98	2.68	0.63	0.31	0.23	1.68
PBT	GO	Type II	14	4	1.72	0.34	1.40	2.04	0.82	0.16	0.58	0.91
PBT	GO	Type III	18	4	1.88	0.66	1.26	2.54	0.62	0.21	0.35	0.86
PBT	A2	All Items	81	41	1.52	0.86	-0.44	3.42	0.60	0.25	0.24	1.08
PBT	A2	SSMC	13	13	1.13	0.97	-0.44	2.86	0.55	0.19	0.25	0.89
PBT	A2	CR	68	28	1.70	0.75	0.02	3.42	0.62	0.28	0.24	1.08
PBT	A2	Type I	49	33	1.41	0.90	-0.44	3.42	0.56	0.24	0.24	1.08
PBT	A2	Type II	14	4	1.92	0.29	1.59	2.29	0.85	0.25	0.54	1.07
PBT	A2	Type III	18	4	2.04	0.63	1.30	2.79	0.66	0.29	0.38	0.92
PBT	M1	All Items	81	42	1.19	0.94	-0.77	3.36	0.51	0.19	0.20	0.99
PBT	M1	SSMC	16	16	0.68	0.77	-0.77	2.21	0.39	0.13	0.20	0.57
PBT	M1	CR	65	26	1.50	0.91	-0.54	3.36	0.57	0.19	0.29	0.99

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	*b* Estimates Summary				*a* Estimates Summary			
					Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
PBT	M1	Type I	49	34	1.02	0.86	-0.77	2.90	0.49	0.20	0.20	0.99
PBT	M1	Type II	14	4	1.61	1.40	0.44	3.36	0.54	0.13	0.40	0.71
PBT	M1	Type III	18	4	2.19	0.28	1.92	2.55	0.59	0.17	0.42	0.78
PBT	M2	All	80	42	1.67	1.14	-0.50	4.87	0.52	0.24	0.18	1.15
PBT	M2	SSMC	19	19	1.11	0.91	-0.17	3.63	0.41	0.14	0.18	0.71
PBT	M2	CR	61	23	2.13	1.13	-0.50	4.87	0.61	0.26	0.22	1.15
PBT	M2	I	48	34	1.65	1.20	-0.17	4.87	0.50	0.25	0.18	1.15
PBT	M2	II	14	4	2.14	0.18	1.94	2.34	0.63	0.09	0.51	0.70
PBT	M2	III	18	4	1.40	1.28	-0.50	2.27	0.60	0.14	0.47	0.79
PBT	M3	All	80	39	1.31	1.33	-2.27	4.61	0.56	0.26	0.13	1.24
PBT	M3	SSMC	12	12	0.73	1.67	-2.27	4.28	0.52	0.27	0.20	1.24
PBT	M3	CR	68	27	1.57	1.08	-0.35	4.61	0.58	0.26	0.13	1.10
PBT	M3	I	48	31	1.20	1.43	-2.27	4.61	0.54	0.27	0.13	1.24
PBT	M3	II	14	4	1.62	0.48	0.96	1.99	0.72	0.14	0.52	0.83
PBT	M3	III	18	4	1.87	0.95	0.71	2.99	0.53	0.23	0.25	0.76

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Appendix 12.1: Form Composition

Table A.12.1 Form Composition for ELA/L Grade 3

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	9 - 12	19 - 25
	Reading Informational Text	9 - 12	19 - 25
	Vocabulary	4 - 7	8 - 14
	Claim Total	22	46
Writing			
	Written Expression	2	27
	Knowledge of Conventions	1	9
	Claim Total	3	36
SUMMATIVE TOTAL		23	82

Note: This table is identical to Table 12.1 in Section 12.

Table A.12.2 Form Composition for ELA/L Grade 4

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 - 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	33
	Knowledge of Conventions	1	9
	Claim Total	3	42
SUMMATIVE TOTAL		33	106

Table A.12.3 Form Composition for ELA/L Grade 5

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	10 - 16	24 - 34
	Reading Informational Text	9 - 13	22 - 28
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	33
	Knowledge of Conventions	1	9
	Claim Total	3	42
SUMMATIVE TOTAL		33	106

Table A.12.4 Form Composition for ELA/L Grade 6

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 - 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	36
	Knowledge of Conventions	1	9
	Claim Total	3	45
SUMMATIVE TOTAL		33	109

Table A.12.5 Form Composition for ELA/L Grade 7

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 - 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	36
	Knowledge of Conventions	1	9
	Claim Total	3	45
SUMMATIVE TOTAL		33	109

Table A.12.6 Form Composition for ELA/L Grade 8

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 - 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	36
	Knowledge of Conventions	1	9
	Claim Total	3	45
SUMMATIVE TOTAL		33	109

Table A.12.7 Form Composition for ELA/L Grade 9

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 - 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	36
	Knowledge of Conventions	1	9
	Claim Total	3	45
SUMMATIVE TOTAL		33	109

Table A.12.8 Form Composition for ELA/L Grade 10

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 - 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	36
	Knowledge of Conventions	1	9
	Claim Total	3	45
SUMMATIVE TOTAL		33	109

Table A.12.9 Form Composition for ELA/L Grade 11

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 19	16 – 40
	Reading Informational Text	6 - 17	14 - 36
	Vocabulary	4 - 8	8 - 16
	Claim Total	30	64
Writing			
	Written Expression	2	36
	Knowledge of Conventions	1	9
	Claim Total	3	45
SUMMATIVE TOTAL		33	109

Table A.12.10 Form Composition for Mathematics Grade 3

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	26	30
	Additional & Supporting Content	10	10
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		43	66

Note: This table is identical to Table 12.3 in Section 12.

Table A.12.11 Form Composition for Mathematics Grade 4

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	25	31
	Additional & Supporting Content	8	9
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		40	66

Table A.12.12 Form Composition for Mathematics Grade 5

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	25	30
	Additional & Supporting Content	8	10
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		40	66

Table A.12.13 Form Composition for Mathematics Grade 6

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	20	26
	Additional & Supporting Content	11	14
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		38	66

Table A.12.14 Form Composition for Mathematics Grade 7

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	23	29
	Additional & Supporting Content	8	11
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		38	66

Table A.12.15 Form Composition for Mathematics Grade 8

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	27
	Additional & Supporting Content	8	13
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		36	66

Table A.12.16 Form Composition for Algebra I

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	28
	Additional & Supporting Content	13	21
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		42	81

Table A.12.17 Form Composition for Geometry

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	30
	Additional & Supporting Content	14	19
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		43	81

Table A.12.18 Form Composition for Algebra II

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	20	29
	Additional & Supporting Content	13	20
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		41	81

Table A.12.19 Form Composition for Integrated Mathematics I

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	31
	Additional & Supporting Content	13	18
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		42	81

Table A.12.20 Form Composition for Integrated Mathematics II

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	22	32
	Additional & Supporting Content	12	17
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		42	81

Table A.12.21 Form Composition for Integrated Mathematics III

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	19	26
	Additional & Supporting Content	13	23
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		40	81

Appendix 12.2: Scaling Constants and Associated Information

Table A.12.22 Threshold Scores and Scaling Constants for ELA/L Grades 3 to 8

PARCC Assessment	Threshold Cut	Theta	Scale Score	*A*	*B*
Grade 3 ELA	Level 2 Cut	-0.9648	700	36.7227	735.4297
	Level 3 Cut	-0.2840	726		
	Level 4 Cut	0.3968	750		
	Level 5 Cut	2.0360	810		
Grade 4 ELA	Level 2 Cut	-1.3004	700	31.5462	741.0214
	Level 3 Cut	-0.5079	725		
	Level 4 Cut	0.2846	750		
	Level 5 Cut	1.5578	790		
Grade 5 ELA	Level 2 Cut	-1.3411	700	29.4580	739.5050
	Level 3 Cut	-0.4924	726		
	Level 4 Cut	0.3563	750		
	Level 5 Cut	2.0224	799		
Grade 6 ELA	Level 2 Cut	-1.3656	700	28.3160	738.6673
	Level 3 Cut	-0.4827	725		
	Level 4 Cut	0.4002	750		
	Level 5 Cut	1.8133	790		
Grade 7 ELA	Level 2 Cut	-1.2488	700	33.9161	742.3542
	Level 3 Cut	-0.5117	725		
	Level 4 Cut	0.2254	750		
	Level 5 Cut	1.2614	785		
Grade 8 ELA	Level 2 Cut	-1.2730	700	34.1183	743.4330
	Level 3 Cut	-0.5402	725		
	Level 4 Cut	0.1925	750		
	Level 5 Cut	1.4696	794		

Table A.12.23 Threshold Scores and Scaling Constants for Mathematics Grades 3 to 8

PARCC Assessment	Threshold Cut	Theta	Scale Score	*A*	*B*
Grade 3 Mathematics	Level 2 Cut	-1.4141	700	32.1135	745.4119
	Level 3 Cut	-0.6356	727		
	Level 4 Cut	0.1429	750		
	Level 5 Cut	1.3931	790		
Grade 4 Mathematics	Level 2 Cut	-1.3840	700	29.9167	741.4049
	Level 3 Cut	-0.5484	727		
	Level 4 Cut	0.2873	750		
	Level 5 Cut	1.8323	796		
Grade 5 Mathematics	Level 2 Cut	-1.4571	700	29.0301	742.2997
	Level 3 Cut	-0.5959	725		
	Level 4 Cut	0.2653	750		
	Level 5 Cut	1.6262	790		
Grade 6 Mathematics	Level 2 Cut	-1.3829	700	28.1465	738.9252
	Level 3 Cut	-0.4948	725		
	Level 4 Cut	0.3935	750		
	Level 5 Cut	1.7567	788		
Grade 7 Mathematics	Level 2 Cut	-1.4464	700	25.1033	736.3102
	Level 3 Cut	-0.4505	725		
	Level 4 Cut	0.5453	750		
	Level 5 Cut	1.9919	786		
Grade 8 Mathematics	Level 2 Cut	-0.8851	700	32.9505	729.1640
	Level 3 Cut	-0.1264	728		
	Level 4 Cut	0.6323	750		
	Level 5 Cut	2.1896	801		

Table A.12.24 Threshold Scores and Scaling Constants for High School ELA

PARCC Assessment	Threshold Cut	Theta	Scale Score	*A*	*B*
Grade 9 ELA	Level 2 Cut	-1.1635	700	34.2174	739.8124
	Level 3 Cut	-0.4329	726		
	Level 4 Cut	0.2977	750		
	Level 5 Cut	1.5065	791		
Grade 10 ELA	Level 2 Cut	-0.8909	700	43.1280	738.4223
	Level 3 Cut	-0.3112	725		
	Level 4 Cut	0.2684	750		
	Level 5 Cut	1.2858	794		
Grade 11 ELA	Level 2 Cut	-1.1017	700	34.9278	738.4801
	Level 3 Cut	-0.3859	726		
	Level 4 Cut	0.3298	750		
	Level 5 Cut	1.5206	792		

Table A.12.25 Threshold Scores and Scaling Constants for High School Mathematics

PARCC Assessment	Threshold Cut	Theta	Scale Score	*A*	*B*
Algebra I	Level 2 Cut	-1.1781	700	31.5325	737.1490
	Level 3 Cut	-0.3853	728		
	Level 4 Cut	0.4075	750		
	Level 5 Cut	2.1651	805		
Algebra II	Level 2 Cut	-0.5759	700	37.7676	721.7509
	Level 3 Cut	0.0860	726		
	Level 4 Cut	0.7480	750		
	Level 5 Cut	2.2728	808		
Geometry	Level 2 Cut	-1.3013	700	25.9775	733.8039
	Level 3 Cut	-0.3389	726		
	Level 4 Cut	0.6235	750		
	Level 5 Cut	1.8940	783		
Integrated Mathematics I	Level 2 Cut	-1.0919	700	32.0043	734.9446
	Level 3 Cut	-0.3107	726		
	Level 4 Cut	0.4704	750		
	Level 5 Cut	1.9934	799		
Integrated Mathematics II	Level 2 Cut	-0.9175	700	29.2865	726.8695
	Level 3 Cut	-0.0638	725		
	Level 4 Cut	0.7898	750		
	Level 5 Cut	1.9817	785		
Integrated Mathematics III	Level 2 Cut	-0.7076	700	37.3549	726.4336
	Level 3 Cut	-0.0384	726		
	Level 4 Cut	0.6309	750		
	Level 5 Cut	2.0689	804		

Table A.12.26 Scaling Constants for Reading and Writing Grades 3 to 11

	Reading		Writing	
	A_R	B_R	A_W	B_W
Grade 3 ELA	14.6891	44.1719	7.3445	32.0859
Grade 4 ELA	12.6184	46.4086	6.3093	33.2043
Grade 5 ELA	11.7832	45.8019	5.8916	32.9010
Grade 6 ELA	11.3264	45.4669	5.6632	32.7335
Grade 7 ELA	13.5664	46.9416	6.7832	33.4708
Grade 8 ELA	13.6472	47.3732	6.8237	33.6866
Grade 9 ELA	13.6870	45.9250	6.8435	32.9625
Grade 10 ELA	17.2512	45.3690	8.6256	32.6845
Grade 11 ELA	13.9712	45.3920	6.9856	32.6961

Appendix 12.3: IRT Test Characteristic Curves, Information Curves, and CSEM Curves

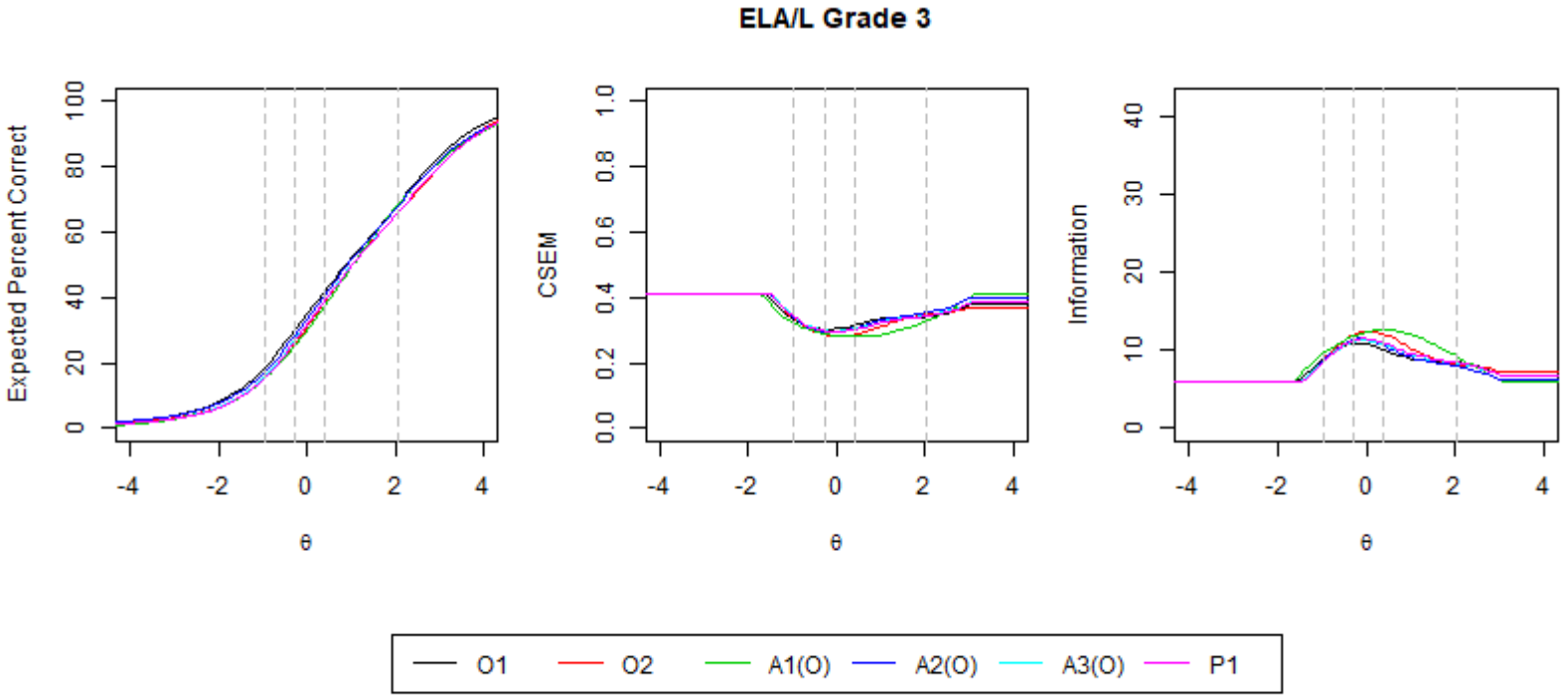

Figure A.12.1 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 3

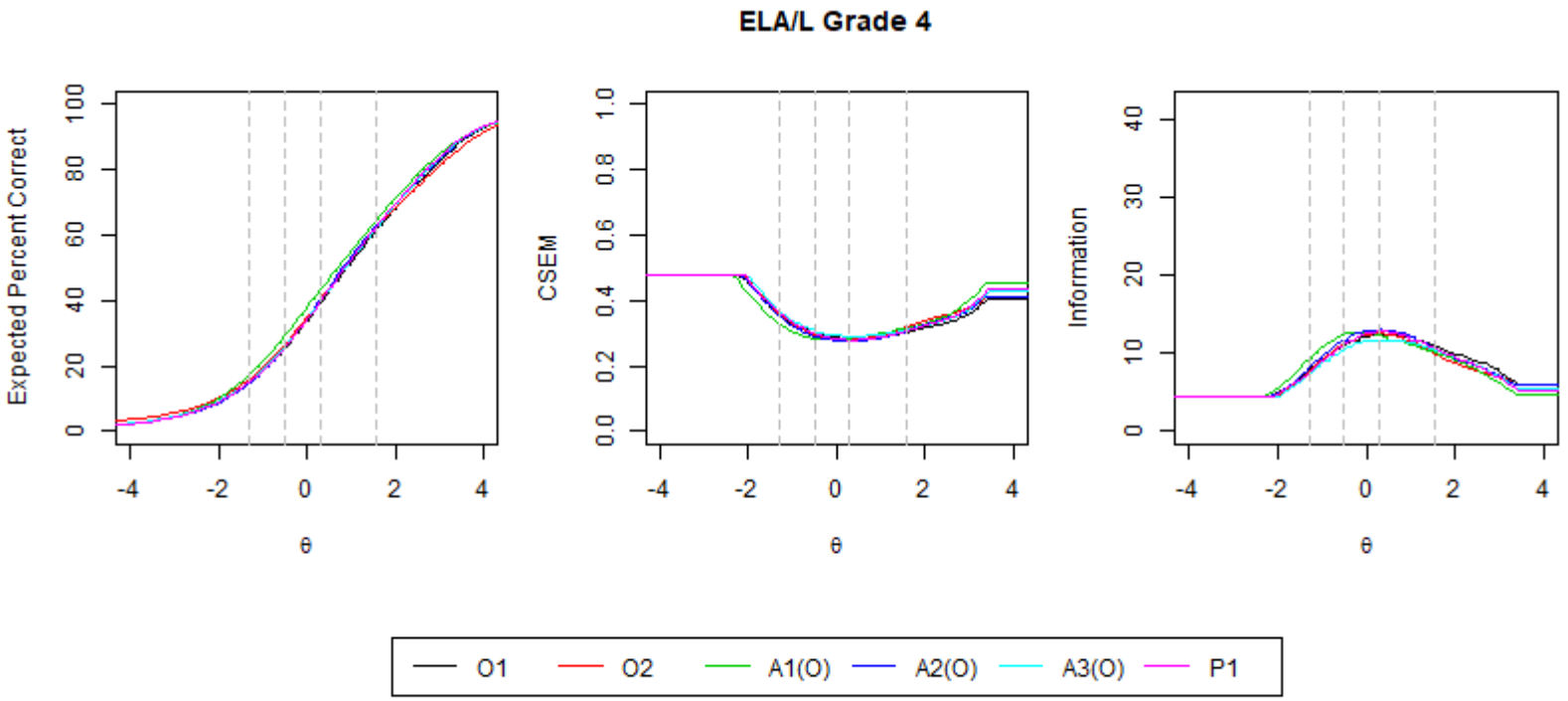

Figure A.12.2 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 4

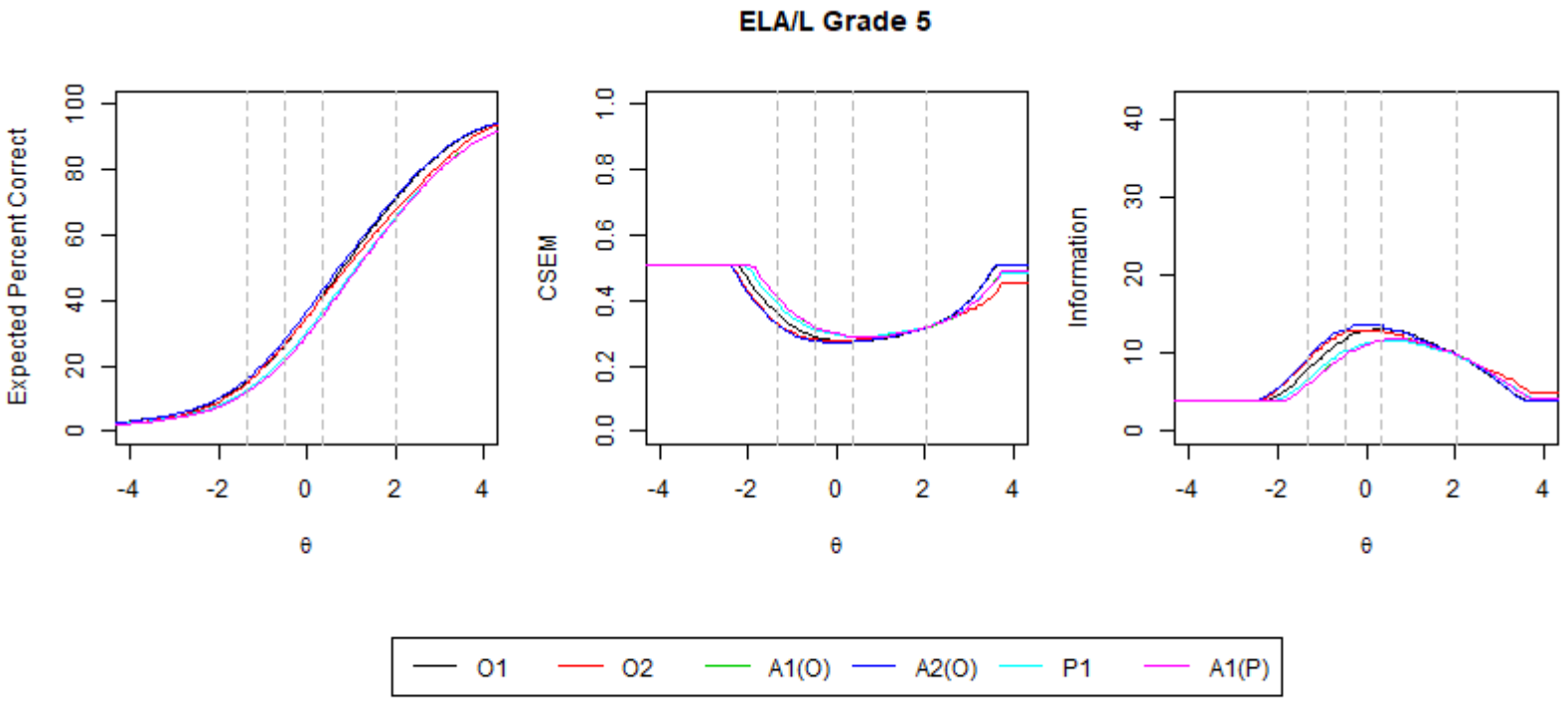

Figure A.12.3 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 5

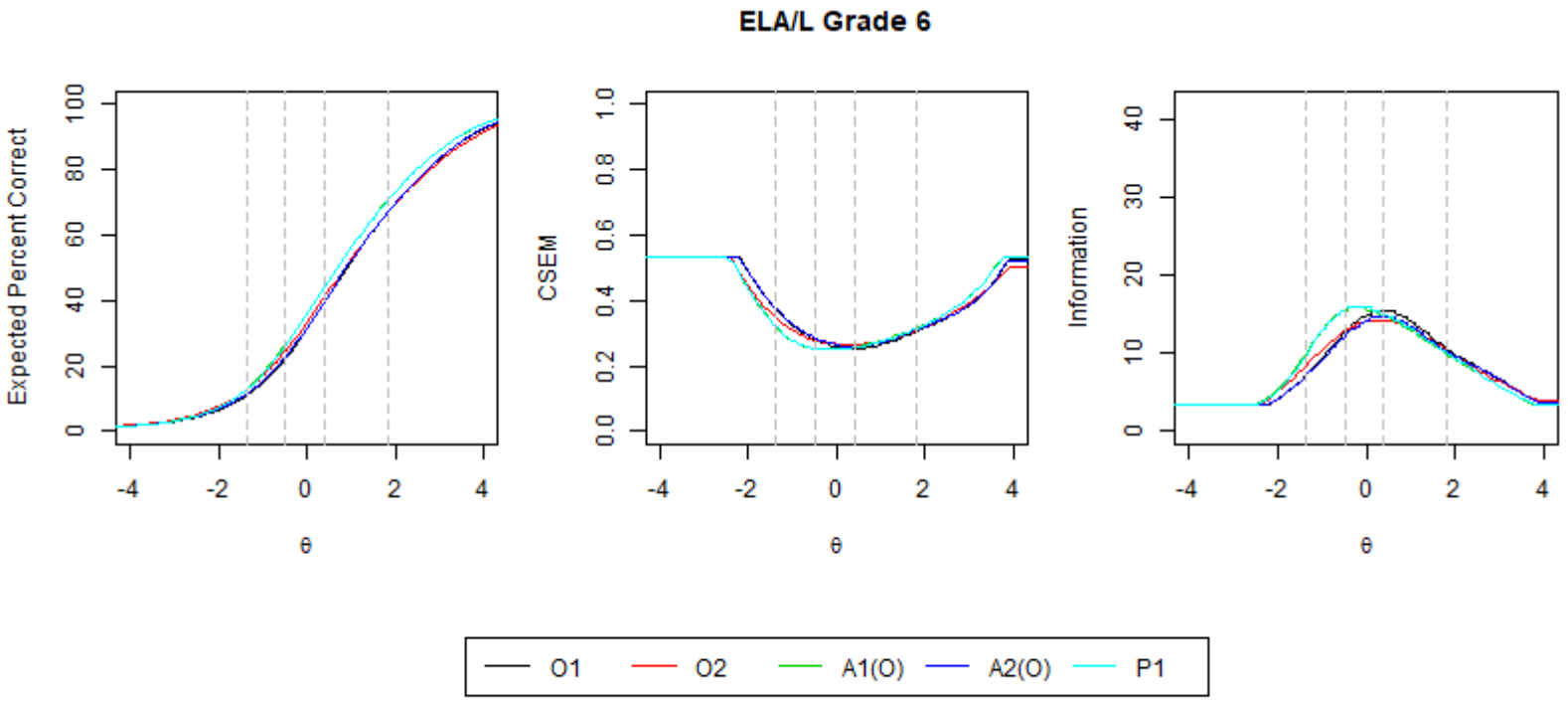

Figure A.12.4 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 6

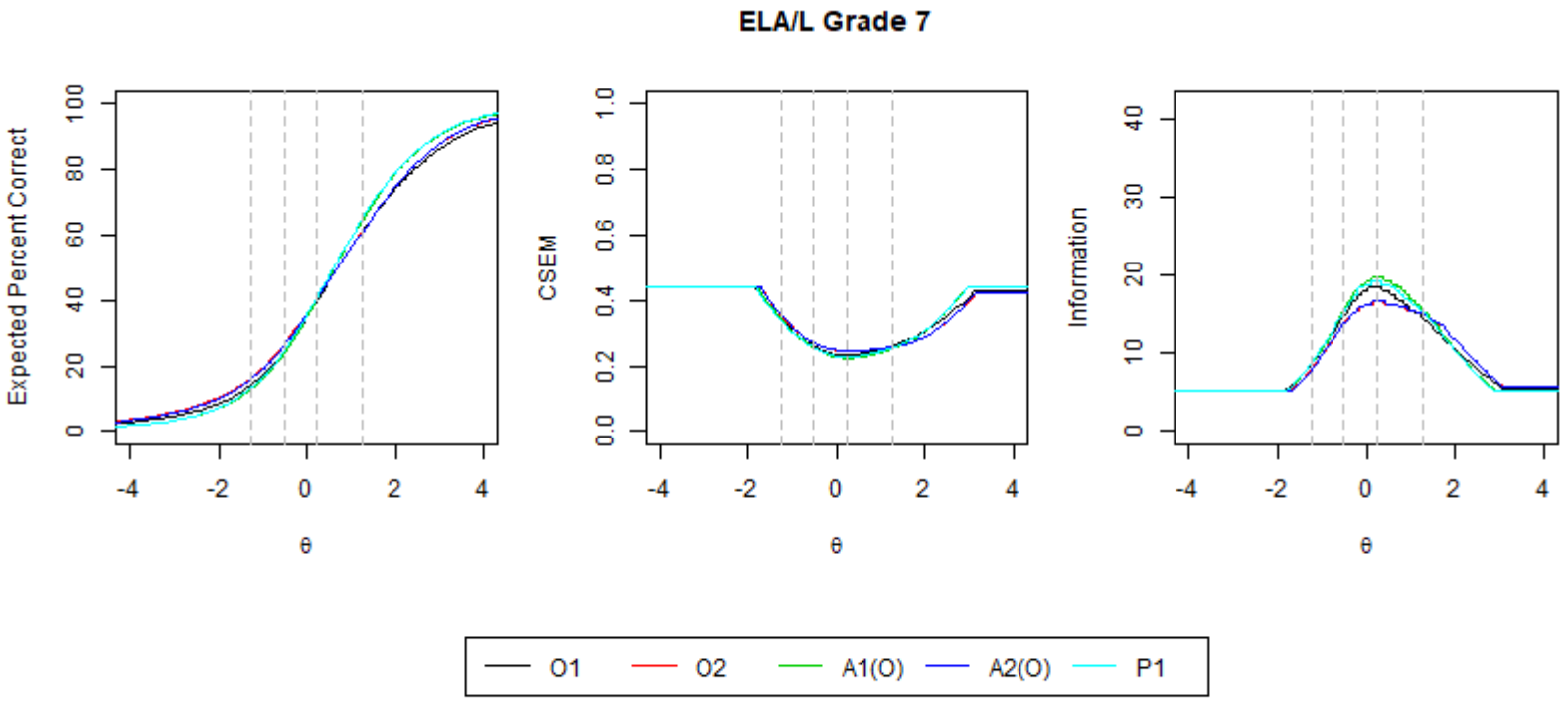

Figure A.12.5 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 7

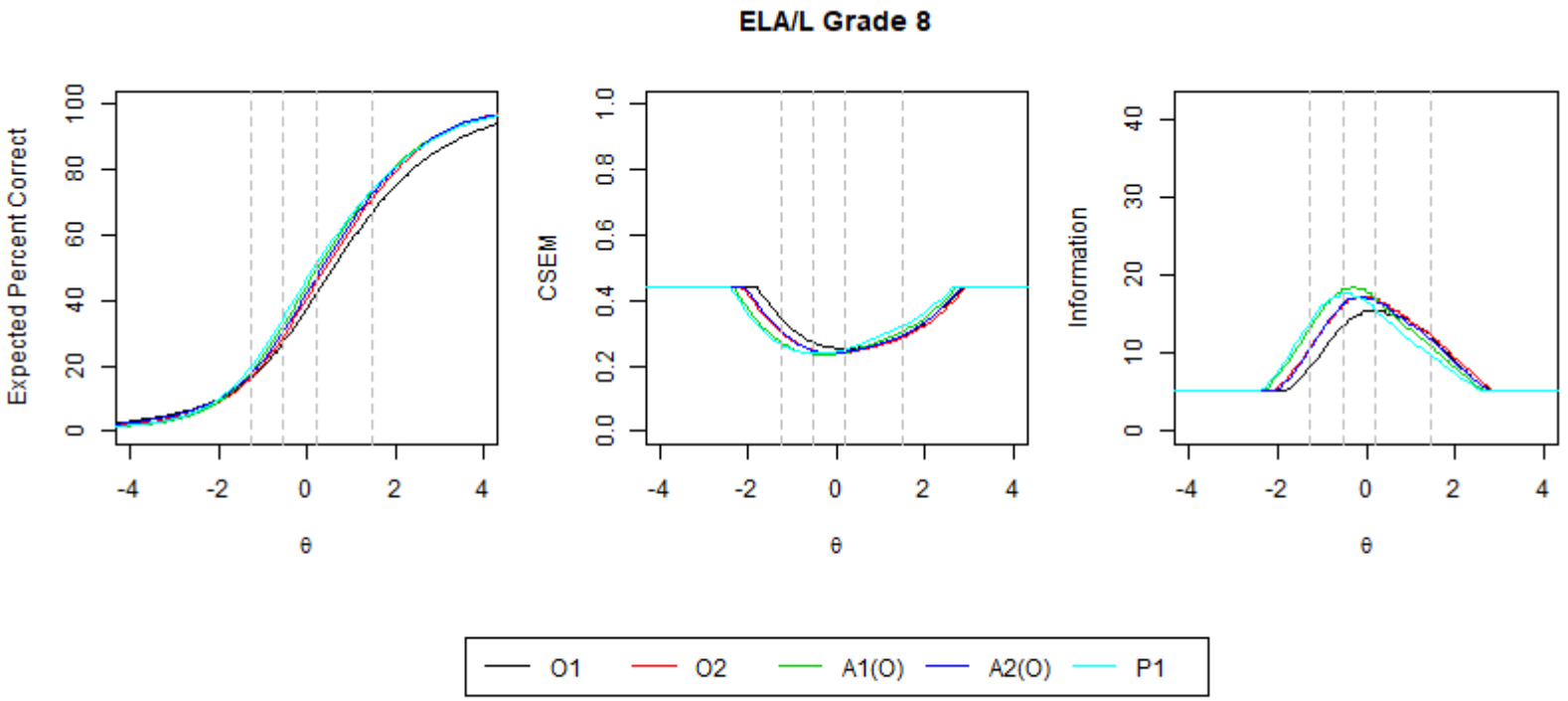

Figure A.12.6 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 8

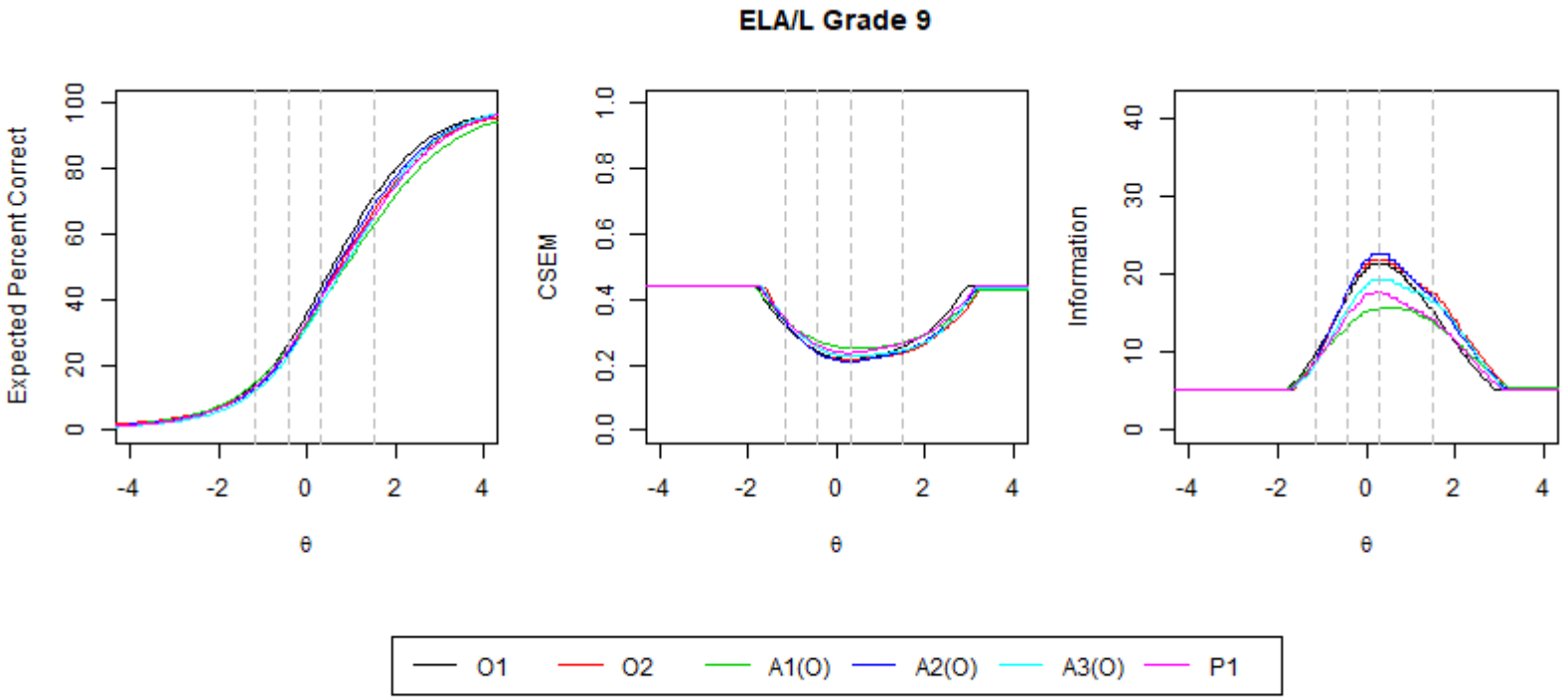

Figure A.12.7 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 9

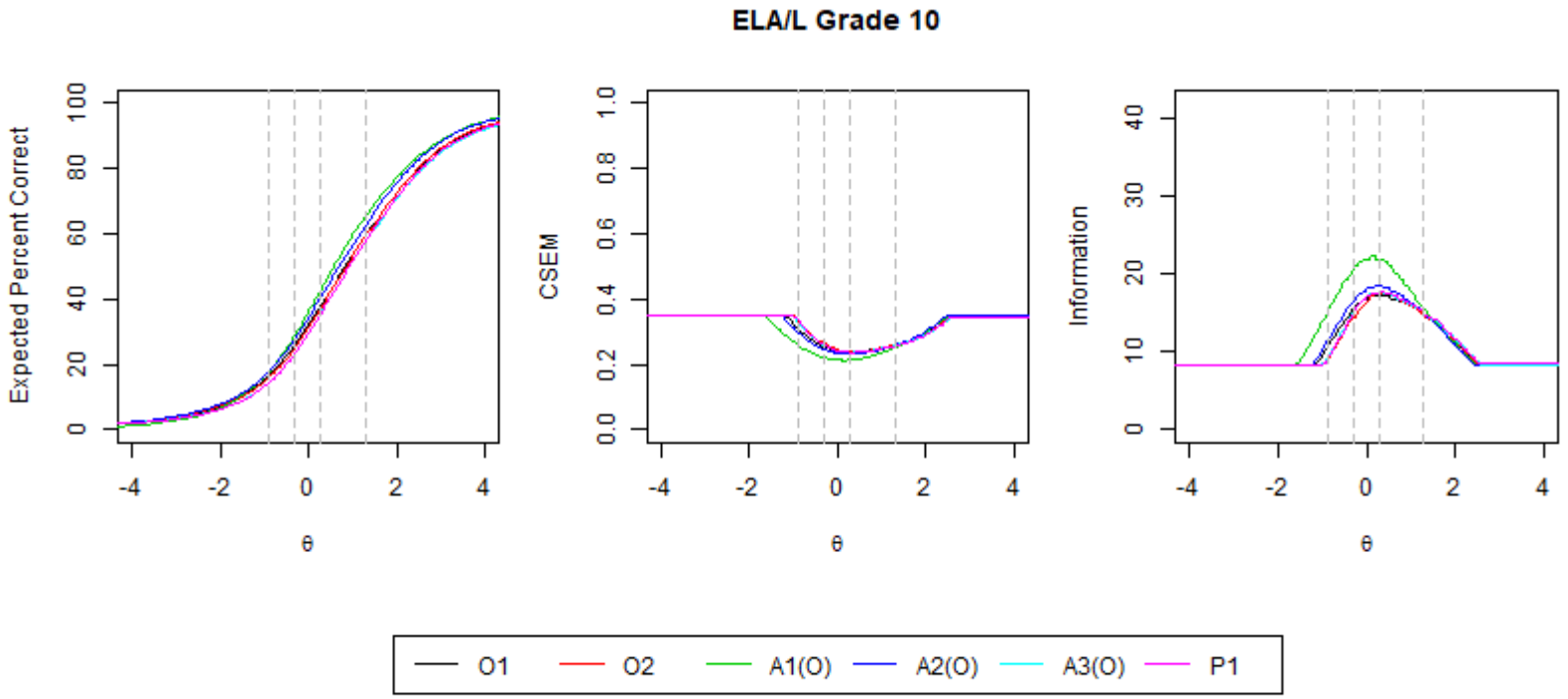

Figure A.12.8 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 10

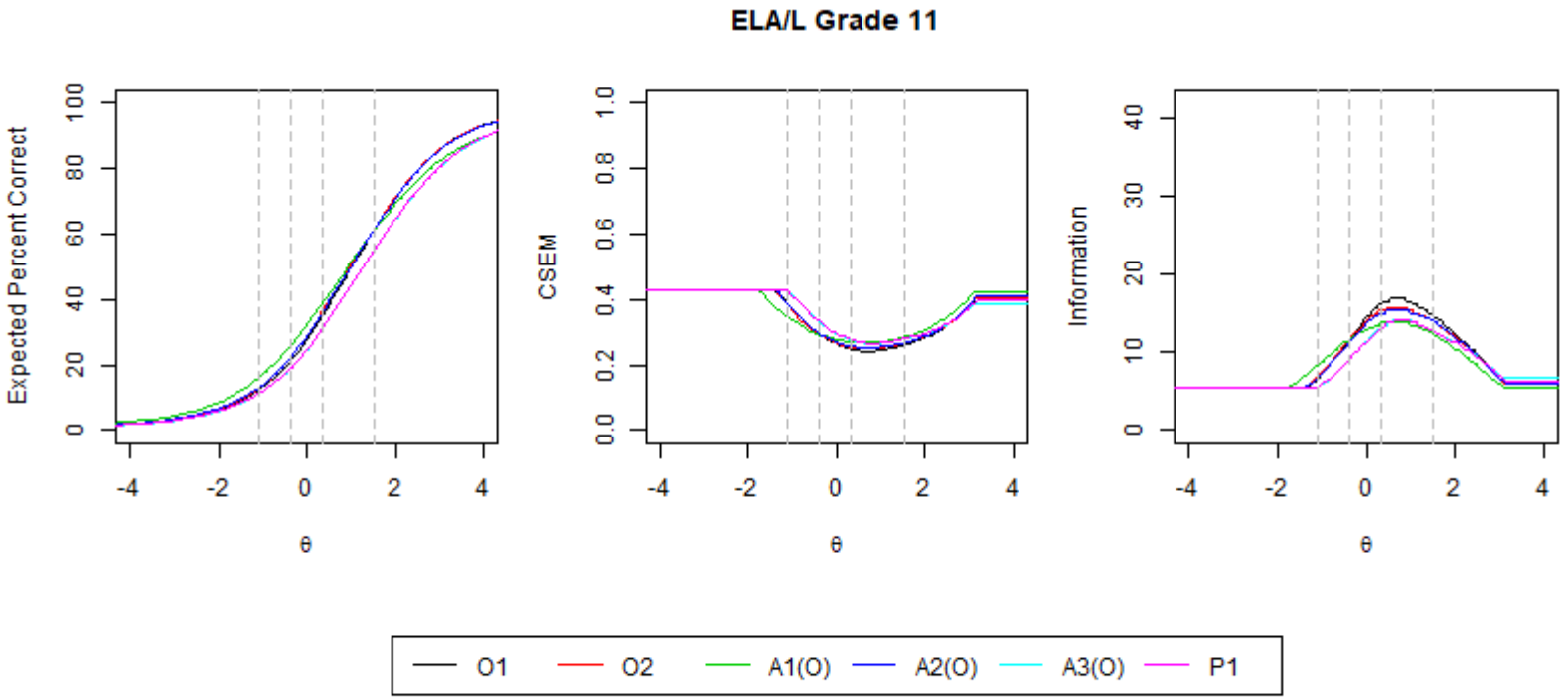

Figure A.12.9 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 11

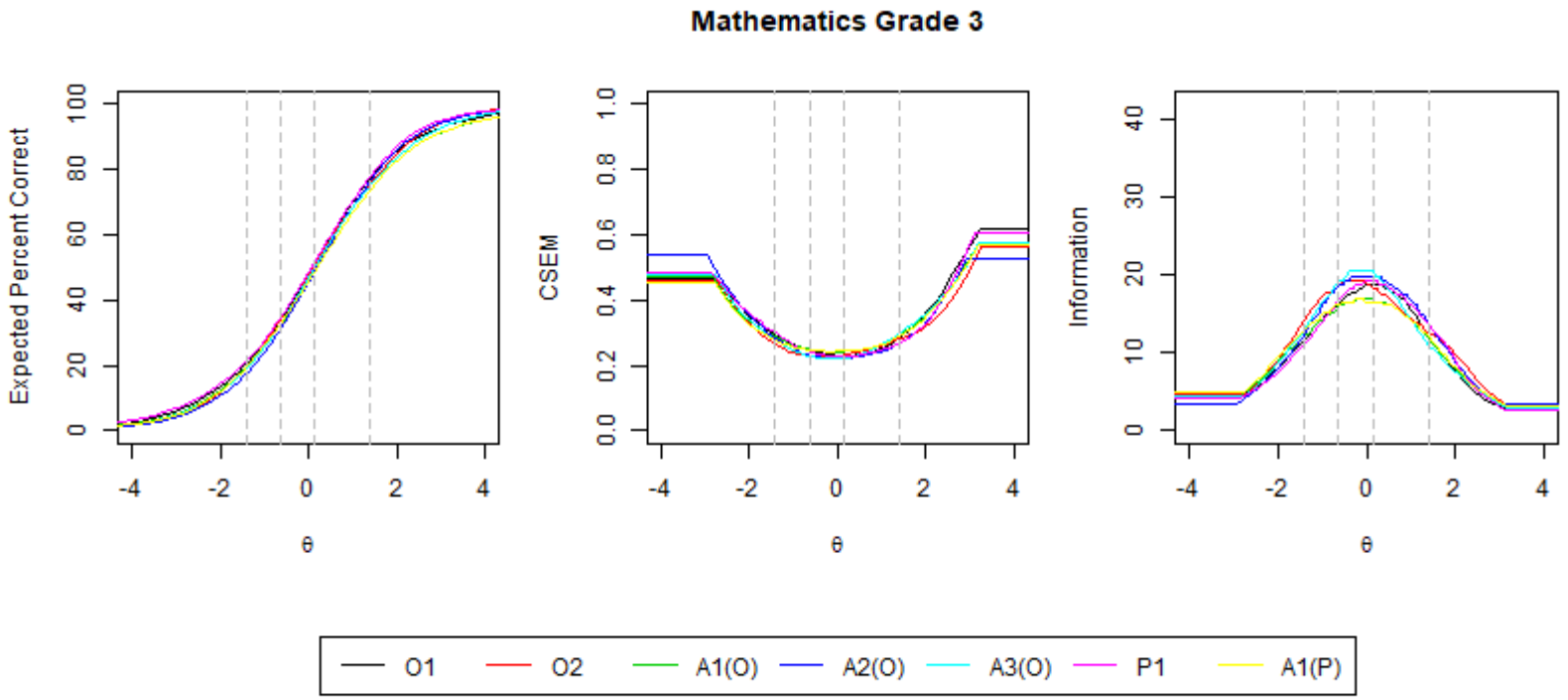

Figure A.12.10 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 3

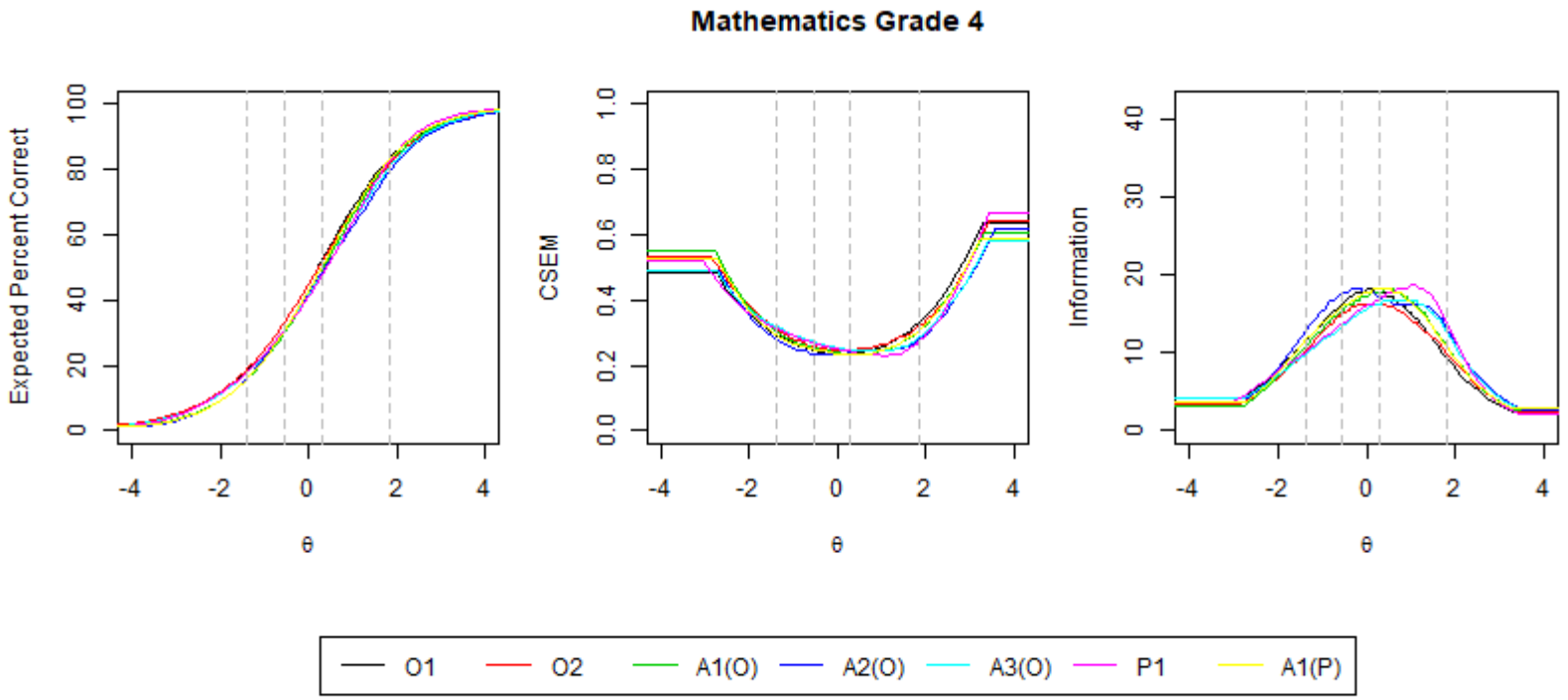

Figure A.12.11 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 4

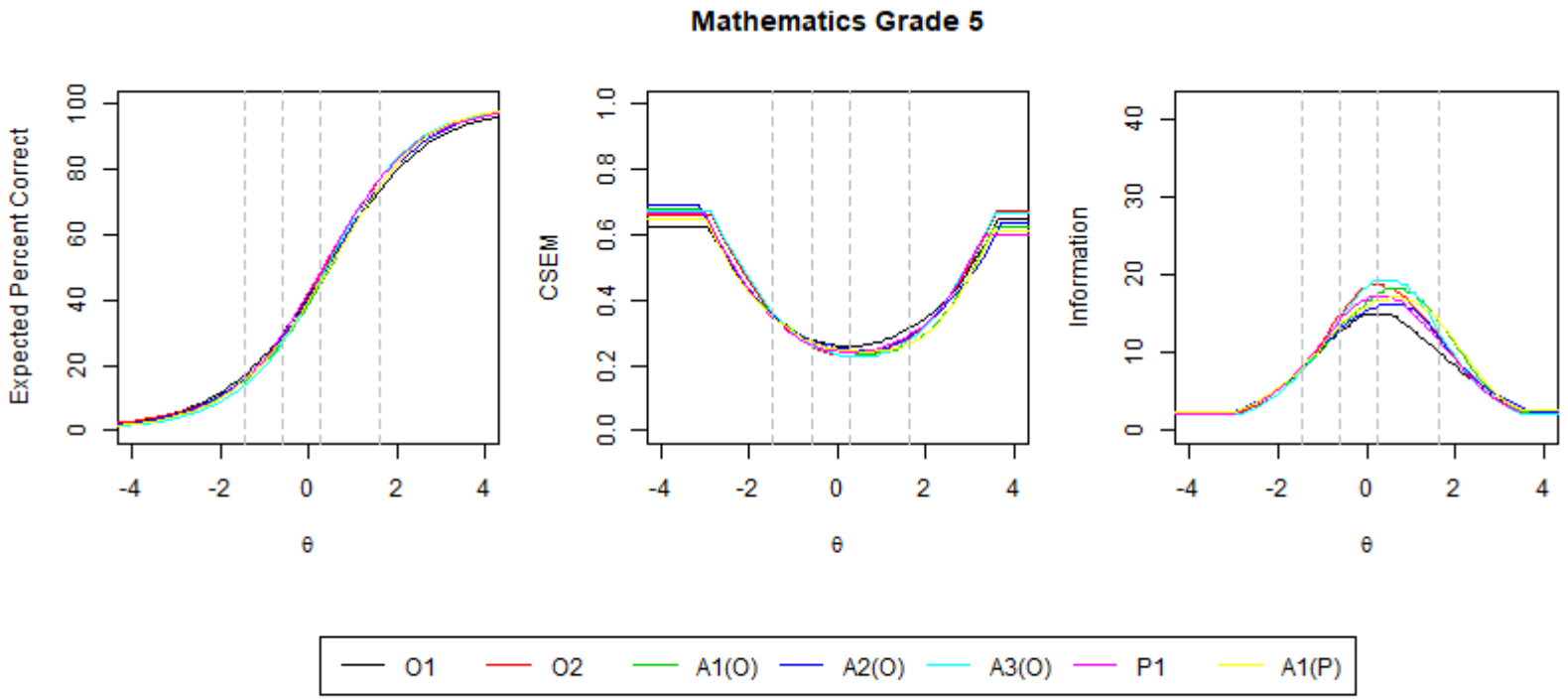

Figure A.12.12 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 5

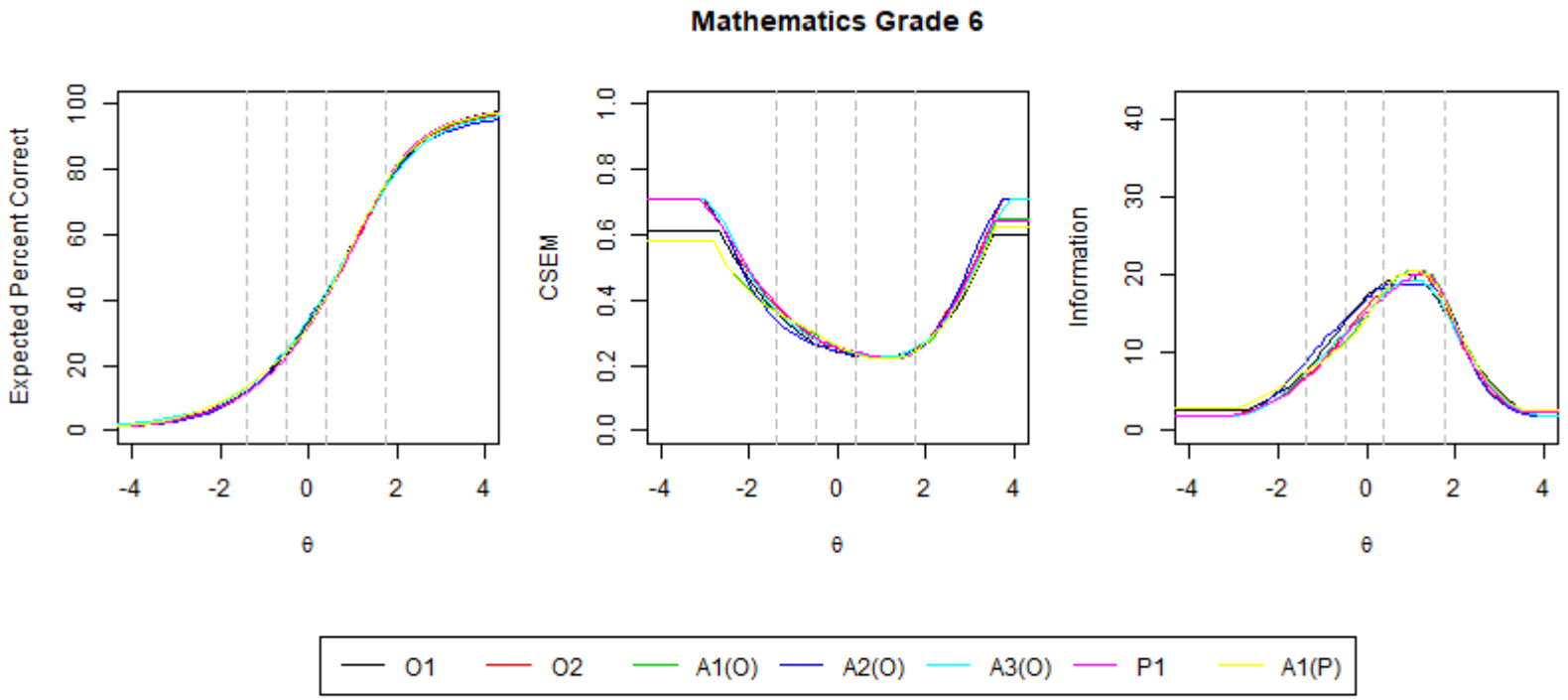

Figure A.12.13 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 6

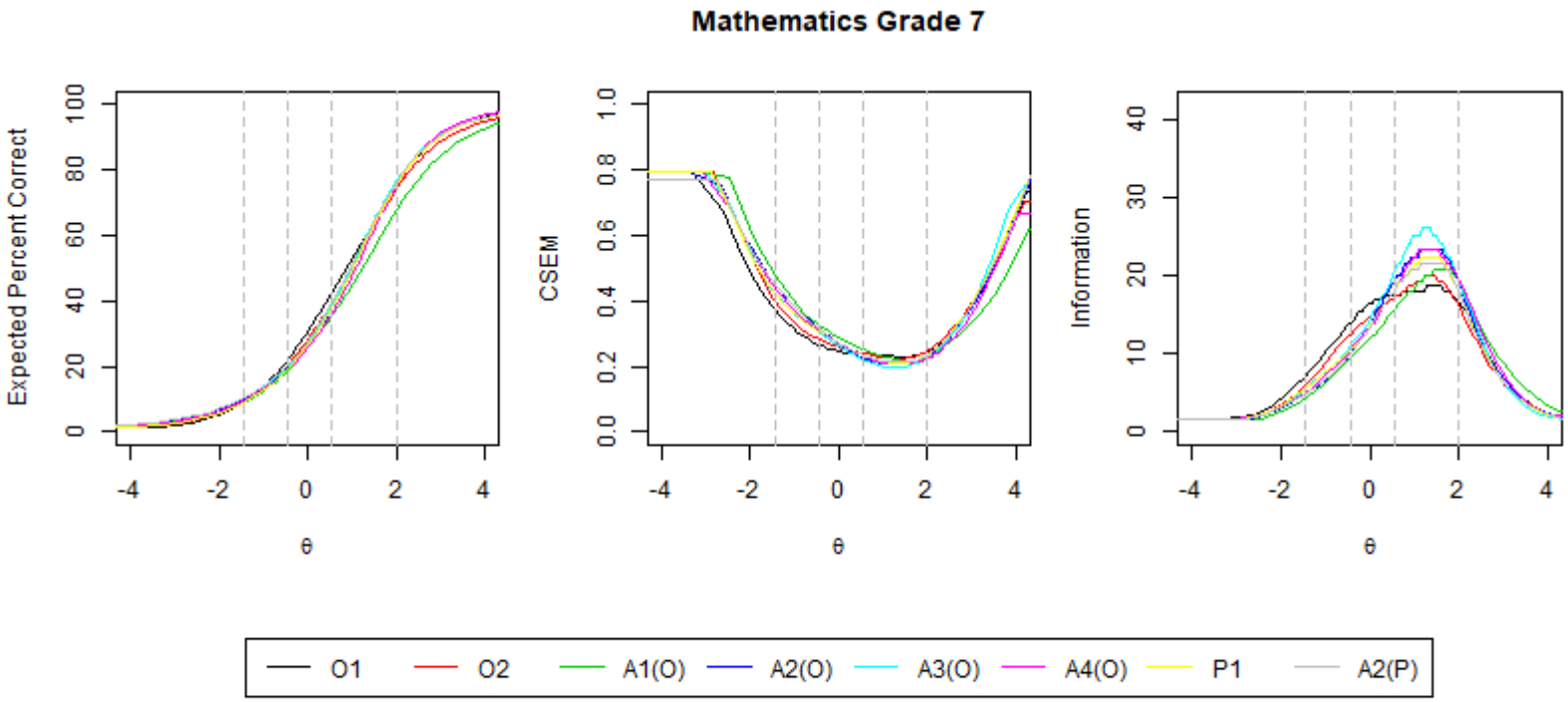

Figure A.12.14 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 7

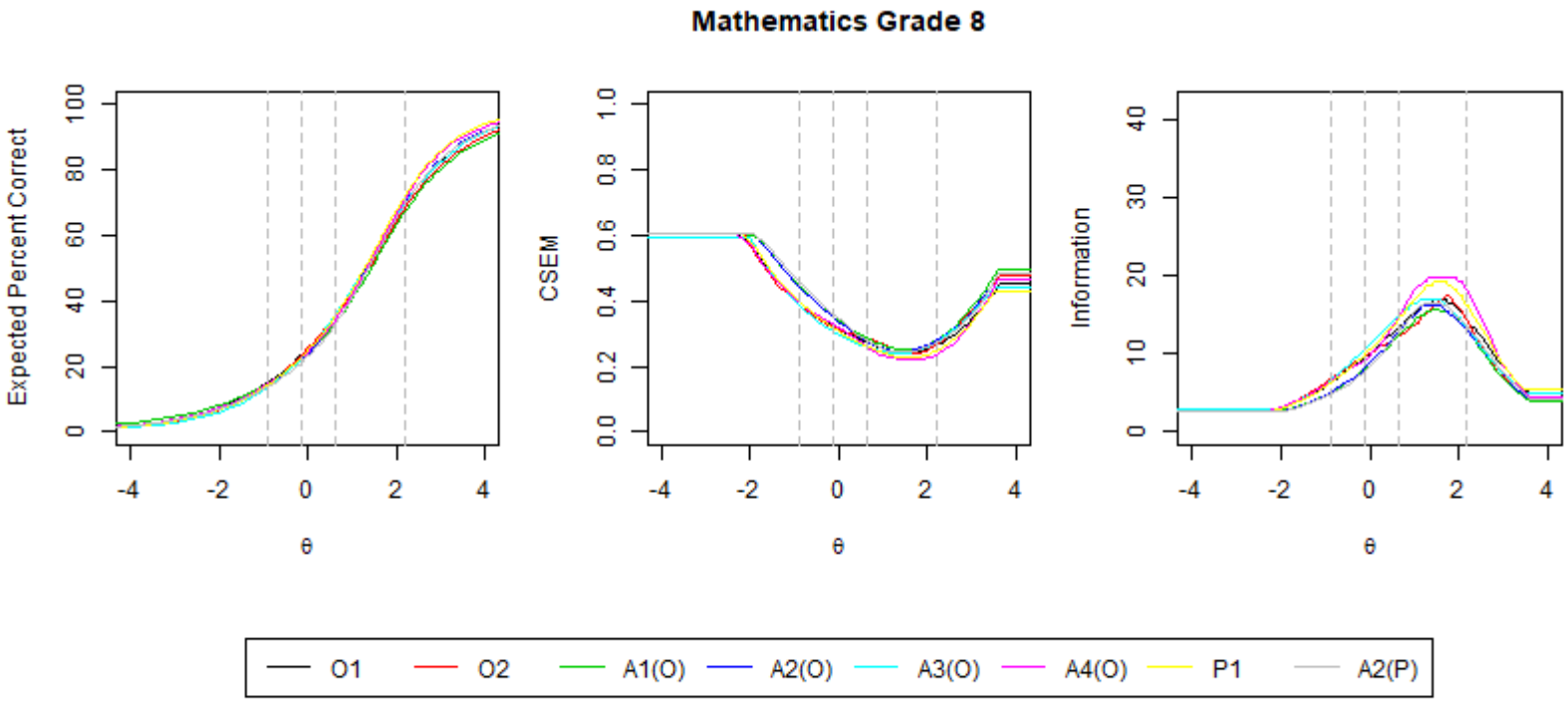

Figure A.12.15 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 8

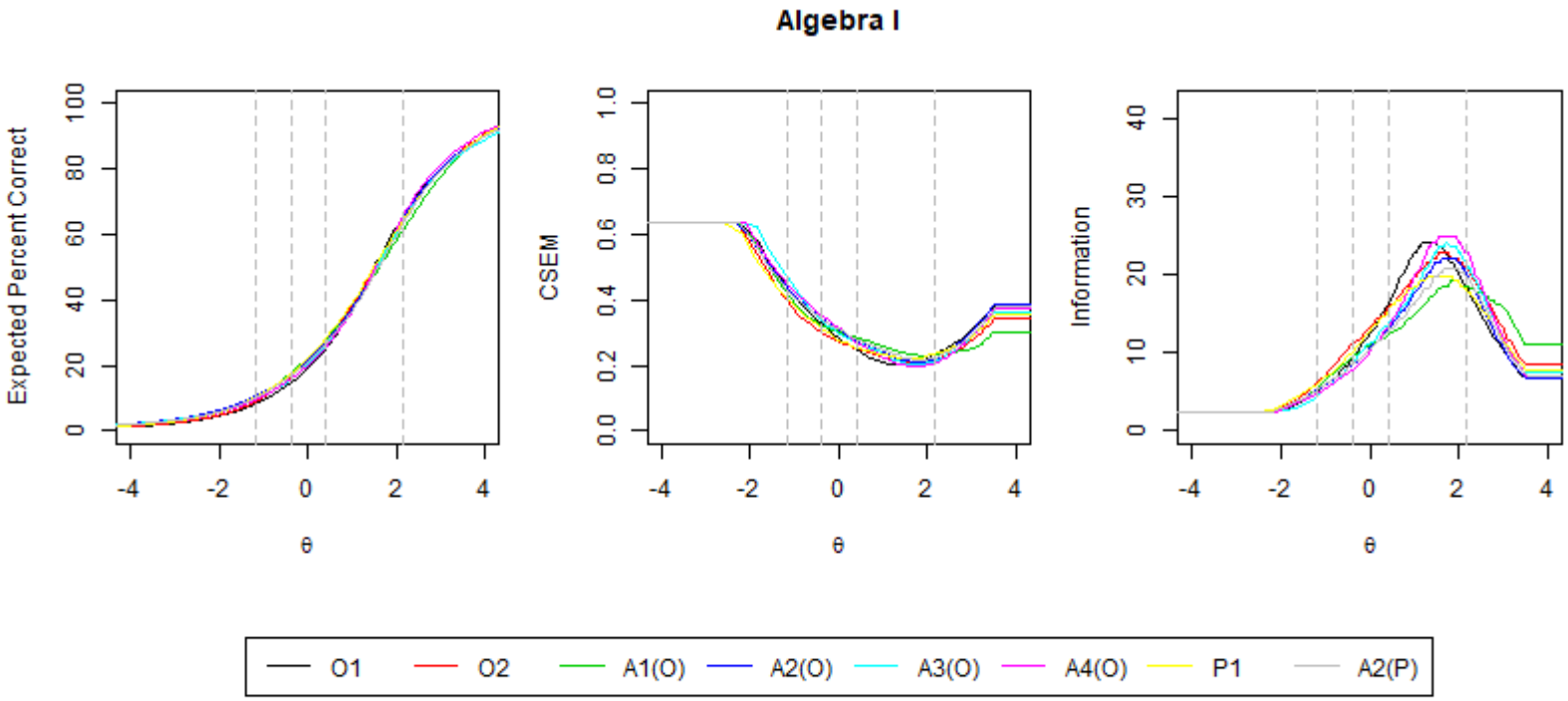

Figure A.12.16 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Algebra I

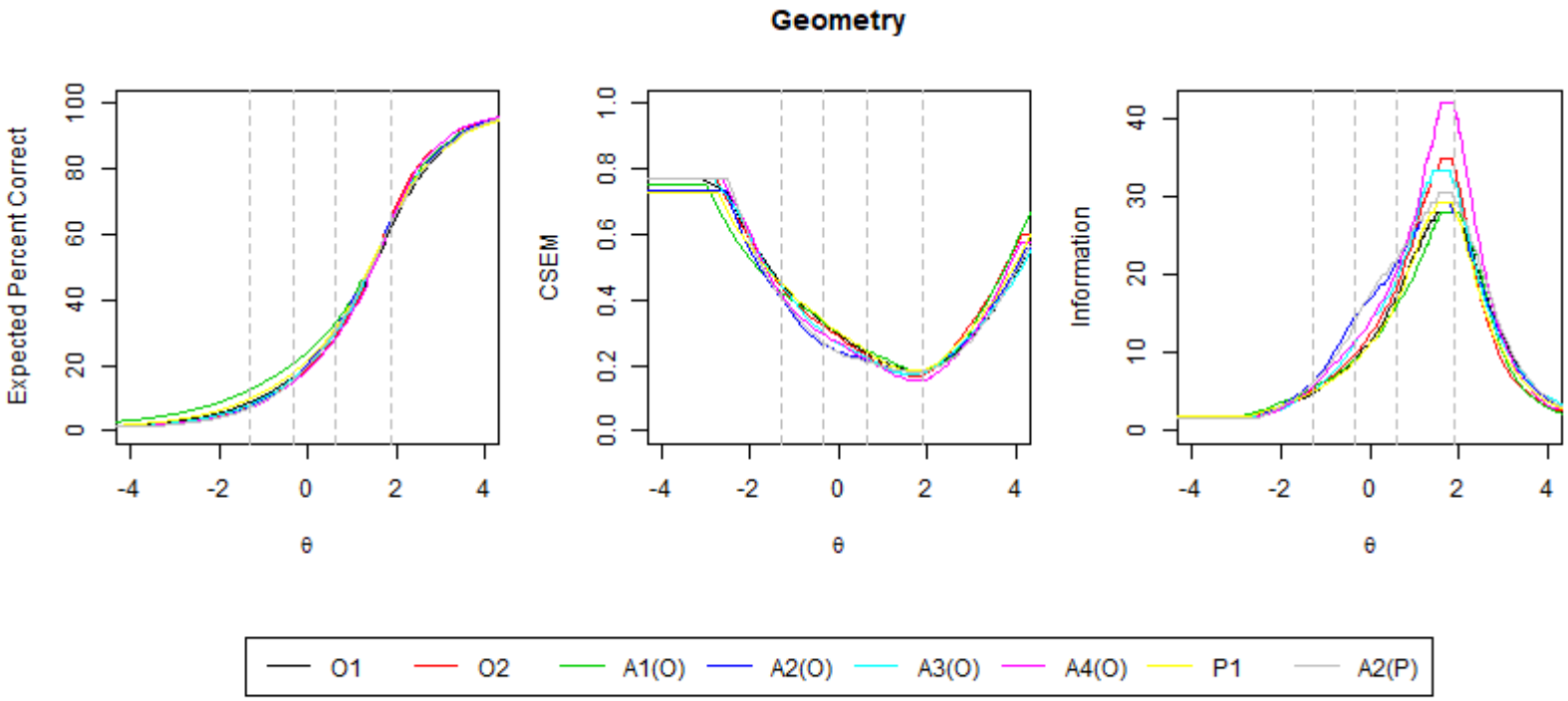

Figure A.12.17 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Geometry

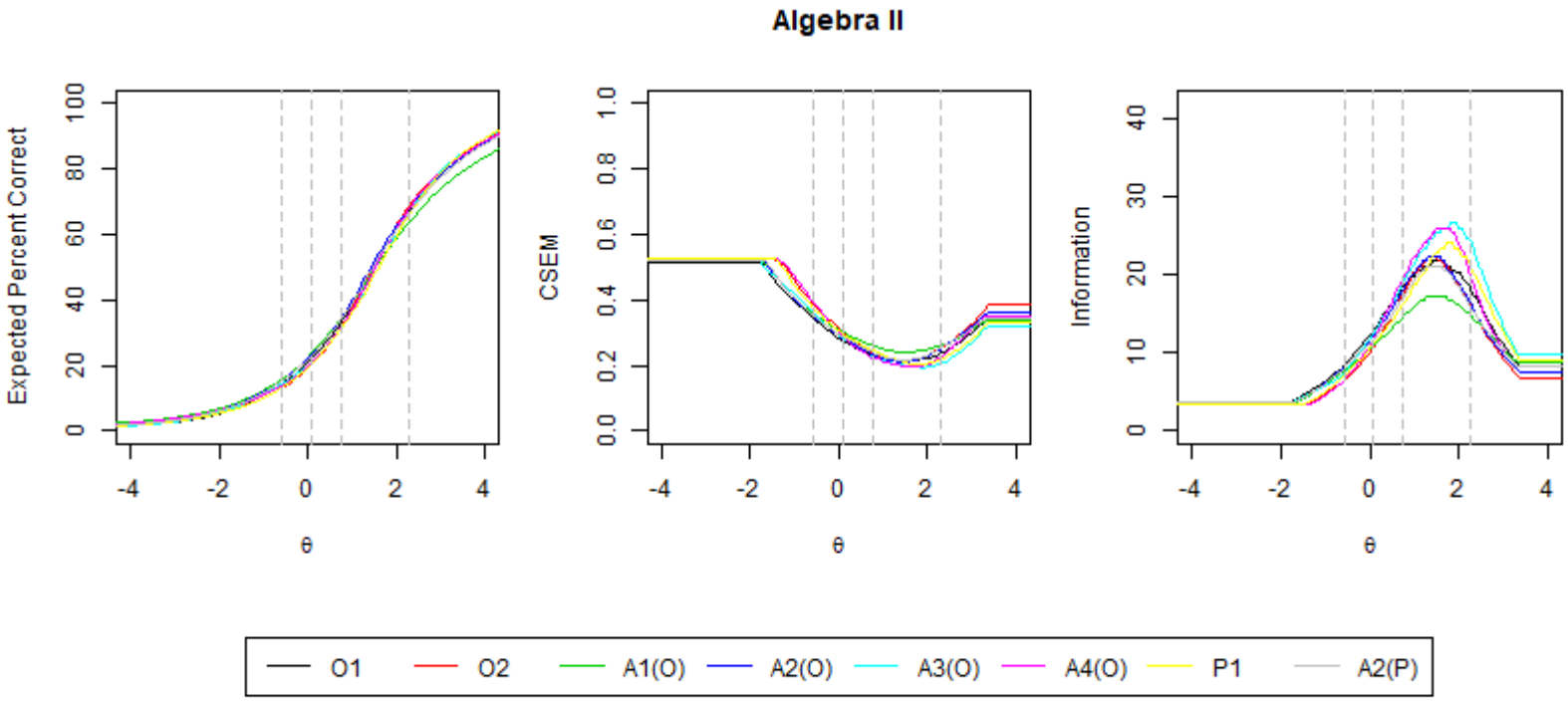

Figure A.12.18 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Algebra II

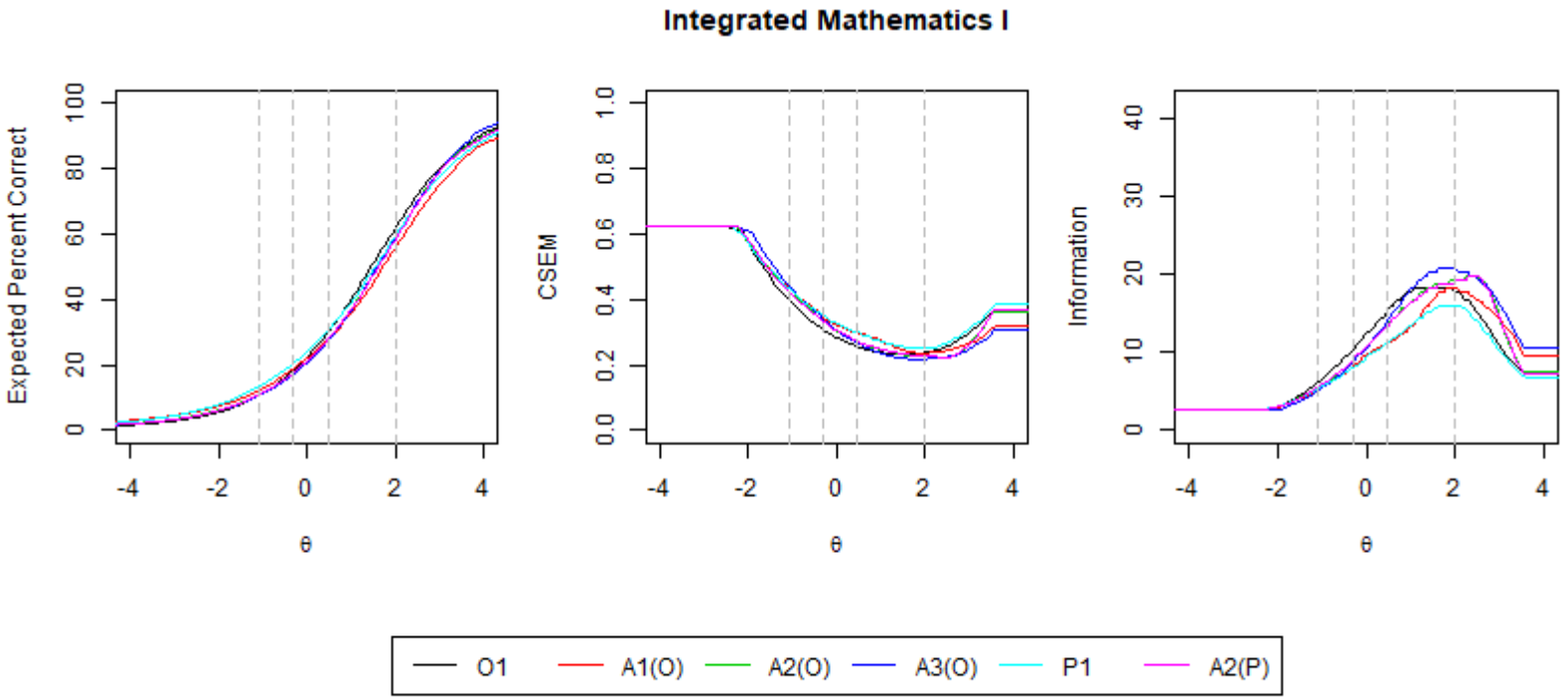

Figure A.12.19 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics I

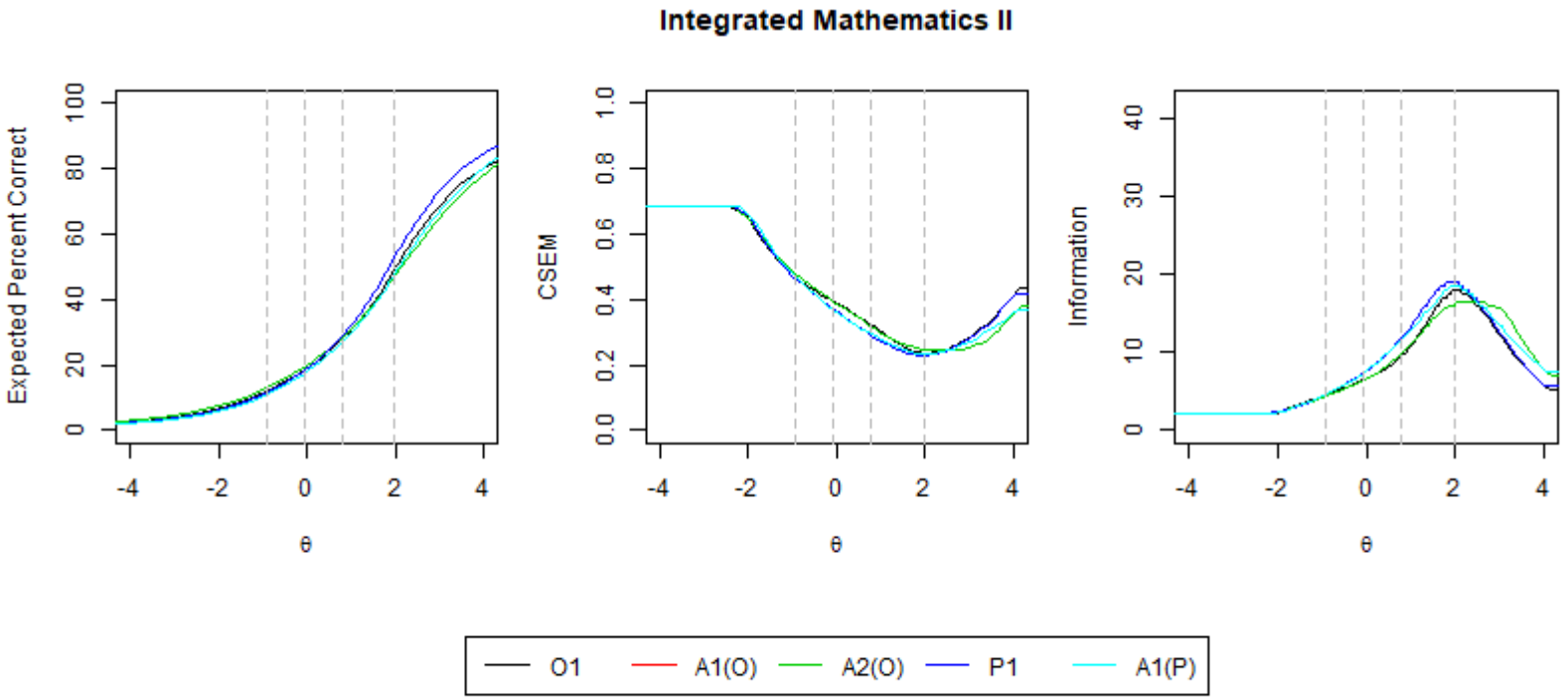

Figure A.12.20 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics II

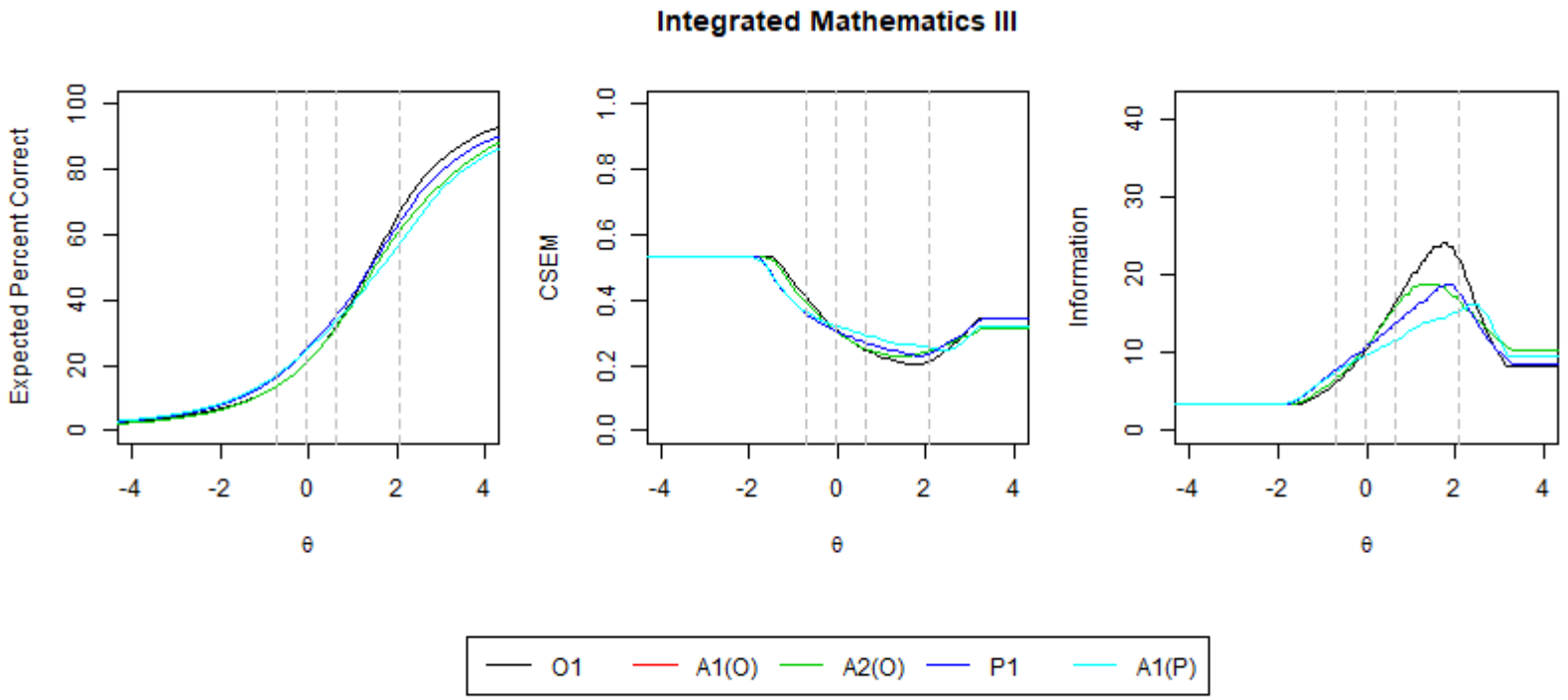

Figure A.12.21 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics III

Appendix 12.4: Subgroup Scale Score Performance

Table A.12.27 Subgroup Performance for ELA/L Scale Scores: Grade 3

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**338,927**	**738.80**	**42.68**	**650**	**850**
Gender	Female	166,189	744.26	43.00	650	850
	Male	172,738	733.55	41.70	650	850
Ethnicity	American Indian/Alaska Native	4,941	718.39	35.86	650	850
	Asian	22,383	769.37	40.35	650	850
	Black or African American	67,641	723.07	40.17	650	850
	Hispanic/Latino	94,348	726.78	40.31	650	850
	Native Hawaiian or Pacific Islander	578	748.27	45.12	650	850
	Multiple Race Selected	11,490	744.74	43.10	650	850
	White	137,498	750.01	39.86	650	850
Economic Status*	Economically Disadvantaged	168,229	723.74	39.45	650	850
	Not Economically Disadvantaged	170,629	753.66	40.47	650	850
English Learner Status	English Learner (EL)	50,459	712.86	35.69	650	850
	Non English Learner	288,272	743.36	42.17	650	850
Disabilities	Students with Disabilities (SWD)	53,040	709.57	40.25	650	850
	Students without Disabilities	284,826	744.31	40.87	650	850
Reading Score		**338,927**	**45.45**	**17.31**	**10**	**90**
Gender	Female	166,189	46.95	17.42	10	90
	Male	172,738	44.00	17.08	10	90
Ethnicity	American Indian/Alaska Native	4,941	36.51	14.37	10	90
	Asian	22,383	56.75	16.54	10	90
	Black or African American	67,641	39.04	15.79	10	90
	Hispanic/Latino	94,348	40.01	15.96	10	90
	Native Hawaiian or Pacific Islander	578	48.48	17.43	10	90
	Multiple Race Selected	11,490	48.49	17.57	10	90
	White	137,498	50.55	16.47	10	90
Economic Status*	Economically Disadvantaged	168,229	39.08	15.59	10	90
	Not Economically Disadvantaged	170,629	51.73	16.61	10	90
English Learner Status	English Learner (EL)	50,459	34.16	13.68	10	90
	Non English Learner	288,272	47.43	17.12	10	90
Disabilities	Students with Disabilities (SWD)	53,040	34.62	16.38	10	90
	Students without Disabilities	284,826	47.49	16.72	10	90
Writing Score		**338,927**	**31.03**	**12.28**	**10**	**60**
Gender	Female	166,189	33.02	11.92	10	60
	Male	172,738	29.12	12.32	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	4,941	27.15	11.41	10	60
	Asian	22,383	38.92	10.39	10	60
	Black or African American	67,641	27.23	12.36	10	60
	Hispanic/Latino	94,348	28.74	12.21	10	60
	Native Hawaiian or Pacific Islander	578	33.48	12.94	10	60
	Multiple Race Selected	11,490	31.82	12.30	10	60
	White	137,498	33.26	11.49	10	60
Economic Status*	Economically Disadvantaged	168,229	27.68	12.19	10	60
	Not Economically Disadvantaged	170,629	34.34	11.44	10	60
English Learner Status	English Learner (EL)	50,459	25.74	11.81	10	60
	Non English Learner	288,272	31.97	12.12	10	60
Disabilities	Students with Disabilities (SWD)	53,040	22.62	12.31	10	60
	Students without Disabilities	284,826	32.62	11.61	10	60

Note: This table is identical to Table 12.7 in Section 12.
*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.28 Subgroup Performance for ELA/L Scale Scores: Grade 4

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**345,483**	**743.92**	**37.34**	**650**	**850**
Gender	Female	169,612	749.13	37.24	650	850
	Male	175,871	738.89	36.74	650	850
Ethnicity	American Indian/Alaska Native	5,240	723.25	31.22	650	850
	Asian	22,872	773.22	36.50	650	850
	Black or African American	68,500	728.92	34.54	650	850
	Hispanic/Latino	95,760	733.67	34.27	650	850
	Native Hawaiian or Pacific Islander	583	751.40	39.20	650	850
	Multiple Race Selected	11,377	748.24	37.79	650	850
	White	141,110	753.80	35.05	650	850
Economic Status*	Economically Disadvantaged	171,130	729.88	33.51	650	850
	Not Economically Disadvantaged	174,288	757.72	35.73	650	850
English Learner Status	English Learner (EL)	47,174	719.44	29.37	650	848
	Non English Learner	298,091	747.81	36.99	650	850
Disabilities	Students with Disabilities (SWD)	57,713	716.82	35.13	650	850
	Students without Disabilities	286,558	749.44	35.33	650	850
Reading Score		**345,483**	**47.48**	**15.04**	**10**	**90**
Gender	Female	169,612	48.68	14.96	10	90
	Male	175,871	46.31	15.02	10	90
Ethnicity	American Indian/Alaska Native	5,240	38.65	12.35	10	90
	Asian	22,872	58.71	14.98	10	90
	Black or African American	68,500	41.56	13.60	10	90
	Hispanic/Latino	95,760	42.95	13.45	10	90
	Native Hawaiian or Pacific Islander	583	49.71	15.47	10	90
	Multiple Race Selected	11,377	49.78	15.35	10	90
	White	141,110	51.74	14.37	10	90
Economic Status*	Economically Disadvantaged	171,130	41.69	13.15	10	90
	Not Economically Disadvantaged	174,288	53.17	14.60	10	90
English Learner Status	English Learner (EL)	47,174	37.22	11.19	10	88
	Non English Learner	298,091	49.11	14.93	10	90
Disabilities	Students with Disabilities (SWD)	57,713	37.65	14.19	10	90
	Students without Disabilities	286,558	49.48	14.42	10	90
Writing Score		**345,483**	**32.45**	**10.95**	**10**	**60**
Gender	Female	169,612	34.61	10.31	10	60
	Male	175,871	30.36	11.14	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	5,240	28.10	10.49	10	60
	Asian	22,872	39.73	9.27	10	60
	Black or African American	68,500	28.61	11.14	10	60
	Hispanic/Latino	95,760	30.38	10.88	10	60
	Native Hawaiian or Pacific Islander	583	34.72	11.29	10	60
	Multiple Race Selected	11,377	32.93	10.92	10	60
	White	141,110	34.65	10.02	10	60
Economic Status*	Economically Disadvantaged	171,130	29.17	10.92	10	60
	Not Economically Disadvantaged	174,288	35.67	9.98	10	60
English Learner Status	English Learner (EL)	47,174	26.95	10.59	10	58
	Non English Learner	298,091	33.32	10.75	10	60
Disabilities	Students with Disabilities (SWD)	57,713	24.18	11.59	10	60
	Students without Disabilities	286,558	34.13	10.01	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.29 Subgroup Performance for ELA/L Scale Scores: Grade 5

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**348,524**	**742.50**	**35.32**	**650**	**850**
Gender	Female	170,846	748.68	35.14	650	850
	Male	177,678	736.55	34.46	650	850
Ethnicity	American Indian/Alaska Native	5,027	724.38	30.49	650	850
	Asian	22,415	769.88	33.63	650	850
	Black or African American	68,022	727.21	33.01	650	850
	Hispanic/Latino	97,632	732.33	32.72	650	850
	Native Hawaiian or Pacific Islander	613	753.10	34.75	650	844
	Multiple Race Selected	10,874	747.22	34.82	650	850
	White	143,921	752.59	32.57	650	850
Economic Status*	Economically Disadvantaged	171,057	728.73	32.20	650	850
	Not Economically Disadvantaged	177,432	755.77	33.04	650	850
English Learner Status	English Learner (EL)	29,756	707.49	25.43	650	850
	Non English Learner	318,575	745.78	34.32	650	850
Disabilities	Students with Disabilities (SWD)	59,471	715.65	33.24	650	850
	Students without Disabilities	287,932	748.10	33.12	650	850
Reading Score		**348,524**	**46.99**	**14.34**	**10**	**90**
Gender	Female	170,846	48.59	14.29	10	90
	Male	177,678	45.44	14.21	10	90
Ethnicity	American Indian/Alaska Native	5,027	38.65	12.26	10	90
	Asian	22,415	57.45	13.84	10	90
	Black or African American	68,022	41.09	13.31	10	90
	Hispanic/Latino	97,632	42.55	13.09	10	90
	Native Hawaiian or Pacific Islander	613	50.57	13.70	10	85
	Multiple Race Selected	10,874	49.45	14.34	10	90
	White	143,921	51.24	13.36	10	90
Economic Status*	Economically Disadvantaged	171,057	41.36	12.90	10	90
	Not Economically Disadvantaged	177,432	52.41	13.54	10	90
English Learner Status	English Learner (EL)	29,756	32.70	9.81	10	90
	Non English Learner	318,575	48.33	13.96	10	90
Disabilities	Students with Disabilities (SWD)	59,471	36.97	13.53	10	90
	Students without Disabilities	287,932	49.07	13.61	10	90
Writing Score		**348,524**	**31.54**	**11.16**	**10**	**60**
Gender	Female	170,846	34.09	10.23	10	60
	Male	177,678	29.09	11.47	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	5,027	28.14	10.70	10	60
	Asian	22,415	38.79	9.08	10	60
	Black or African American	68,022	27.20	11.45	10	60
	Hispanic/Latino	97,632	29.29	11.10	10	60
	Native Hawaiian or Pacific Islander	613	34.50	10.97	10	60
	Multiple Race Selected	10,874	32.30	10.91	10	60
	White	143,921	34.04	10.13	10	60
Economic Status*	Economically Disadvantaged	171,057	28.02	11.24	10	60
	Not Economically Disadvantaged	177,432	34.94	9.97	10	60
English Learner Status	English Learner (EL)	29,756	22.35	10.69	10	60
	Non English Learner	318,575	32.40	10.81	10	60
Disabilities	Students with Disabilities (SWD)	59,471	23.01	11.80	10	60
	Students without Disabilities	287,932	33.32	10.17	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.30 Subgroup Performance for ELA/L Scale Scores: Grade 6

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**344,520**	**741.96**	**33.38**	**650**	**850**
Gender	Female	169,028	748.09	33.17	650	850
	Male	175,492	736.06	32.51	650	850
Ethnicity	American Indian/Alaska Native	4,934	727.52	27.59	650	850
	Asian	22,424	769.00	33.56	650	850
	Black or African American	65,792	726.73	30.10	650	850
	Hispanic/Latino	94,427	732.48	30.18	650	850
	Native Hawaiian or Pacific Islander	638	749.61	34.09	663	850
	Multiple Race Selected	10,766	745.60	33.48	650	850
	White	145,411	751.03	31.36	650	850
Economic Status*	Economically Disadvantaged	162,719	728.79	29.61	650	850
	Not Economically Disadvantaged	176,999	753.94	32.25	650	850
English Learner Status	English Learner (EL)	20,967	706.79	22.39	650	837
	Non English Learner	323,236	744.26	32.69	650	850
Disabilities	Students with Disabilities (SWD)	58,100	716.76	29.59	650	850
	Students without Disabilities	285,248	747.15	31.73	650	850
Reading Score		**344,520**	**46.77**	**13.35**	**10**	**90**
Gender	Female	169,028	48.35	13.23	10	90
	Male	175,492	45.25	13.29	10	90
Ethnicity	American Indian/Alaska Native	4,934	39.85	10.87	10	90
	Asian	22,424	56.99	13.61	10	90
	Black or African American	65,792	40.83	11.77	10	90
	Hispanic/Latino	94,427	42.72	11.78	10	90
	Native Hawaiian or Pacific Islander	638	49.24	13.45	16	90
	Multiple Race Selected	10,766	48.93	13.67	10	90
	White	145,411	50.58	12.77	10	90
Economic Status*	Economically Disadvantaged	162,719	41.40	11.55	10	90
	Not Economically Disadvantaged	176,999	51.61	13.05	10	90
English Learner Status	English Learner (EL)	20,967	32.75	8.31	10	84
	Non English Learner	323,236	47.69	13.11	10	90
Disabilities	Students with Disabilities (SWD)	58,100	37.36	11.92	10	90
	Students without Disabilities	285,248	48.71	12.80	10	90
Writing Score		**344,520**	**30.59**	**11.93**	**10**	**60**
Gender	Female	169,028	33.45	10.90	10	60
	Male	175,492	27.83	12.23	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	4,934	28.40	10.90	10	60
	Asian	22,424	38.46	9.84	10	60
	Black or African American	65,792	25.90	12.07	10	60
	Hispanic/Latino	94,427	28.20	11.83	10	60
	Native Hawaiian or Pacific Islander	638	33.40	11.34	10	60
	Multiple Race Selected	10,766	30.95	11.82	10	60
	White	145,411	33.09	11.00	10	60
Economic Status*	Economically Disadvantaged	162,719	26.91	11.90	10	60
	Not Economically Disadvantaged	176,999	33.97	10.96	10	60
English Learner Status	English Learner (EL)	20,967	20.10	10.91	10	56
	Non English Learner	323,236	31.28	11.67	10	60
Disabilities	Students with Disabilities (SWD)	58,100	21.77	12.05	10	60
	Students without Disabilities	285,248	32.40	11.07	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.31 Subgroup Performance for ELA/L Scale Scores: Grade 7

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**338,731**	**745.09**	**40.18**	**650**	**850**
Gender	Female	165,429	753.66	39.02	650	850
	Male	173,302	736.92	39.56	650	850
Ethnicity	American Indian/Alaska Native	4,480	724.97	34.06	650	850
	Asian	22,488	777.45	38.23	650	850
	Black or African American	63,195	727.62	37.17	650	850
	Hispanic/Latino	91,718	733.37	37.41	650	850
	Native Hawaiian or Pacific Islander	656	759.50	36.68	650	850
	Multiple Race Selected	9,964	748.84	39.69	650	850
	White	146,095	755.33	37.32	650	850
Economic Status*	Economically Disadvantaged	154,717	729.00	36.63	650	850
	Not Economically Disadvantaged	179,649	758.83	38.14	650	850
English Learner Status	English Learner (EL)	18,912	699.63	29.46	650	839
	Non English Learner	319,563	747.80	39.10	650	850
Disabilities	Students with Disabilities (SWD)	57,248	713.57	37.52	650	850
	Students without Disabilities	280,353	751.60	37.57	650	850
Reading Score		**338,731**	**48.02**	**16.44**	**10**	**90**
Gender	Female	165,429	50.46	16.04	10	90
	Male	173,302	45.69	16.47	10	90
Ethnicity	American Indian/Alaska Native	4,480	38.84	14.03	10	89
	Asian	22,488	60.22	15.67	10	90
	Black or African American	63,195	41.15	15.18	10	90
	Hispanic/Latino	91,718	43.17	15.17	10	90
	Native Hawaiian or Pacific Islander	656	53.04	15.02	10	90
	Multiple Race Selected	9,964	50.37	16.53	10	90
	White	146,095	52.26	15.48	10	90
Economic Status*	Economically Disadvantaged	154,717	41.50	14.92	10	90
	Not Economically Disadvantaged	179,649	53.52	15.68	10	90
English Learner Status	English Learner (EL)	18,912	29.61	11.62	10	86
	Non English Learner	319,563	49.12	16.03	10	90
Disabilities	Students with Disabilities (SWD)	57,248	35.75	15.41	10	90
	Students without Disabilities	280,353	50.55	15.48	10	90
Writing Score		**338,731**	**31.92**	**12.20**	**10**	**60**
Gender	Female	165,429	35.05	11.03	10	60
	Male	173,302	28.92	12.50	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	4,480	27.92	11.28	10	60
	Asian	22,488	40.57	9.98	10	60
	Black or African American	63,195	27.09	12.25	10	60
	Hispanic/Latino	91,718	28.99	12.09	10	60
	Native Hawaiian or Pacific Islander	656	36.54	10.31	10	60
	Multiple Race Selected	9,964	32.29	11.99	10	60
	White	146,095	34.59	11.13	10	60
Economic Status*	Economically Disadvantaged	154,717	27.71	12.09	10	60
	Not Economically Disadvantaged	179,649	35.54	11.13	10	60
English Learner Status	English Learner (EL)	18,912	19.70	10.84	10	60
	Non English Learner	319,563	32.65	11.88	10	60
Disabilities	Students with Disabilities (SWD)	57,248	22.74	12.42	10	60
	Students without Disabilities	280,353	33.81	11.26	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.32 Subgroup Performance for ELA/L Scale Scores: Grade 8

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**339,283**	**743.30**	**40.42**	**650**	**850**
Gender	Female	165,771	751.76	39.09	650	850
	Male	173,512	735.22	40.01	650	850
Ethnicity	American Indian/Alaska Native	4,393	726.60	32.34	650	850
	Asian	22,684	776.40	39.54	650	850
	Black or African American	62,997	725.35	36.47	650	850
	Hispanic/Latino	89,335	731.37	37.05	650	850
	Native Hawaiian or Pacific Islander	565	756.48	38.76	650	850
	Multiple Race Selected	9,685	745.07	40.20	650	850
	White	149,491	753.31	38.09	650	850
Economic Status*	Economically Disadvantaged	150,381	727.24	36.20	650	850
	Not Economically Disadvantaged	184,857	756.26	39.17	650	850
English Learner Status	English Learner (EL)	18,008	698.53	27.80	650	845
	Non English Learner	321,054	745.83	39.54	650	850
Disabilities	Students with Disabilities (SWD)	56,818	712.43	36.86	650	850
	Students without Disabilities	281,396	749.61	38.17	650	850
Reading Score		**339,283**	**46.94**	**16.19**	**10**	**90**
Gender	Female	165,771	49.02	15.82	10	90
	Male	173,512	44.96	16.29	10	90
Ethnicity	American Indian/Alaska Native	4,393	39.12	12.78	10	90
	Asian	22,684	59.21	16.28	10	90
	Black or African American	62,997	40.11	14.35	10	90
	Hispanic/Latino	89,335	41.93	14.58	10	90
	Native Hawaiian or Pacific Islander	565	51.55	15.09	10	90
	Multiple Race Selected	9,685	48.74	16.48	10	90
	White	149,491	51.05	15.53	10	90
Economic Status*	Economically Disadvantaged	150,381	40.49	14.21	10	90
	Not Economically Disadvantaged	184,857	52.09	15.88	10	90
English Learner Status	English Learner (EL)	18,008	29.39	10.21	10	90
	Non English Learner	321,054	47.93	15.90	10	90
Disabilities	Students with Disabilities (SWD)	56,818	35.50	14.65	10	90
	Students without Disabilities	281,396	49.28	15.49	10	90
Writing Score		**339,283**	**32.49**	**11.69**	**10**	**60**
Gender	Female	165,771	35.59	10.61	10	60
	Male	173,512	29.53	11.90	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	4,393	29.69	10.20	10	60
	Asian	22,684	40.98	10.14	10	60
	Black or African American	62,997	27.73	11.56	10	60
	Hispanic/Latino	89,335	29.80	11.39	10	60
	Native Hawaiian or Pacific Islander	565	36.33	10.64	10	60
	Multiple Race Selected	9,685	32.19	11.52	10	60
	White	149,491	34.91	10.84	10	60
Economic Status*	Economically Disadvantaged	150,381	28.56	11.43	10	60
	Not Economically Disadvantaged	184,857	35.69	10.96	10	60
English Learner Status	English Learner (EL)	18,008	20.86	10.44	10	60
	Non English Learner	321,054	33.15	11.41	10	60
Disabilities	Students with Disabilities (SWD)	56,818	23.75	11.94	10	60
	Students without Disabilities	281,396	34.28	10.80	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.33 Subgroup Performance for ELA/L Scale Scores: Grade 9

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**128,229**	**746.64**	**39.64**	**650**	**850**
Gender	Female	63,018	754.49	38.30	650	850
	Male	65,211	739.06	39.44	650	850
Ethnicity	American Indian/Alaska Native	3,151	725.14	30.21	650	833
	Asian	10,602	780.68	36.46	650	850
	Black or African American	17,883	731.61	36.13	650	850
	Hispanic/Latino	41,613	731.21	37.00	650	850
	Native Hawaiian or Pacific Islander	309	753.49	37.70	659	850
	Multiple Race Selected	1,753	757.91	39.13	650	850
	White	52,907	757.92	35.44	650	850
Economic Status*	Economically Disadvantaged	53,989	729.41	35.70	650	850
	Not Economically Disadvantaged	74,237	759.18	37.61	650	850
English Learner Status	English Learner (EL)	8,463	695.59	28.04	650	811
	Non English Learner	119,469	750.30	37.79	650	850
Disabilities	Students with Disabilities (SWD)	22,846	718.94	35.60	650	850
	Students without Disabilities	105,209	752.70	37.86	650	850
Reading Score		**128,229**	**48.87**	**16.08**	**10**	**90**
Gender	Female	63,018	50.99	15.69	10	90
	Male	65,211	46.82	16.19	10	90
Ethnicity	American Indian/Alaska Native	3,151	39.63	12.52	10	88
	Asian	10,602	61.52	15.10	10	90
	Black or African American	17,883	42.97	14.72	10	90
	Hispanic/Latino	41,613	42.69	14.99	10	90
	Native Hawaiian or Pacific Islander	309	50.78	15.25	10	90
	Multiple Race Selected	1,753	53.59	15.88	10	90
	White	52,907	53.57	14.50	10	90
Economic Status*	Economically Disadvantaged	53,989	41.96	14.50	10	90
	Not Economically Disadvantaged	74,237	53.89	15.29	10	90
English Learner Status	English Learner (EL)	8,463	28.30	11.01	10	79
	Non English Learner	119,469	50.34	15.36	10	90
Disabilities	Students with Disabilities (SWD)	22,846	38.36	14.66	10	90
	Students without Disabilities	105,209	51.17	15.45	10	90
Writing Score		**128,229**	**32.19**	**12.07**	**10**	**60**
Gender	Female	63,018	35.23	10.83	10	60
	Male	65,211	29.26	12.47	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	3,151	27.59	10.58	10	60
	Asian	10,602	41.39	9.53	10	60
	Black or African American	17,883	28.06	11.96	10	60
	Hispanic/Latino	41,613	28.15	12.03	10	60
	Native Hawaiian or Pacific Islander	309	34.67	11.10	10	60
	Multiple Race Selected	1,753	34.86	11.73	10	60
	White	52,907	35.09	10.73	10	60
Economic Status*	Economically Disadvantaged	53,989	27.64	11.89	10	60
	Not Economically Disadvantaged	74,237	35.50	11.07	10	60
English Learner Status	English Learner (EL)	8,463	18.34	10.37	10	52
	Non English Learner	119,469	33.19	11.56	10	60
Disabilities	Students with Disabilities (SWD)	22,846	23.71	12.21	10	60
	Students without Disabilities	105,209	34.05	11.21	10	60

Note: This table is identical to Table 12.8 in Section 12.
*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.34 Subgroup Performance for ELA/L Scale Scores: Grade 10

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**188,597**	**744.90**	**48.60**	**650**	**850**
Gender	Female	92,340	754.80	46.81	650	850
	Male	96,257	735.41	48.39	650	850
Ethnicity	American Indian/Alaska Native	3,285	725.13	39.77	650	850
	Asian	14,658	780.42	45.48	650	850
	Black or African American	40,543	725.75	44.02	650	850
	Hispanic/Latino	50,303	728.94	45.45	650	850
	Native Hawaiian or Pacific Islander	469	753.84	45.06	650	850
	Multiple Race Selected	4,138	755.67	46.78	650	850
	White	75,150	759.22	45.26	650	850
Economic Status*	Economically Disadvantaged	69,718	725.74	44.46	650	850
	Not Economically Disadvantaged	115,753	756.49	47.52	650	850
English Learner Status	English Learner (EL)	12,507	690.86	32.68	650	843
	Non English Learner	175,587	748.80	47.22	650	850
Disabilities	Students with Disabilities (SWD)	32,052	712.07	44.15	650	850
	Students without Disabilities	156,223	751.68	46.69	650	850
Reading Score		**188,597**	**48.27**	**19.87**	**10**	**90**
Gender	Female	92,340	50.93	19.31	10	90
	Male	96,257	45.72	20.07	10	90
Ethnicity	American Indian/Alaska Native	3,285	38.23	16.39	10	90
	Asian	14,658	61.65	19.15	10	90
	Black or African American	40,543	40.95	18.01	10	90
	Hispanic/Latino	50,303	41.43	18.17	10	90
	Native Hawaiian or Pacific Islander	469	51.38	19.02	10	90
	Multiple Race Selected	4,138	53.38	19.60	10	90
	White	75,150	54.33	18.72	10	90
Economic Status*	Economically Disadvantaged	69,718	40.13	17.79	10	90
	Not Economically Disadvantaged	115,753	53.11	19.51	10	90
English Learner Status	English Learner (EL)	12,507	26.26	12.48	10	90
	Non English Learner	175,587	49.86	19.36	10	90
Disabilities	Students with Disabilities (SWD)	32,052	35.85	18.32	10	90
	Students without Disabilities	156,223	50.84	19.21	10	90
Writing Score		**188,597**	**32.28**	**13.03**	**10**	**60**
Gender	Female	92,340	35.55	12.08	10	60
	Male	96,257	29.14	13.13	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	3,285	29.26	11.31	10	60
	Asian	14,658	41.11	11.25	10	60
	Black or African American	40,543	27.51	12.43	10	60
	Hispanic/Latino	50,303	28.81	12.77	10	60
	Native Hawaiian or Pacific Islander	469	34.98	11.47	10	60
	Multiple Race Selected	4,138	34.36	12.32	10	60
	White	75,150	35.46	12.16	10	60
Economic Status*	Economically Disadvantaged	69,718	28.04	12.66	10	60
	Not Economically Disadvantaged	115,753	34.90	12.60	10	60
English Learner Status	English Learner (EL)	12,507	19.49	10.64	10	60
	Non English Learner	175,587	33.20	12.70	10	60
Disabilities	Students with Disabilities (SWD)	32,052	23.79	12.50	10	60
	Students without Disabilities	156,223	34.03	12.44	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.35 Subgroup Performance for ELA/L Scale Scores: Grade 11

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**102,755**	**736.83**	**40.65**	**650**	**850**
Gender	Female	48,814	746.06	38.98	650	850
	Male	53,941	728.48	40.32	650	850
Ethnicity	American Indian/Alaska Native	3,124	736.17	33.91	650	850
	Asian	6,300	759.17	42.48	650	850
	Black or African American	19,900	727.57	38.39	650	850
	Hispanic/Latino	34,523	733.54	38.81	650	850
	Native Hawaiian or Pacific Islander	241	749.91	38.81	650	839
	Multiple Race Selected	1,019	738.14	41.20	650	850
	White	37,632	740.96	41.71	650	850
Economic Status*	Economically Disadvantaged	45,411	731.20	38.74	650	850
	Not Economically Disadvantaged	57,331	741.30	41.55	650	850
English Learner Status	English Learner (EL)	6,470	703.61	31.08	650	821
	Non English Learner	96,181	739.10	40.23	650	850
Disabilities	Students with Disabilities (SWD)	18,641	711.67	36.64	650	850
	Students without Disabilities	84,045	742.43	39.36	650	850
Reading Score		**102,755**	**45.14**	**15.99**	**10**	**90**
Gender	Female	48,814	47.82	15.39	10	90
	Male	53,941	42.72	16.13	10	90
Ethnicity	American Indian/Alaska Native	3,124	42.61	13.45	10	90
	Asian	6,300	53.72	16.83	10	90
	Black or African American	19,900	41.50	14.95	10	90
	Hispanic/Latino	34,523	43.34	15.04	10	90
	Native Hawaiian or Pacific Islander	241	49.75	15.22	10	89
	Multiple Race Selected	1,019	46.34	16.18	10	90
	White	37,632	47.43	16.51	10	90
Economic Status*	Economically Disadvantaged	45,411	42.34	14.91	10	90
	Not Economically Disadvantaged	57,331	47.36	16.45	10	90
English Learner Status	English Learner (EL)	6,470	31.13	11.39	10	81
	Non English Learner	96,181	46.10	15.80	10	90
Disabilities	Students with Disabilities (SWD)	18,641	35.57	14.40	10	90
	Students without Disabilities	84,045	47.27	15.53	10	90
Writing Score		**102,755**	**28.99**	**13.17**	**10**	**60**
Gender	Female	48,814	32.39	12.35	10	60
	Male	53,941	25.92	13.14	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	3,124	31.44	10.83	10	58
	Asian	6,300	34.96	12.97	10	60
	Black or African American	19,900	26.52	12.90	10	60
	Hispanic/Latino	34,523	28.73	12.76	10	60
	Native Hawaiian or Pacific Islander	241	32.83	12.72	10	56
	Multiple Race Selected	1,019	28.60	13.49	10	60
	White	37,632	29.33	13.51	10	60
Economic Status*	Economically Disadvantaged	45,411	28.12	12.87	10	60
	Not Economically Disadvantaged	57,331	29.69	13.36	10	60
English Learner Status	English Learner (EL)	6,470	21.37	11.49	10	56
	Non English Learner	96,181	29.52	13.12	10	60
Disabilities	Students with Disabilities (SWD)	18,641	21.82	12.40	10	60
	Students without Disabilities	84,045	30.59	12.80	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.36 Subgroup Performance for Mathematics Scale Scores: Grade 3

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**348,117**	**742.56**	**37.24**	**650**	**850**
Gender	Female	170,632	742.81	36.27	650	850
	Male	177,485	742.31	38.15	650	850
Ethnicity	American Indian/Alaska Native	4,971	724.10	30.77	650	850
	Asian	23,146	774.35	35.70	650	850
	Black or African American	68,498	725.50	34.68	650	850
	Hispanic/Latino	97,908	732.12	33.34	650	850
	Native Hawaiian or Pacific Islander	688	748.82	36.49	650	850
	Multiple Race Selected	12,281	747.03	38.10	650	850
	White	140,467	753.14	34.47	650	850
Economic Status*	Economically Disadvantaged	170,494	728.39	33.53	650	850
	Not Economically Disadvantaged	172,005	756.35	35.63	650	850
English Learner Status	English Learner (EL)	54,563	724.20	31.41	650	850
	Non English Learner	293,337	745.98	37.25	650	850
Disabilities	Students with Disabilities (SWD)	54,044	720.44	37.06	650	850
	Students without Disabilities	292,963	746.69	35.81	650	850
Language Form	Spanish	4,906	715.52	30.48	650	832

Note: This table is identical to Table 12.9 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.37 Subgroup Performance for Mathematics Scale Scores: Grade 4

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**354,080**	**738.12**	**33.85**	**650**	**850**
Gender	Female	173,910	738.45	33.23	650	850
	Male	180,170	737.80	34.43	650	850
Ethnicity	American Indian/Alaska Native	5,261	720.62	28.45	650	850
	Asian	23,664	768.44	33.10	650	850
	Black or African American	69,253	721.71	30.41	650	850
	Hispanic/Latino	98,952	728.40	29.97	650	850
	Native Hawaiian or Pacific Islander	699	742.75	33.70	650	850
	Multiple Race Selected	12,133	741.54	33.61	650	850
	White	143,987	748.04	31.42	650	850
Economic Status*	Economically Disadvantaged	172,822	724.68	30.01	650	850
	Not Economically Disadvantaged	175,695	751.17	32.42	650	850
English Learner Status	English Learner (EL)	50,758	718.50	27.10	650	850
	Non English Learner	303,085	741.42	33.75	650	850
Disabilities	Students with Disabilities (SWD)	58,575	717.12	32.23	650	850
	Students without Disabilities	294,253	742.36	32.57	650	850
Language Form	Spanish	3,647	710.84	25.37	650	838

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.38 Subgroup Performance for Mathematics Scale Scores: Grade 5

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**355,854**	**738.22**	**33.85**	**650**	**850**
Gender	Female	174,514	739.18	32.62	650	850
	Male	181,340	737.29	34.98	650	850
Ethnicity	American Indian/Alaska Native	5,056	721.51	28.84	650	847
	Asian	23,053	770.69	34.15	650	850
	Black or African American	68,672	721.85	29.98	650	850
	Hispanic/Latino	100,022	728.56	29.81	650	850
	Native Hawaiian or Pacific Islander	725	745.92	31.43	650	845
	Multiple Race Selected	11,647	741.97	34.46	650	850
	White	146,558	747.61	31.44	650	850
Economic Status*	Economically Disadvantaged	171,884	724.73	29.78	650	850
	Not Economically Disadvantaged	178,461	750.95	32.66	650	850
English Learner Status	English Learner (EL)	32,382	711.69	26.08	650	850
	Non English Learner	323,270	740.89	33.39	650	850
Disabilities	Students with Disabilities (SWD)	60,194	717.65	31.23	650	850
	Students without Disabilities	294,518	742.47	32.81	650	850
Language Form	Spanish	2,960	706.68	24.00	650	790

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.39 Subgroup Performance for Mathematics Scale Scores: Grade 6

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**345,712**	**734.45**	**31.92**	**650**	**850**
Gender	Female	169,707	735.54	31.08	650	850
	Male	176,005	733.40	32.68	650	850
Ethnicity	American Indian/Alaska Native	4,941	719.32	26.67	650	818
	Asian	22,533	764.62	31.92	650	850
	Black or African American	65,796	718.17	27.79	650	850
	Hispanic/Latino	95,501	724.91	28.02	650	850
	Native Hawaiian or Pacific Islander	640	740.88	32.89	650	850
	Multiple Race Selected	10,744	737.88	32.51	650	850
	White	145,430	743.64	29.62	650	850
Economic Status*	Economically Disadvantaged	163,718	721.54	27.97	650	850
	Not Economically Disadvantaged	177,206	746.26	30.86	650	850
English Learner Status	English Learner (EL)	22,981	705.86	24.52	650	838
	Non English Learner	322,412	736.51	31.40	650	850
Disabilities	Students with Disabilities (SWD)	58,010	712.33	29.16	650	850
	Students without Disabilities	286,527	738.98	30.55	650	850
Language Form	Spanish	2,536	702.97	21.78	650	808

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.40 Subgroup Performance for Mathematics Scale Scores: Grade 7

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**323,440**	**735.75**	**29.36**	**650**	**850**
Gender	Female	158,086	737.33	28.47	650	850
	Male	165,354	734.24	30.12	650	850
Ethnicity	American Indian/Alaska Native	4,450	719.79	24.48	650	814
	Asian	18,846	761.37	29.43	650	850
	Black or African American	61,250	720.85	25.56	650	841
	Hispanic/Latino	90,916	727.46	26.26	650	850
	Native Hawaiian or Pacific Islander	540	744.54	29.42	650	847
	Multiple Race Selected	8,848	738.04	29.87	650	850
	White	138,562	744.63	27.45	650	850
Economic Status*	Economically Disadvantaged	154,536	724.48	26.08	650	850
	Not Economically Disadvantaged	168,750	746.08	28.39	650	850
English Learner Status	English Learner (EL)	20,668	709.10	22.70	650	823
	Non English Learner	302,529	737.59	28.88	650	850
Disabilities	Students with Disabilities (SWD)	55,972	714.53	27.49	650	842
	Students without Disabilities	266,420	740.26	27.75	650	850
Language Form	Spanish	2,586	708.68	20.64	640	789

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.41 Subgroup Performance for Mathematics Scale Scores: Grade 8

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**263,809**	**725.69**	**37.06**	**650**	**850**
Gender	Female	127,232	728.34	36.10	650	850
	Male	136,577	723.22	37.78	650	850
Ethnicity	American Indian/Alaska Native	3,853	710.70	30.24	650	834
	Asian	11,514	757.25	41.25	650	850
	Black or African American	52,795	708.20	31.51	650	850
	Hispanic/Latino	76,084	717.48	33.32	650	850
	Native Hawaiian or Pacific Islander	349	734.82	35.19	650	850
	Multiple Race Selected	6,982	726.90	38.27	650	850
	White	112,208	736.66	35.89	650	850
Economic Status*	Economically Disadvantaged	133,175	714.08	33.08	650	850
	Not Economically Disadvantaged	130,490	737.53	37.15	650	850
English Learner Status	English Learner (EL)	18,290	697.39	27.82	650	850
	Non English Learner	245,304	727.81	36.80	650	850
Disabilities	Students with Disabilities (SWD)	51,091	701.86	32.33	650	850
	Students without Disabilities	211,843	731.50	35.82	650	850
Language Form	Spanish	2,108	697.98	25.23	650	812

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.42 Subgroup Performance for Mathematics Scale Scores: Algebra I

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**221,242**	**741.16**	**36.64**	**650**	**850**
Gender	Female	107,530	742.58	35.06	650	850
	Male	113,712	739.82	38.02	650	850
Ethnicity	American Indian/Alaska Native	3,677	719.20	26.53	650	827
	Asian	17,417	775.45	37.13	650	850
	Black or African American	46,624	723.61	30.07	650	850
	Hispanic/Latino	61,703	727.11	31.35	650	850
	Native Hawaiian or Pacific Islander	507	747.04	36.34	650	850
	Multiple Race Selected	5,406	749.11	35.36	650	850
	White	85,812	754.25	33.49	650	850
Economic Status*	Economically Disadvantaged	86,722	725.73	30.96	650	850
	Not Economically Disadvantaged	130,987	751.35	36.72	650	850
English Learner Status	English Learner (EL)	17,105	708.36	26.13	650	850
	Non English Learner	203,950	743.94	36.05	650	850
Disabilities	Students with Disabilities (SWD)	37,607	718.31	32.60	650	850
	Students without Disabilities	183,295	745.86	35.65	650	850
Language Form	Spanish	3,579	704.51	21.04	650	802

Note: This table is identical to Table 12.10 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.43 Subgroup Performance for Mathematics Scale Scores: Geometry

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**130,412**	**734.03**	**27.61**	**650**	**850**
Gender	Female	64,117	734.70	26.63	650	850
	Male	66,295	733.39	28.52	650	850
Ethnicity	American Indian/Alaska Native	3,043	721.36	21.09	650	811
	Asian	11,269	760.52	29.13	650	850
	Black or African American	19,482	720.43	22.71	650	833
	Hispanic/Latino	39,809	723.51	22.70	650	828
	Native Hawaiian or Pacific Islander	390	735.84	25.78	669	797
	Multiple Race Selected	2,235	743.88	28.69	650	845
	White	54,141	741.45	25.87	650	850
Economic Status*	Economically Disadvantaged	50,368	721.82	22.56	650	850
	Not Economically Disadvantaged	77,023	742.00	27.91	650	850
English Learner Status	English Learner (EL)	8,246	710.71	19.49	650	827
	Non English Learner	121,663	735.66	27.37	650	850
Disabilities	Students with Disabilities (SWD)	21,545	716.69	23.87	650	850
	Students without Disabilities	108,538	737.50	27.00	650	850
Language Form	Spanish	1,961	708.41	17.21	650	774

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.44 Subgroup Performance for Mathematics Scale Scores: Algebra II

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**123,787**	**724.65**	**41.33**	**650**	**850**
Gender	Female	62,071	725.19	39.27	650	850
	Male	61,716	724.11	43.31	650	850
Ethnicity	American Indian/Alaska Native	2,976	707.34	28.64	650	850
	Asian	11,391	764.48	44.12	650	850
	Black or African American	20,983	705.48	32.23	650	850
	Hispanic/Latino	34,578	709.62	33.38	650	850
	Native Hawaiian or Pacific Islander	315	728.08	39.89	650	850
	Multiple Race Selected	2,218	734.73	41.94	650	850
	White	51,271	734.34	40.13	650	850
Economic Status*	Economically Disadvantaged	44,202	707.71	32.86	650	850
	Not Economically Disadvantaged	77,190	734.29	42.82	650	850
English Learner Status	English Learner (EL)	5,607	690.72	27.26	650	850
	Non English Learner	118,091	726.28	41.19	650	850
Disabilities	Students with Disabilities (SWD)	16,150	699.25	34.59	650	850
	Students without Disabilities	107,507	728.46	40.91	650	850
Language Form	Spanish	1,300	681.21	22.77	650	782

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.45 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics I

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**750**	**722.43**	**30.43**	**650**	**850**
Gender	Female	338	723.38	29.09	650	812
	Male	412	721.65	31.50	650	850
Ethnicity	American Indian/Alaska Native	40	725.25	30.50	677	812
	Asian	n/r	n/r	n/r	n/r	n/r
	Black or African American	35	708.14	25.84	650	766
	Hispanic/Latino	479	718.14	27.02	650	806
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	n/r	n/r	n/r	n/r	n/r
	White	186	735.39	35.21	674	850
Economic Status*	Economically Disadvantaged	505	717.88	28.24	650	814
	Not Economically Disadvantaged	245	731.81	32.62	650	850
English Learner Status	English Learner (EL)	110	709.45	20.69	662	758
	Non English Learner	640	724.66	31.28	650	850
Disabilities	Students with Disabilities (SWD)	160	709.23	30.00	650	850
	Students without Disabilities	590	726.01	29.57	650	818
Language Form	Spanish	n/r	n/r	n/r	n/r	n/r

Note: This table is identical to Table 12.11 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.46 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics II

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**873**	**716.45**	**31.08**	**650**	**850**
Gender	Female	450	715.83	30.30	650	850
	Male	423	717.12	31.92	650	850
Ethnicity	American Indian/Alaska Native	34	729.85	29.79	673	776
	Asian	n/r	n/r	n/r	n/r	n/r
	Black or African American	98	716.67	26.79	650	788
	Hispanic/Latino	538	708.07	24.14	650	850
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	n/r	n/r	n/r	n/r	n/r
	White	183	735.42	38.23	655	850
Economic Status*	Economically Disadvantaged	534	708.13	22.54	650	784
	Not Economically Disadvantaged	339	729.57	37.53	650	850
English Learner Status	English Learner (EL)	134	699.70	15.12	655	746
	Non English Learner	739	719.49	32.25	650	850
Disabilities	Students with Disabilities (SWD)	122	706.43	31.30	655	850
	Students without Disabilities	751	718.08	30.76	650	850
Language Form	Spanish	n/r	n/r	n/r	n/r	n/r

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.47 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics III

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**191**	**689.08**	**26.85**	**650**	**796**
Gender	Female	94	690.96	28.52	650	796
	Male	97	687.27	25.13	650	752
Ethnicity	American Indian/Alaska Native	n/r	n/r	n/r	n/r	n/r
	Asian	n/a	n/a	n/a	n/a	n/a
	Black or African American	n/r	n/r	n/r	n/r	n/r
	Hispanic/Latino	153	685.65	23.04	650	754
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	n/a	n/a	n/a	n/a	n/a
	White	n/r	n/r	n/r	n/r	n/r
Economic Status*	Economically Disadvantaged	185	688.64	26.73	650	796
	Not Economically Disadvantaged	n/r	n/r	n/r	n/r	n/r
English Learner Status	English Learner (EL)	25	682.08	16.34	650	713
	Non English Learner	166	690.14	27.97	650	796
Disabilities	Students with Disabilities (SWD)	n/r	n/r	n/r	n/r	n/r
	Students without Disabilities	176	689.40	26.26	650	758
Language Form	Spanish	n/r	n/r	n/r	n/r	n/r

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25. n/a = not applicable.

Appendix 12.5: Scale Score Cumulative Frequencies

Table A.12.48 Scale Score Cumulative Frequencies: ELA/L Grade 3

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	8,648	2.55	8,648	2.55
655-659	1,901	0.56	10,549	3.11
660-664	3,768	1.11	14,317	4.22
665-669	6,087	1.80	20,404	6.02
670-674	3,441	1.02	23,845	7.04
675-679	7,545	2.23	31,390	9.26
680-684	7,749	2.29	39,139	11.55
685-689	7,985	2.36	47,124	13.90
690-694	9,409	2.78	56,533	16.68
695-699	10,654	3.14	67,187	19.82
700-704	9,442	2.79	76,629	22.61
705-709	12,778	3.77	89,407	26.38
710-714	7,611	2.25	97,018	28.63
715-719	12,562	3.71	109,580	32.33
720-724	13,800	4.07	123,380	36.40
725-729	14,047	4.14	137,427	40.55
730-734	14,552	4.29	151,979	44.84
735-739	14,966	4.42	166,945	49.26
740-744	18,130	5.35	185,075	54.61
745-749	15,335	4.52	200,410	59.13
750-754	15,307	4.52	215,717	63.65
755-759	15,082	4.45	230,799	68.10
760-764	11,363	3.35	242,162	71.45
765-769	14,318	4.22	256,480	75.67
770-774	13,499	3.98	269,979	79.66
775-779	9,550	2.82	279,529	82.47
780-784	10,680	3.15	290,209	85.63
785-789	7,856	2.32	298,065	87.94
790-794	8,806	2.60	306,871	90.54
795-799	5,326	1.57	312,197	92.11
800-804	5,053	1.49	317,250	93.60
805-809	5,316	1.57	322,566	95.17
810-814	4,198	1.24	326,764	96.41
815-819	1,784	0.53	328,548	96.94
820-824	2,801	0.83	331,349	97.76
825-829	2,180	0.64	333,529	98.41
830-834	1,314	0.39	334,843	98.80
835-839	1,097	0.32	335,940	99.12
840-844	1,016	0.30	336,956	99.42
845-850	1,971	0.58	338,927	100.00

Table A.12.49 Scale Score Cumulative Frequencies: ELA/L Grade 4

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,325	0.38	1,325	0.38
655-659	1,303	0.38	2,628	0.76
660-664	624	0.18	3,252	0.94
665-669	2,300	0.67	5,552	1.61
670-674	3,242	0.94	8,794	2.55
675-679	4,232	1.22	13,026	3.77
680-684	7,565	2.19	20,591	5.96
685-689	6,060	1.75	26,651	7.71
690-694	8,746	2.53	35,397	10.25
695-699	8,565	2.48	43,962	12.72
700-704	11,166	3.23	55,128	15.96
705-709	11,185	3.24	66,313	19.19
710-714	11,013	3.19	77,326	22.38
715-719	14,609	4.23	91,935	26.61
720-724	13,888	4.02	105,823	30.63
725-729	17,792	5.15	123,615	35.78
730-734	15,412	4.46	139,027	40.24
735-739	17,905	5.18	156,932	45.42
740-744	17,994	5.21	174,926	50.63
745-749	17,706	5.12	192,632	55.76
750-754	17,524	5.07	210,156	60.83
755-759	16,858	4.88	227,014	65.71
760-764	15,879	4.60	242,893	70.31
765-769	17,526	5.07	260,419	75.38
770-774	11,989	3.47	272,408	78.85
775-779	13,142	3.80	285,550	82.65
780-784	10,148	2.94	295,698	85.59
785-789	10,573	3.06	306,271	88.65
790-794	7,841	2.27	314,112	90.92
795-799	6,873	1.99	320,985	92.91
800-804	5,581	1.62	326,566	94.52
805-809	3,965	1.15	330,531	95.67
810-814	4,640	1.34	335,171	97.02
815-819	2,477	0.72	337,648	97.73
820-824	2,257	0.65	339,905	98.39
825-829	1,510	0.44	341,415	98.82
830-834	1,058	0.31	342,473	99.13
835-839	970	0.28	343,443	99.41
840-844	664	0.19	344,107	99.60
845-850	1,376	0.40	345,483	100.00

Table A.12.50 Scale Score Cumulative Frequencies: ELA/L Grade 5

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,406	0.40	1,406	0.40
655-659	747	0.21	2,153	0.62
660-664	1,624	0.47	3,777	1.08
665-669	2,537	0.73	6,314	1.81
670-674	3,389	0.97	9,703	2.78
675-679	4,208	1.21	13,911	3.99
680-684	4,637	1.33	18,548	5.32
685-689	7,615	2.18	26,163	7.51
690-694	7,710	2.21	33,873	9.72
695-699	7,888	2.26	41,761	11.98
700-704	10,388	2.98	52,149	14.96
705-709	15,230	4.37	67,379	19.33
710-714	10,431	2.99	77,810	22.33
715-719	15,981	4.59	93,791	26.91
720-724	13,722	3.94	107,513	30.85
725-729	14,396	4.13	121,909	34.98
730-734	17,453	5.01	139,362	39.99
735-739	17,906	5.14	157,268	45.12
740-744	23,644	6.78	180,912	51.91
745-749	17,886	5.13	198,798	57.04
750-754	17,761	5.10	216,559	62.14
755-759	19,795	5.68	236,354	67.82
760-764	16,748	4.81	253,102	72.62
765-769	17,731	5.09	270,833	77.71
770-774	14,074	4.04	284,907	81.75
775-779	12,498	3.59	297,405	85.33
780-784	10,957	3.14	308,362	88.48
785-789	7,940	2.28	316,302	90.75
790-794	7,976	2.29	324,278	93.04
795-799	6,713	1.93	330,991	94.97
800-804	4,337	1.24	335,328	96.21
805-809	4,142	1.19	339,470	97.40
810-814	2,213	0.63	341,683	98.04
815-819	2,446	0.70	344,129	98.74
820-824	1,254	0.36	345,383	99.10
825-829	939	0.27	346,322	99.37
830-834	879	0.25	347,201	99.62
835-839	357	0.10	347,558	99.72
840-844	419	0.12	347,977	99.84
845-850	547	0.16	348,524	100.00

Table A.12.51 Scale Score Cumulative Frequencies: ELA/L Grade 6

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	413	0.12	413	0.12
655-659	291	0.08	704	0.20
660-664	484	0.14	1,188	0.34
665-669	823	0.24	2,011	0.58
670-674	1,668	0.48	3,679	1.07
675-679	2,888	0.84	6,567	1.91
680-684	3,989	1.16	10,556	3.06
685-689	7,767	2.25	18,323	5.32
690-694	8,744	2.54	27,067	7.86
695-699	9,495	2.76	36,562	10.61
700-704	12,137	3.52	48,699	14.14
705-709	11,780	3.42	60,479	17.55
710-714	16,544	4.80	77,023	22.36
715-719	13,324	3.87	90,347	26.22
720-724	16,365	4.75	106,712	30.97
725-729	18,591	5.40	125,303	36.37
730-734	18,849	5.47	144,152	41.84
735-739	18,279	5.31	162,431	47.15
740-744	20,591	5.98	183,022	53.12
745-749	20,112	5.84	203,134	58.96
750-754	19,482	5.65	222,616	64.62
755-759	16,308	4.73	238,924	69.35
760-764	17,697	5.14	256,621	74.49
765-769	16,681	4.84	273,302	79.33
770-774	13,104	3.80	286,406	83.13
775-779	12,269	3.56	298,675	86.69
780-784	10,659	3.09	309,334	89.79
785-789	7,951	2.31	317,285	92.09
790-794	6,754	1.96	324,039	94.06
795-799	5,487	1.59	329,526	95.65
800-804	2,972	0.86	332,498	96.51
805-809	3,724	1.08	336,222	97.59
810-814	2,312	0.67	338,534	98.26
815-819	1,487	0.43	340,021	98.69
820-824	1,728	0.50	341,749	99.20
825-829	902	0.26	342,651	99.46
830-834	661	0.19	343,312	99.65
835-839	375	0.11	343,687	99.76
840-844	282	0.08	343,969	99.84
845-850	551	0.16	344,520	100.00

Table A.12.52 Scale Score Cumulative Frequencies: ELA/L Grade 7

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	3,959	1.17	3,959	1.17
655-659	1,545	0.46	5,504	1.62
660-664	3,154	0.93	8,658	2.56
665-669	3,418	1.01	12,076	3.57
670-674	4,573	1.35	16,649	4.92
675-679	4,938	1.46	21,587	6.37
680-684	5,802	1.71	27,389	8.09
685-689	5,464	1.61	32,853	9.70
690-694	7,330	2.16	40,183	11.86
695-699	8,033	2.37	48,216	14.23
700-704	9,870	2.91	58,086	17.15
705-709	9,789	2.89	67,875	20.04
710-714	9,501	2.80	77,376	22.84
715-719	11,822	3.49	89,198	26.33
720-724	12,043	3.56	101,241	29.89
725-729	14,398	4.25	115,639	34.14
730-734	14,560	4.30	130,199	38.44
735-739	14,664	4.33	144,863	42.77
740-744	16,990	5.02	161,853	47.78
745-749	17,562	5.18	179,415	52.97
750-754	17,315	5.11	196,730	58.08
755-759	17,228	5.09	213,958	63.16
760-764	14,467	4.27	228,425	67.44
765-769	18,155	5.36	246,580	72.80
770-774	13,125	3.87	259,705	76.67
775-779	12,040	3.55	271,745	80.22
780-784	11,216	3.31	282,961	83.54
785-789	10,038	2.96	292,999	86.50
790-794	9,187	2.71	302,186	89.21
795-799	7,915	2.34	310,101	91.55
800-804	5,690	1.68	315,791	93.23
805-809	5,150	1.52	320,941	94.75
810-814	4,344	1.28	325,285	96.03
815-819	2,942	0.87	328,227	96.90
820-824	3,115	0.92	331,342	97.82
825-829	2,108	0.62	333,450	98.44
830-834	1,584	0.47	335,034	98.91
835-839	1,332	0.39	336,366	99.30
840-844	509	0.15	336,875	99.45
845-850	1,856	0.55	338,731	100.00

Table A.12.53 Scale Score Cumulative Frequencies: ELA/L Grade 8

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	3,620	1.07	3,620	1.07
655-659	998	0.29	4,618	1.36
660-664	2,852	0.84	7,470	2.20
665-669	3,684	1.09	11,154	3.29
670-674	4,392	1.29	15,546	4.58
675-679	4,706	1.39	20,252	5.97
680-684	7,778	2.29	28,030	8.26
685-689	7,438	2.19	35,468	10.45
690-694	7,644	2.25	43,112	12.71
695-699	9,390	2.77	52,502	15.47
700-704	8,974	2.64	61,476	18.12
705-709	11,209	3.30	72,685	21.42
710-714	11,313	3.33	83,998	24.76
715-719	9,409	2.77	93,407	27.53
720-724	16,228	4.78	109,635	32.31
725-729	14,409	4.25	124,044	36.56
730-734	14,696	4.33	138,740	40.89
735-739	17,012	5.01	155,752	45.91
740-744	16,760	4.94	172,512	50.85
745-749	17,218	5.07	189,730	55.92
750-754	16,447	4.85	206,177	60.77
755-759	14,135	4.17	220,312	64.93
760-764	16,043	4.73	236,355	69.66
765-769	15,117	4.46	251,472	74.12
770-774	12,126	3.57	263,598	77.69
775-779	11,465	3.38	275,063	81.07
780-784	11,994	3.54	287,057	84.61
785-789	8,187	2.41	295,244	87.02
790-794	8,704	2.57	303,948	89.59
795-799	6,490	1.91	310,438	91.50
800-804	6,966	2.05	317,404	93.55
805-809	4,030	1.19	321,434	94.74
810-814	3,664	1.08	325,098	95.82
815-819	3,162	0.93	328,260	96.75
820-824	2,750	0.81	331,010	97.56
825-829	2,238	0.66	333,248	98.22
830-834	1,409	0.42	334,657	98.64
835-839	1,496	0.44	336,153	99.08
840-844	669	0.20	336,822	99.27
845-850	2,461	0.73	339,283	100.00

Table A.12.54 Scale Score Cumulative Frequencies: ELA/L Grade 9

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,354	1.06	1,354	1.06
655-659	579	0.45	1,933	1.51
660-664	626	0.49	2,559	2.00
665-669	1,255	0.98	3,814	2.97
670-674	1,646	1.28	5,460	4.26
675-679	1,942	1.51	7,402	5.77
680-684	1,979	1.54	9,381	7.32
685-689	1,944	1.52	11,325	8.83
690-694	2,816	2.20	14,141	11.03
695-699	3,434	2.68	17,575	13.71
700-704	2,567	2.00	20,142	15.71
705-709	3,262	2.54	23,404	18.25
710-714	4,685	3.65	28,089	21.91
715-719	3,142	2.45	31,231	24.36
720-724	5,368	4.19	36,599	28.54
725-729	4,734	3.69	41,333	32.23
730-734	5,631	4.39	46,964	36.63
735-739	5,673	4.42	52,637	41.05
740-744	6,492	5.06	59,129	46.11
745-749	6,557	5.11	65,686	51.23
750-754	6,574	5.13	72,260	56.35
755-759	5,789	4.51	78,049	60.87
760-764	6,506	5.07	84,555	65.94
765-769	6,546	5.10	91,101	71.05
770-774	5,411	4.22	96,512	75.27
775-779	4,554	3.55	101,066	78.82
780-784	4,963	3.87	106,029	82.69
785-789	4,424	3.45	110,453	86.14
790-794	3,517	2.74	113,970	88.88
795-799	3,041	2.37	117,011	91.25
800-804	2,242	1.75	119,253	93.00
805-809	2,470	1.93	121,723	94.93
810-814	1,430	1.12	123,153	96.04
815-819	1,736	1.35	124,889	97.40
820-824	658	0.51	125,547	97.91
825-829	830	0.65	126,377	98.56
830-834	670	0.52	127,047	99.08
835-839	353	0.28	127,400	99.35
840-844	199	0.16	127,599	99.51
845-850	630	0.49	128,229	100.00

Table A.12.55 Scale Score Cumulative Frequencies: ELA/L Grade 10

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	7,707	4.09	7,707	4.09
655-659	1,497	0.79	9,204	4.88
660-664	3,183	1.69	12,387	6.57
665-669	3,293	1.75	15,680	8.31
670-674	3,146	1.67	18,826	9.98
675-679	3,171	1.68	21,997	11.66
680-684	3,174	1.68	25,171	13.35
685-689	2,995	1.59	28,166	14.93
690-694	4,216	2.24	32,382	17.17
695-699	3,976	2.11	36,358	19.28
700-704	3,967	2.10	40,325	21.38
705-709	5,252	2.78	45,577	24.17
710-714	5,270	2.79	50,847	26.96
715-719	5,209	2.76	56,056	29.72
720-724	6,641	3.52	62,697	33.24
725-729	6,657	3.53	69,354	36.77
730-734	6,770	3.59	76,124	40.36
735-739	6,820	3.62	82,944	43.98
740-744	8,374	4.44	91,318	48.42
745-749	8,294	4.40	99,612	52.82
750-754	5,444	2.89	105,056	55.70
755-759	8,346	4.43	113,402	60.13
760-764	8,066	4.28	121,468	64.41
765-769	6,515	3.45	127,983	67.86
770-774	7,575	4.02	135,558	71.88
775-779	6,046	3.21	141,604	75.08
780-784	5,711	3.03	147,315	78.11
785-789	6,416	3.40	153,731	81.51
790-794	4,083	2.16	157,814	83.68
795-799	5,478	2.90	163,292	86.58
800-804	3,436	1.82	166,728	88.40
805-809	4,548	2.41	171,276	90.82
810-814	2,755	1.46	174,031	92.28
815-819	2,458	1.30	176,489	93.58
820-824	2,196	1.16	178,685	94.74
825-829	1,962	1.04	180,647	95.78
830-834	1,689	0.90	182,336	96.68
835-839	1,449	0.77	183,785	97.45
840-844	1,182	0.63	184,967	98.08
845-850	3,630	1.92	188,597	100.00

Table A.12.56 Scale Score Cumulative Frequencies: ELA/L Grade 11

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	2,516	2.45	2,516	2.45
655-659	114	0.11	2,630	2.56
660-664	1,725	1.68	4,355	4.24
665-669	1,144	1.11	5,499	5.35
670-674	1,105	1.08	6,604	6.43
675-679	2,506	2.44	9,110	8.87
680-684	2,627	2.56	11,737	11.42
685-689	2,559	2.49	14,296	13.91
690-694	3,526	3.43	17,822	17.34
695-699	3,433	3.34	21,255	20.69
700-704	2,897	2.82	24,152	23.50
705-709	3,709	3.61	27,861	27.11
710-714	3,522	3.43	31,383	30.54
715-719	3,503	3.41	34,886	33.95
720-724	3,333	3.24	38,219	37.19
725-729	4,927	4.79	43,146	41.99
730-734	3,988	3.88	47,134	45.87
735-739	4,818	4.69	51,952	50.56
740-744	4,746	4.62	56,698	55.18
745-749	5,203	5.06	61,901	60.24
750-754	5,020	4.89	66,921	65.13
755-759	4,710	4.58	71,631	69.71
760-764	4,411	4.29	76,042	74.00
765-769	4,198	4.09	80,240	78.09
770-774	4,310	4.19	84,550	82.28
775-779	2,926	2.85	87,476	85.13
780-784	3,089	3.01	90,565	88.14
785-789	2,290	2.23	92,855	90.37
790-794	2,320	2.26	95,175	92.62
795-799	1,710	1.66	96,885	94.29
800-804	1,443	1.40	98,328	95.69
805-809	1,095	1.07	99,423	96.76
810-814	800	0.78	100,223	97.54
815-819	774	0.75	100,997	98.29
820-824	431	0.42	101,428	98.71
825-829	484	0.47	101,912	99.18
830-834	273	0.27	102,185	99.45
835-839	188	0.18	102,373	99.63
840-844	67	0.07	102,440	99.69
845-850	315	0.31	102,755	100.00

Table A.12.57 Scale Score Cumulative Frequencies: Mathematics Grade 3

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	2,593	0.74	2,593	0.74
655-659	1,258	0.36	3,851	1.11
660-664	2,287	0.66	6,138	1.76
665-669	1,988	0.57	8,126	2.33
670-674	2,808	0.81	10,934	3.14
675-679	4,545	1.31	15,479	4.45
680-684	5,130	1.47	20,609	5.92
685-689	5,766	1.66	26,375	7.58
690-694	11,393	3.27	37,768	10.85
695-699	6,684	1.92	44,452	12.77
700-704	10,233	2.94	54,685	15.71
705-709	13,013	3.74	67,698	19.45
710-714	14,720	4.23	82,418	23.68
715-719	15,117	4.34	97,535	28.02
720-724	15,299	4.39	112,834	32.41
725-729	15,449	4.44	128,283	36.85
730-734	15,621	4.49	143,904	41.34
735-739	20,079	5.77	163,983	47.11
740-744	16,636	4.78	180,619	51.88
745-749	18,322	5.26	198,941	57.15
750-754	15,659	4.50	214,600	61.65
755-759	21,618	6.21	236,218	67.86
760-764	14,889	4.28	251,107	72.13
765-769	13,750	3.95	264,857	76.08
770-774	15,406	4.43	280,263	80.51
775-779	12,693	3.65	292,956	84.15
780-784	8,671	2.49	301,627	86.65
785-789	8,779	2.52	310,406	89.17
790-794	9,465	2.72	319,871	91.89
795-799	7,640	2.19	327,511	94.08
800-804	4,078	1.17	331,589	95.25
805-809	3,335	0.96	334,924	96.21
810-814	4,309	1.24	339,233	97.45
815-819	2,279	0.65	341,512	98.10
820-824	1,925	0.55	343,437	98.66
825-829	983	0.28	344,420	98.94
830-834	1,262	0.36	345,682	99.30
835-839	534	0.15	346,216	99.45
840-844	491	0.14	346,707	99.59
845-850	1,410	0.41	348,117	100.00

Table A.12.58 Scale Score Cumulative Frequencies: Mathematics Grade 4

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,179	0.33	1,179	0.33
655-659	989	0.28	2,168	0.61
660-664	1,020	0.29	3,188	0.90
665-669	2,124	0.60	5,312	1.50
670-674	2,790	0.79	8,102	2.29
675-679	4,766	1.35	12,868	3.63
680-684	6,013	1.70	18,881	5.33
685-689	6,971	1.97	25,852	7.30
690-694	9,332	2.64	35,184	9.94
695-699	12,188	3.44	47,372	13.38
700-704	12,716	3.59	60,088	16.97
705-709	14,543	4.11	74,631	21.08
710-714	17,334	4.90	91,965	25.97
715-719	17,314	4.89	109,279	30.86
720-724	17,325	4.89	126,604	35.76
725-729	16,777	4.74	143,381	40.49
730-734	22,409	6.33	165,790	46.82
735-739	16,457	4.65	182,247	51.47
740-744	22,471	6.35	204,718	57.82
745-749	15,845	4.47	220,563	62.29
750-754	21,124	5.97	241,687	68.26
755-759	16,967	4.79	258,654	73.05
760-764	16,257	4.59	274,911	77.64
765-769	15,306	4.32	290,217	81.96
770-774	11,089	3.13	301,306	85.10
775-779	11,031	3.12	312,337	88.21
780-784	9,303	2.63	321,640	90.84
785-789	8,087	2.28	329,727	93.12
790-794	6,817	1.93	336,544	95.05
795-799	5,846	1.65	342,390	96.70
800-804	2,857	0.81	345,247	97.51
805-809	2,244	0.63	347,491	98.14
810-814	1,829	0.52	349,320	98.66
815-819	1,443	0.41	350,763	99.06
820-824	1,094	0.31	351,857	99.37
825-829	452	0.13	352,309	99.50
830-834	767	0.22	353,076	99.72
835-839	20	0.01	353,096	99.72
840-844	492	0.14	353,588	99.86
845-850	492	0.14	354,080	100.00

Table A.12.59 Scale Score Cumulative Frequencies: Mathematics Grade 5

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,442	0.41	1,442	0.41
655-659	984	0.28	2,426	0.68
660-664	1,201	0.34	3,627	1.02
665-669	1,735	0.49	5,362	1.51
670-674	3,393	0.95	8,755	2.46
675-679	4,005	1.13	12,760	3.59
680-684	5,187	1.46	17,947	5.04
685-689	5,745	1.61	23,692	6.66
690-694	8,544	2.40	32,236	9.06
695-699	16,603	4.67	48,839	13.72
700-704	10,177	2.86	59,016	16.58
705-709	14,914	4.19	73,930	20.78
710-714	15,053	4.23	88,983	25.01
715-719	19,168	5.39	108,151	30.39
720-724	18,565	5.22	126,716	35.61
725-729	17,710	4.98	144,426	40.59
730-734	18,100	5.09	162,526	45.67
735-739	24,103	6.77	186,629	52.45
740-744	19,610	5.51	206,239	57.96
745-749	16,446	4.62	222,685	62.58
750-754	21,089	5.93	243,774	68.50
755-759	16,591	4.66	260,365	73.17
760-764	18,327	5.15	278,692	78.32
765-769	12,516	3.52	291,208	81.83
770-774	13,116	3.69	304,324	85.52
775-779	10,011	2.81	314,335	88.33
780-784	10,733	3.02	325,068	91.35
785-789	7,290	2.05	332,358	93.40
790-794	6,208	1.74	338,566	95.14
795-799	4,024	1.13	342,590	96.27
800-804	3,374	0.95	345,964	97.22
805-809	2,835	0.80	348,799	98.02
810-814	1,676	0.47	350,475	98.49
815-819	1,374	0.39	351,849	98.87
820-824	1,151	0.32	353,000	99.20
825-829	884	0.25	353,884	99.45
830-834	358	0.10	354,242	99.55
835-839	622	0.17	354,864	99.72
840-844	27	0.01	354,891	99.73
845-850	963	0.27	355,854	100.00

Table A.12.60 Scale Score Cumulative Frequencies: Mathematics Grade 6

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,821	0.53	1,821	0.53
655-659	635	0.18	2,456	0.71
660-664	805	0.23	3,261	0.94
665-669	3,081	0.89	6,342	1.83
670-674	1,452	0.42	7,794	2.25
675-679	4,780	1.38	12,574	3.64
680-684	6,746	1.95	19,320	5.59
685-689	5,257	1.52	24,577	7.11
690-694	11,197	3.24	35,774	10.35
695-699	11,643	3.37	47,417	13.72
700-704	15,901	4.60	63,318	18.32
705-709	16,731	4.84	80,049	23.15
710-714	18,486	5.35	98,535	28.50
715-719	17,512	5.07	116,047	33.57
720-724	21,126	6.11	137,173	39.68
725-729	19,834	5.74	157,007	45.42
730-734	18,784	5.43	175,791	50.85
735-739	20,769	6.01	196,560	56.86
740-744	17,338	5.02	213,898	61.87
745-749	20,326	5.88	234,224	67.75
750-754	18,610	5.38	252,834	73.13
755-759	16,574	4.79	269,408	77.93
760-764	14,561	4.21	283,969	82.14
765-769	12,415	3.59	296,384	85.73
770-774	10,791	3.12	307,175	88.85
775-779	9,664	2.80	316,839	91.65
780-784	7,241	2.09	324,080	93.74
785-789	6,839	1.98	330,919	95.72
790-794	4,947	1.43	335,866	97.15
795-799	3,163	0.91	339,029	98.07
800-804	1,939	0.56	340,968	98.63
805-809	1,638	0.47	342,606	99.10
810-814	1,068	0.31	343,674	99.41
815-819	636	0.18	344,310	99.59
820-824	249	0.07	344,559	99.67
825-829	366	0.11	344,925	99.77
830-834	222	0.06	345,147	99.84
835-839	98	0.03	345,245	99.86
840-844	171	0.05	345,416	99.91
845-850	296	0.09	345,712	100.00

Table A.12.61 Scale Score Cumulative Frequencies: Mathematics Grade 7

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	970	0.30	970	0.30
655-659	466	0.14	1,436	0.44
660-664	769	0.24	2,205	0.68
665-669	770	0.24	2,975	0.92
670-674	1,579	0.49	4,554	1.41
675-679	2,948	0.91	7,502	2.32
680-684	2,954	0.91	10,456	3.23
685-689	6,398	1.98	16,854	5.21
690-694	10,366	3.20	27,220	8.42
695-699	7,833	2.42	35,053	10.84
700-704	16,092	4.98	51,145	15.81
705-709	12,417	3.84	63,562	19.65
710-714	15,255	4.72	78,817	24.37
715-719	20,115	6.22	98,932	30.59
720-724	20,011	6.19	118,943	36.77
725-729	18,623	5.76	137,566	42.53
730-734	20,537	6.35	158,103	48.88
735-739	22,446	6.94	180,549	55.82
740-744	18,228	5.64	198,777	61.46
745-749	17,638	5.45	216,415	66.91
750-754	18,284	5.65	234,699	72.56
755-759	18,773	5.80	253,472	78.37
760-764	14,319	4.43	267,791	82.79
765-769	12,682	3.92	280,473	86.72
770-774	10,667	3.30	291,140	90.01
775-779	9,622	2.97	300,762	92.99
780-784	6,308	1.95	307,070	94.94
785-789	5,924	1.83	312,994	96.77
790-794	3,180	0.98	316,174	97.75
795-799	2,575	0.80	318,749	98.55
800-804	1,963	0.61	320,712	99.16
805-809	979	0.30	321,691	99.46
810-814	778	0.24	322,469	99.70
815-819	348	0.11	322,817	99.81
820-824	265	0.08	323,082	99.89
825-829	13	0.00	323,095	99.89
830-834	165	0.05	323,260	99.94
835-839	2	0.00	323,262	99.94
840-844	83	0.03	323,345	99.97
845-850	95	0.03	323,440	100.00

Table A.12.62 Scale Score Cumulative Frequencies: Mathematics Grade 8

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	5,761	2.18	5,761	2.18
655-659	2,110	0.80	7,871	2.98
660-664	4,286	1.62	12,157	4.61
665-669	3,386	1.28	15,543	5.89
670-674	5,692	2.16	21,235	8.05
675-679	4,708	1.78	25,943	9.83
680-684	12,538	4.75	38,481	14.59
685-689	7,137	2.71	45,618	17.29
690-694	10,941	4.15	56,559	21.44
695-699	12,913	4.89	69,472	26.33
700-704	10,017	3.80	79,489	30.13
705-709	14,449	5.48	93,938	35.61
710-714	11,205	4.25	105,143	39.86
715-719	16,762	6.35	121,905	46.21
720-724	11,199	4.25	133,104	50.45
725-729	12,122	4.59	145,226	55.05
730-734	14,899	5.65	160,125	60.70
735-739	14,887	5.64	175,012	66.34
740-744	7,693	2.92	182,705	69.26
745-749	12,730	4.83	195,435	74.08
750-754	11,129	4.22	206,564	78.30
755-759	6,641	2.52	213,205	80.82
760-764	8,485	3.22	221,690	84.03
765-769	7,866	2.98	229,556	87.02
770-774	6,142	2.33	235,698	89.34
775-779	7,353	2.79	243,051	92.13
780-784	4,143	1.57	247,194	93.70
785-789	3,440	1.30	250,634	95.01
790-794	3,782	1.43	254,416	96.44
795-799	2,198	0.83	256,614	97.27
800-804	1,820	0.69	258,434	97.96
805-809	1,434	0.54	259,868	98.51
810-814	625	0.24	260,493	98.74
815-819	1,043	0.40	261,536	99.14
820-824	418	0.16	261,954	99.30
825-829	630	0.24	262,584	99.54
830-834	293	0.11	262,877	99.65
835-839	234	0.09	263,111	99.74
840-844	200	0.08	263,311	99.81
845-850	498	0.19	263,809	100.00

Table A.12.63 Scale Score Cumulative Frequencies: Algebra I

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	1,409	0.64	1,409	0.64
655-659	196	0.09	1,605	0.73
660-664	13	0.01	1,618	0.73
665-669	2,321	1.05	3,939	1.78
670-674	407	0.18	4,346	1.96
675-679	4,260	1.93	8,606	3.89
680-684	617	0.28	9,223	4.17
685-689	7,309	3.30	16,532	7.47
690-694	3,137	1.42	19,669	8.89
695-699	9,307	4.21	28,976	13.10
700-704	6,533	2.95	35,509	16.05
705-709	13,168	5.95	48,677	22.00
710-714	9,251	4.18	57,928	26.18
715-719	8,816	3.98	66,744	30.17
720-724	11,414	5.16	78,158	35.33
725-729	11,609	5.25	89,767	40.57
730-734	10,206	4.61	99,973	45.19
735-739	12,553	5.67	112,526	50.86
740-744	11,137	5.03	123,663	55.89
745-749	10,252	4.63	133,915	60.53
750-754	9,252	4.18	143,167	64.71
755-759	8,626	3.90	151,793	68.61
760-764	10,604	4.79	162,397	73.40
765-769	8,338	3.77	170,735	77.17
770-774	8,482	3.83	179,217	81.00
775-779	8,058	3.64	187,275	84.65
780-784	5,445	2.46	192,720	87.11
785-789	5,510	2.49	198,230	89.60
790-794	4,850	2.19	203,080	91.79
795-799	3,946	1.78	207,026	93.57
800-804	3,391	1.53	210,417	95.11
805-809	2,728	1.23	213,145	96.34
810-814	2,308	1.04	215,453	97.38
815-819	1,285	0.58	216,738	97.96
820-824	1,308	0.59	218,046	98.56
825-829	858	0.39	218,904	98.94
830-834	539	0.24	219,443	99.19
835-839	467	0.21	219,910	99.40
840-844	459	0.21	220,369	99.61
845-850	873	0.39	221,242	100.00

Table A.12.64 Scale Score Cumulative Frequencies: Geometry

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	205	0.16	205	0.16
655-659	0	0.00	205	0.16
660-664	206	0.16	411	0.32
665-669	308	0.24	719	0.55
670-674	153	0.12	872	0.67
675-679	474	0.36	1,346	1.03
680-684	1,129	0.87	2,475	1.90
685-689	1,891	1.45	4,366	3.35
690-694	4,003	3.07	8,369	6.42
695-699	3,551	2.72	11,920	9.14
700-704	6,182	4.74	18,102	13.88
705-709	7,024	5.39	25,126	19.27
710-714	10,085	7.73	35,211	27.00
715-719	7,765	5.95	42,976	32.95
720-724	11,525	8.84	54,501	41.79
725-729	5,771	4.43	60,272	46.22
730-734	9,372	7.19	69,644	53.40
735-739	7,860	6.03	77,504	59.43
740-744	9,162	7.03	86,666	66.46
745-749	6,348	4.87	93,014	71.32
750-754	6,178	4.74	99,192	76.06
755-759	5,737	4.40	104,929	80.46
760-764	4,948	3.79	109,877	84.25
765-769	5,393	4.14	115,270	88.39
770-774	3,823	2.93	119,093	91.32
775-779	3,485	2.67	122,578	93.99
780-784	2,684	2.06	125,262	96.05
785-789	1,795	1.38	127,057	97.43
790-794	1,352	1.04	128,409	98.46
795-799	699	0.54	129,108	99.00
800-804	460	0.35	129,568	99.35
805-809	379	0.29	129,947	99.64
810-814	206	0.16	130,153	99.80
815-819	74	0.06	130,227	99.86
820-824	80	0.06	130,307	99.92
825-829	35	0.03	130,342	99.95
830-834	26	0.02	130,368	99.97
835-839	8	0.01	130,376	99.97
840-844	11	0.01	130,387	99.98
845-850	25	0.02	130,412	100.00

Table A.12.65 Scale Score Cumulative Frequencies: Algebra II

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	2,727	2.20	2,727	2.20
655-659	2,339	1.89	5,066	4.09
660-664	1,876	1.52	6,942	5.61
665-669	2,080	1.68	9,022	7.29
670-674	2,395	1.93	11,417	9.22
675-679	5,691	4.60	17,108	13.82
680-684	3,581	2.89	20,689	16.71
685-689	2,940	2.38	23,629	19.09
690-694	7,012	5.66	30,641	24.75
695-699	6,765	5.47	37,406	30.22
700-704	6,376	5.15	43,782	35.37
705-709	5,876	4.75	49,658	40.12
710-714	5,159	4.17	54,817	44.28
715-719	6,617	5.35	61,434	49.63
720-724	5,705	4.61	67,139	54.24
725-729	5,222	4.22	72,361	58.46
730-734	4,288	3.46	76,649	61.92
735-739	5,342	4.32	81,991	66.24
740-744	4,587	3.71	86,578	69.94
745-749	4,073	3.29	90,651	73.23
750-754	3,490	2.82	94,141	76.05
755-759	4,697	3.79	98,838	79.85
760-764	3,396	2.74	102,234	82.59
765-769	2,525	2.04	104,759	84.63
770-774	3,286	2.65	108,045	87.28
775-779	2,315	1.87	110,360	89.15
780-784	2,603	2.10	112,963	91.26
785-789	1,539	1.24	114,502	92.50
790-794	2,017	1.63	116,519	94.13
795-799	1,192	0.96	117,711	95.09
800-804	1,236	1.00	118,947	96.09
805-809	1,058	0.85	120,005	96.94
810-814	771	0.62	120,776	97.57
815-819	540	0.44	121,316	98.00
820-824	476	0.38	121,792	98.39
825-829	524	0.42	122,316	98.81
830-834	267	0.22	122,583	99.03
835-839	165	0.13	122,748	99.16
840-844	172	0.14	122,920	99.30
845-850	867	0.70	123,787	100.00

Table A.12.66 Scale Score Cumulative Frequencies: Integrated Mathematics I

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	7	0.93	7	0.93
655-659	0	0.00	7	0.93
660-664	4	0.53	11	1.47
665-669	5	0.67	16	2.13
670-674	6	0.80	22	2.93
675-679	18	2.40	40	5.33
680-684	8	1.07	48	6.40
685-689	26	3.47	74	9.87
690-694	46	6.13	120	16.00
695-699	57	7.60	177	23.60
700-704	44	5.87	221	29.47
705-709	46	6.13	267	35.60
710-714	54	7.20	321	42.80
715-719	89	11.87	410	54.67
720-724	34	4.53	444	59.20
725-729	51	6.80	495	66.00
730-734	29	3.87	524	69.87
735-739	23	3.07	547	72.93
740-744	26	3.47	573	76.40
745-749	35	4.67	608	81.07
750-754	20	2.67	628	83.73
755-759	26	3.47	654	87.20
760-764	21	2.80	675	90.00
765-769	16	2.13	691	92.13
770-774	15	2.00	706	94.13
775-779	10	1.33	716	95.47
780-784	8	1.07	724	96.53
785-789	7	0.93	731	97.47
790-794	5	0.67	736	98.13
795-799	4	0.53	740	98.67
800-804	2	0.27	742	98.93
805-809	4	0.53	746	99.47
810-814	2	0.27	748	99.73
815-819	1	0.13	749	99.87
820-824	0	0.00	749	99.87
825-829	0	0.00	749	99.87
830-834	0	0.00	749	99.87
835-839	0	0.00	749	99.87
840-844	0	0.00	749	99.87
845-850	1	0.13	750	100.00

Table A.12.67 Scale Score Cumulative Frequencies: Integrated Mathematics II

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	3	0.34	3	0.34
655-659	6	0.69	9	1.03
660-664	0	0.00	9	1.03
665-669	14	1.60	23	2.63
670-674	1	0.11	24	2.75
675-679	35	4.01	59	6.76
680-684	49	5.61	108	12.37
685-689	11	1.26	119	13.63
690-694	66	7.56	185	21.19
695-699	86	9.85	271	31.04
700-704	96	11.00	367	42.04
705-709	66	7.56	433	49.60
710-714	77	8.82	510	58.42
715-719	59	6.76	569	65.18
720-724	37	4.24	606	69.42
725-729	52	5.96	658	75.37
730-734	22	2.52	680	77.89
735-739	19	2.18	699	80.07
740-744	27	3.09	726	83.16
745-749	20	2.29	746	85.45
750-754	8	0.92	754	86.37
755-759	24	2.75	778	89.12
760-764	18	2.06	796	91.18
765-769	14	1.60	810	92.78
770-774	16	1.83	826	94.62
775-779	7	0.80	833	95.42
780-784	8	0.92	841	96.33
785-789	5	0.57	846	96.91
790-794	9	1.03	855	97.94
795-799	4	0.46	859	98.40
800-804	3	0.34	862	98.74
805-809	2	0.23	864	98.97
810-814	3	0.34	867	99.31
815-819	0	0.00	867	99.31
820-824	1	0.11	868	99.43
825-829	1	0.11	869	99.54
830-834	0	0.00	869	99.54
835-839	0	0.00	869	99.54
840-844	1	0.11	870	99.66
845-850	3	0.34	873	100.00

Table A.12.68 Scale Score Cumulative Frequencies: Integrated Mathematics III

Score Band	Count	Percent	Cumulative Count	Cumulative Percent
650-654	16	8.38	16	8.38
655-659	0	0.00	16	8.38
660-664	24	12.57	40	20.94
665-669	0	0.00	40	20.94
670-674	24	12.57	64	33.51
675-679	20	10.47	84	43.98
680-684	2	1.05	86	45.03
685-689	23	12.04	109	57.07
690-694	15	7.85	124	64.92
695-699	20	10.47	144	75.39
700-704	0	0.00	144	75.39
705-709	10	5.24	154	80.63
710-714	7	3.66	161	84.29
715-719	5	2.62	166	86.91
720-724	3	1.57	169	88.48
725-729	2	1.05	171	89.53
730-734	6	3.14	177	92.67
735-739	5	2.62	182	95.29
740-744	2	1.05	184	96.34
745-749	1	0.52	185	96.86
750-754	4	2.09	189	98.95
755-759	1	0.52	190	99.48
760-764	0	0.00	190	99.48
765-769	0	0.00	190	99.48
770-774	0	0.00	190	99.48
775-779	0	0.00	190	99.48
780-784	0	0.00	190	99.48
785-789	0	0.00	190	99.48
790-794	0	0.00	190	99.48
795-799	1	0.52	191	100.00
800-804	0	0.00	191	100.00
805-809	0	0.00	191	100.00
810-814	0	0.00	191	100.00
815-819	0	0.00	191	100.00
820-824	0	0.00	191	100.00
825-829	0	0.00	191	100.00
830-834	0	0.00	191	100.00
835-839	0	0.00	191	100.00
840-844	0	0.00	191	100.00
845-850	0	0.00	191	100.00

Appendix 13: Growth

Appendix 13 provides the summary growth results for subgroups for grade 4 – 11 ELA/L and mathematics 4 – 8 and high school. Integrated mathematics I, II, and III do not have sufficient sample sizes for subgroup summary analysis.

Table A.13.1 Summary of SGP Estimates for Subgroups: Grade 4 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	166,982	47.84	13.49	47
Female	161,281	52.21	13.25	53
Ethnicity				
White	135,737	50.89	13.14	51
African American	64,219	45.70	13.76	44
Asian/Pacific Islander	21,899	60.41	12.64	64
American Indian/Alaska Native	4,702	44.04	13.97	42
Hispanic	90,846	49.54	13.60	49
Multiple	10,834	49.40	13.28	49
Special Instructional Needs				
Economically Disadvantaged	162,052	47.15	13.70	46
Not-economically Disadvantaged	166,211	52.75	13.06	54
English Learner (EL)	43,019	48.83	14.09	49
Non English Learner	285,244	50.16	13.27	50
Students with Disabilities (SWD)	54,975	42.72	14.23	40
Students without Disabilities	273,288	51.45	13.20	52

Table A.13.2 Summary of SGP Estimates for Subgroups: Grade 5 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	168,872	46.56	13.91	45
Female	163,028	53.70	13.68	55
Ethnicity				
White	138,625	51.18	13.70	52
African American	64,130	45.60	14.02	44
Asian/Pacific Islander	21,604	57.51	13.59	60
American Indian/Alaska Native	4,548	50.36	14.23	51
Hispanic	92,619	49.80	13.81	50
Multiple	10,362	49.57	13.79	49
Special Instructional Needs				
Economically Disadvantaged	162,259	47.90	13.93	47
Not-economically Disadvantaged	169,641	52.13	13.67	53
English Learner (EL)	25,902	47.21	14.79	46
Non English Learner	305,998	50.31	13.71	50
Students with Disabilities (SWD)	56,743	44.67	14.67	42
Students without Disabilities	275,157	51.18	13.62	52

Table A.13.3 Summary of SGP Estimates for Subgroups: Grade 6 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	163,653	47.54	14.00	46
Female	158,016	52.59	13.58	54
Ethnicity				
White	137,019	50.92	13.43	51
African American	60,830	45.40	14.36	44
Asian/Pacific Islander	21,270	58.37	13.34	62
American Indian/Alaska Native	4,451	53.09	14.03	55
Hispanic	88,486	49.81	14.06	50
Multiple	9,588	48.44	13.77	48
Special Instructional Needs				
Economically Disadvantaged	153,808	48.28	14.19	48
Not-economically Disadvantaged	167,861	51.61	13.43	52
English Learner (EL)	17,371	47.87	16.16	47
Non English Learner	304,298	50.14	13.66	50
Students with Disabilities (SWD)	54,560	45.40	15.37	43
Students without Disabilities	267,109	50.96	13.47	51

Table A.13.4 Summary of SGP Estimates for Subgroups: Grade 7 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	162,366	47.03	13.48	46
Female	155,404	53.07	13.47	54
Ethnicity				
White	138,429	50.29	13.40	50
African American	58,817	47.92	13.57	47
Asian/Pacific Islander	21,551	57.54	13.86	60
American Indian/Alaska Native	3,872	46.99	13.50	46
Hispanic	86,139	49.23	13.44	49
Multiple	8,949	49.17	13.37	49
Special Instructional Needs				
Economically Disadvantaged	146,412	48.25	13.47	47
Not-economically Disadvantaged	171,358	51.46	13.48	52
English Learner (EL)	15,283	46.74	14.38	45
Non English Learner	302,487	50.15	13.43	50
Students with Disabilities (SWD)	53,872	45.53	13.99	43
Students without Disabilities	263,898	50.89	13.37	51

Table A.13.5 Summary of SGP Estimates for Subgroups: Grade 8 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	163,179	48.78	14.19	48
Female	156,202	51.23	14.17	52
Ethnicity				
White	142,111	50.52	14.02	51
African American	58,828	46.67	14.31	45
Asian/Pacific Islander	21,804	57.12	14.45	60
American Indian/Alaska Native	3,970	53.26	14.29	54
Hispanic	83,936	49.52	14.29	50
Multiple	8,719	48.41	14.06	48
Special Instructional Needs				
Economically Disadvantaged	142,342	48.31	14.27	48
Not-economically Disadvantaged	177,039	51.32	14.10	52
English Learner (EL)	14,598	48.68	15.42	48
Non English Learner	304,783	50.04	14.12	50
Students with Disabilities (SWD)	53,372	46.76	14.81	45
Students without Disabilities	266,009	50.62	14.05	51

Table A.13.6 Summary of SGP Estimates for Subgroups: Grade 9 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	54,874	48.96	13.60	49
Female	53,442	50.25	13.60	50
Ethnicity				
White	46,533	50.05	13.45	50
African American	13,062	45.62	13.47	44
Asian/Pacific Islander	9,849	57.50	14.35	60
American Indian/Alaska Native	2,538	53.84	14.07	56
Hispanic	34,994	47.88	13.59	47
Multiple	1,338	51.52	13.71	53
Special Instructional Needs				
Economically Disadvantaged	43,367	47.60	13.58	47
Not-economically Disadvantaged	64,949	50.93	13.61	51
English Learner (EL)	4,201	46.70	15.13	45
Non English Learner	104,115	49.71	13.54	50
Students with Disabilities (SWD)	19,396	45.55	14.04	43
Students without Disabilities	88,920	50.48	13.50	51

Table A.13.7 Summary of SGP Estimates for Subgroups: Grade 10 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	83,002	50.66	12.27	51
Female	81,377	54.36	12.41	56
Ethnicity				
White	69,116	53.87	12.20	56
African American	32,537	48.81	11.90	48
Asian/Pacific Islander	13,480	57.60	12.90	61
American Indian/Alaska Native	2,976	55.28	13.10	58
Hispanic	42,930	50.91	12.77	51
Multiple	3,334	57.09	11.39	61
Special Instructional Needs				
Economically Disadvantaged	59,691	50.23	12.47	50
Not-economically Disadvantaged	104,688	53.78	12.27	56
English Learner (EL)	7,270	49.56	12.94	50
Non English Learner	157,109	52.63	12.32	54
Students with Disabilities (SWD)	27,755	47.28	12.43	46
Students without Disabilities	136,624	53.55	12.33	55

Table A.13.8 Summary of SGP Estimates for Subgroups: Grade 11 ELA/L

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	48,460	46.70	11.70	45
Female	43,817	53.89	11.46	56
Ethnicity				
White	34,379	45.96	10.98	43
African American	17,539	49.68	12.05	49
Asian/Pacific Islander	6,020	49.67	11.16	49
American Indian/Alaska Native	2,396	60.37	12.55	63.5
Hispanic	31,017	54.34	12.02	56
Multiple	917	47.42	11.18	47
Special Instructional Needs				
Economically Disadvantaged	40,118	53.24	12.06	55
Not-economically Disadvantaged	52,159	47.71	11.22	46
English Learner (EL)	4,756	53.87	13.77	55
Non English Learner	87,521	49.91	11.47	50
Students with Disabilities (SWD)	16,554	44.62	12.87	42
Students without Disabilities	75,723	51.32	11.31	52

Table A.13.9 Summary of SGP Estimates for Subgroups: Grade 4 Mathematics

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	166,998	49.72	13.00	50
Female	161,247	50.35	13.01	50
Ethnicity				
White	135,939	52.61	12.87	54
African American	64,202	43.69	13.19	41
Asian/Pacific Islander	22,280	59.54	13.32	63
American Indian/Alaska Native	4,686	45.90	13.30	44
Hispanic	90,286	48.48	12.98	48
Multiple	10,827	50.42	13.09	50
Special Instructional Needs				
Economically Disadvantaged	161,672	46.31	13.05	45
Not-economically Disadvantaged	166,573	53.64	12.96	55
English Learner (EL)	43,449	46.58	13.31	45
Non English Learner	284,796	50.56	12.96	51
Students with Disabilities (SWD)	54,672	45.72	13.38	44
Students without Disabilities	273,573	50.89	12.93	51
Spanish Language Form	2,734	45.64	12.91	44

Table A.13.10 Summary of SGP Estimates for Subgroups: Grade 5 Mathematics

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	168,803	48.80	14.38	48
Female	162,942	51.45	14.42	52
Ethnicity				
White	138,658	50.84	14.15	51
African American	64,031	46.04	14.89	44
Asian/Pacific Islander	21,918	59.13	14.20	63
American Indian/Alaska Native	4,521	49.64	14.78	49
Hispanic	92,259	49.78	14.45	50
Multiple	10,345	49.27	14.43	49
Special Instructional Needs				
Economically Disadvantaged	161,987	47.96	14.62	47
Not-economically Disadvantaged	169,758	52.15	14.18	53
English Learner (EL)	26,376	49.79	15.35	49
Non English Learner	305,369	50.13	14.32	50
Students with Disabilities (SWD)	56,362	49.34	15.05	49
Students without Disabilities	275,383	50.26	14.27	50
Spanish Language Form	2,091	47.15	15.49	46

Table A.13.11 Summary of SGP Estimates for Subgroups: Grade 6 Mathematics

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	163,422	48.89	14.74	48
Female	157,978	51.26	14.72	52
Ethnicity				
White	137,048	51.63	14.24	52
African American	60,867	45.81	15.60	44
Asian/Pacific Islander	21,352	57.84	14.11	61
American Indian/Alaska Native	4,451	49.34	15.60	49
Hispanic	88,086	48.75	15.01	48
Multiple	9,572	49.38	14.63	49
Special Instructional Needs				
Economically Disadvantaged	153,670	47.96	15.25	47
Not-economically Disadvantaged	167,730	51.96	14.25	53
English Learner (EL)	17,795	46.12	16.47	44
Non English Learner	303,605	50.28	14.63	50
Students with Disabilities (SWD)	54,290	44.52	15.84	42
Students without Disabilities	267,110	51.18	14.51	52
Spanish Language Form	1,617	48.67	16.20	48

Table A.13.12 Summary of SGP Estimates for Subgroups: Grade 7 Mathematics

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	156,052	49.34	15.33	49
Female	149,841	50.68	15.24	51
Ethnicity				
White	132,963	51.57	14.83	52
African American	57,621	46.75	16.10	45
Asian/Pacific Islander	18,169	56.64	14.74	59
American Indian/Alaska Native	3,854	42.99	16.34	40
Hispanic	84,870	48.60	15.53	48
Multiple	8,403	50.35	15.20	50
Special Instructional Needs				
Economically Disadvantaged	145,280	47.87	15.77	47
Not-economically Disadvantaged	160,613	51.93	14.85	53
English Learner (EL)	15,816	47.26	17.32	46
Non English Learner	290,077	50.15	15.18	50
Students with Disabilities (SWD)	53,177	44.18	16.73	41
Students without Disabilities	252,716	51.22	14.98	52
Spanish Language Form	1,567	54.96	17.01	59

Table A.13.13 Summary of SGP Estimates for Subgroups: Grade 8 Mathematics

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	127,997	47.96	16.11	47
Female	119,495	52.15	15.97	53
Ethnicity				
White	106,729	50.87	15.42	51
African American	49,102	46.59	17.09	45
Asian/Pacific Islander	10,857	57.04	14.80	60
American Indian/Alaska Native	3,454	50.99	16.92	52
Hispanic	70,773	49.98	16.40	50
Multiple	6,563	48.95	16.09	49
Special Instructional Needs				
Economically Disadvantaged	124,944	48.57	16.64	48
Not-economically Disadvantaged	122,548	51.43	15.44	52
English Learner (EL)	14,378	49.44	18.01	49
Non English Learner	233,114	50.02	15.92	50
Students with Disabilities (SWD)	48,242	45.62	17.63	44
Students without Disabilities	199,250	51.04	15.66	51
Spanish Language Form	1,210	54.52	17.57	56

Table A.13.14 Summary of SGP Estimates for Subgroups: Algebra I

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	86,066	50.42	14.62	51
Female	83,708	49.59	14.61	49
Ethnicity				
White	71,431	52.78	13.69	54
African American	32,368	46.73	16.05	45
Asian/Pacific Islander	15,321	56.25	12.62	58
American Indian/Alaska Native	2,709	45.36	16.95	43
Hispanic	43,981	45.80	15.68	44
Multiple	3,961	52.74	13.86	54
Special Instructional Needs				
Economically Disadvantaged	62,620	46.28	15.83	45
Not-economically Disadvantaged	107,154	52.20	13.90	53
English Learner (EL)	6,470	46.93	17.95	45
Non English Learner	163,304	50.14	14.48	50
Students with Disabilities (SWD)	26,583	46.59	16.78	45
Students without Disabilities	143,191	50.65	14.21	51
Spanish Language Form	1,219	43.79	19.33	41

Table A.13.15 Summary of SGP Estimates for Subgroups: Geometry

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	60,998	49.92	15.23	50
Female	58,966	50.85	15.05	51
Ethnicity				
White	50,937	51.84	14.09	53
African American	18,029	44.65	16.90	43
Asian/Pacific Islander	10,755	58.90	12.47	63
American Indian/Alaska Native	2,785	55.89	16.85	58
Hispanic	35,705	48.08	16.49	47
Multiple	1,751	52.53	13.70	53
Special Instructional Needs				
Economically Disadvantaged	46,661	47.23	16.70	46
Not-economically Disadvantaged	73,303	52.38	14.15	53
English Learner (EL)	5,396	45.11	18.42	43
Non English Learner	114,568	50.63	14.98	51
Students with Disabilities (SWD)	20,433	45.03	17.58	43
Students without Disabilities	99,531	51.48	14.64	52
Spanish Language Form	1,439	43.08	18.42	40

Table A.13.16 Summary of SGP Estimates for Subgroups: Algebra II

	Total Sample Size	Average SGP	Average Standard Error	Median SGP
Gender				
Male	48,272	48.79	15.61	48
Female	48,418	50.83	15.78	51
Ethnicity				
White	41,984	48.36	14.84	48
African American	12,345	48.28	17.14	47
Asian/Pacific Islander	10,134	56.53	13.71	59
American Indian/Alaska Native	2,254	53.86	17.73	55
Hispanic	28,749	49.90	16.88	50
Multiple	1,217	49.78	14.76	50
Special Instructional Needs				
Economically Disadvantaged	34,386	49.62	17.06	49
Not-economically Disadvantaged	62,304	49.91	14.94	50
English Learner (EL)	3,149	48.04	18.71	47
Non English Learner	93,541	49.87	15.59	50
Students with Disabilities (SWD)	13,019	44.48	17.90	42
Students without Disabilities	83,671	50.64	15.35	51
Spanish Language Form	975	36.02	18.30	29

Addendum: Statistical Summary of the Fall/Winter Block 2017 Administration

The addendum presents the results of analyses for the fall/winter block 2017 operational administration. These results are reported separately from the spring 2018 results because fall testing involved a nonrepresentative subset of students testing only for ELA/L grades 9, 10, and 11, as well as Algebra I, Geometry, and Algebra II. Both online and paper test forms were administered for each test.

To organize the addendum, tables are numbered sequentially according to the section represented by the tables. The reader can refer back to the corresponding section in the technical report for related information on the topic. For example, the first addendum table provides participation counts similar to those provided for Section 5; therefore it is numbered ADD.5.1. The second addendum table for Section 5 is numbered ADD.5.2, and so on.

Addendum 5: Test Taker Characteristics

Table ADD.5.1 State Participation in ELA/L Fall 2017 Operational Tests, by Grade

State	Category	English Language Arts/Literacy			
		Total	**Grade 9**	**Grade 10**	**Grade 11**
PARCC	N of Students	38,174	4,513	18,740	14,921
	N of CBT	37,919	4,509	18,691	14,719
	% of CBT	99.3	99.9	99.7	98.6
	N of PBT	255	4	49	202
	% of PBT	0.7	0.1	0.3	1.4
BI	N of Students	91	-	-	91
	% of PARCC Data	0.2	-	-	0.2
	N of CBT	25	-	-	25
	% of CBT	27.5	-	-	27.5
	N of PBT	66	-	-	66
	% of PBT	72.5	-	-	72.5
MD	N of Students	18,547	231	14,654	3,662
	% of PARCC Data	48.6	0.6	38.4	9.6
	N of CBT	18,488	231	14,614	3,643
	% of CBT	99.7	100	99.7	99.5
	N of PBT	59	0	40	19
	% of PBT	0.3	0.0	0.3	0.5
NJ	N of Students	11,741	4,037	3,887	3,817
	% of PARCC Data	30.8	10.6	10.2	10.0
	N of CBT	11,717	4,033	3,878	3,806
	% of CBT	99.8	99.9	99.8	99.7
	N of PBT	24	4	9	11
	% of PBT	0.2	0.1	0.2	0.3
NM	N of Students	7,795	245	199	7,351
	% of PARCC Data	20.4	0.6	0.5	19.3
	N of CBT	7,689	245	199	7,245
	% of CBT	98.6	100	100	98.6
	N of PBT	106	0	0	106
	% of PBT	1.4	0.0	0.0	1.4

Table ADD.5.2 State Participation in Mathematics Fall 2017 Operational Tests, by Grade

State	Category	Mathematics			
		Total	A1	GO	A2
PARCC	N of Students	36,387	17,304	6,306	12,777
	N of CBT	36,154	17,273	6,255	12,626
	% of CBT	99.4	99.8	99.2	98.8
	N of PBT	233	31	51	151
	% of PBT	0.6	0.2	0.8	1.2
BI	N of Students	107	-	26	81
	% of PARCC Data	0.3	-	0.1	0.2
	N of CBT	36	-	12	24
	% of CBT	33.6	-	46.2	29.6
	N of PBT	71	-	14	57
	% of PBT	66.4	-	53.8	70.4
MD	N of Students	16,875	13,126	224	3,525
	% of PARCC Data	46.4	36.1	0.6	9.7
	N of CBT	16,844	13,101	224	3,519
	% of CBT	99.8	99.8	100	99.8
	N of PBT	31	25	0	6
	% of PBT	0.2	0.2	0.0	0.2
NJ	N of Students	11,524	3,747	4,207	3,570
	% of PARCC Data	31.7	10.3	11.6	9.8
	N of CBT	11,502	3,743	4,199	3,560
	% of CBT	99.8	99.9	99.8	99.7
	N of PBT	22	4	8	10
	% of PBT	0.2	0.1	0.2	0.3
NM	N of Students	7,881	431	1,849	5,601
	% of PARCC Data	21.7	1.2	5.1	15.4
	N of CBT	7,772	429	1,820	5,523
	% of CBT	98.6	99.5	98.4	98.6
	N of PBT	109	2	29	78
	% of PBT	1.4	0.5	1.6	1.4

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table ADD.5.3 State Participation in Spanish Mathematics Fall 2017 Operational Tests, by Grade

State	Category	Mathematics			
		Total	A1	GO	A2
PARCC	N of Students	300	130	46	124
	N of CBT	298	128	46	124
	% of CBT	99.3	98.5	100	100
	N of PBT	2	2	-	-
	% of PBT	0.7	1.5	-	-
BI	N of Students				
	% of PARCC Data	-	-	-	-
	N of CBT	-	-	-	-
	% of CBT	-	-	-	-
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-
MD	N of Students	108	103	4	1
	% of PARCC Data	36.0	34.3	1.3	0.3
	N of CBT	106	101	4	1
	% of CBT	98.1	98.1	100	100
	N of PBT	2	2	-	-
	% of PBT	1.9	1.9	-	-
NJ	N of Students	87	27	14	46
	% of PARCC Data	29.0	9.0	4.7	15.3
	N of CBT	87	27	14	46
	% of CBT	100	100	100	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-
NM	N of Students	105	-	28	77
	% of PARCC Data	35.0	-	9.3	25.7
	N of CBT	105	-	28	77
	% of CBT	100	-	100	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table ADD.5.4 All States Combined: Fall 2017 ELA/L Test Takers by Grade and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
9	ALL	4,513	2,307	51.1	2,206	48.9
	CBT	4,509	2,305	51.1	2,204	48.9
	PBT	4	2	50.0	2	50.0
10	ALL	18,740	8,052	43.0	10,688	57.0
	CBT	18,691	8,042	43.0	10,649	57.0
	PBT	49	10	20.4	39	79.6
11	ALL	14,921	6,410	43.0	8,511	57.0
	CBT	14,719	6,332	43.0	8,387	57.0
	PBT	202	78	38.6	124	61.4

Table ADD.5.5 All States Combined: Fall 2017 Mathematics Test Takers by Grade and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
A1	ALL	17,304	8,464	48.9	8,840	51.1
	CBT	17,273	8,456	49.0	8,817	51.0
	PBT	31	8	25.8	23	74.2
GO	ALL	6,306	3,211	50.9	3,095	49.1
	CBT	6,255	3,182	50.9	3,073	49.1
	PBT	51	29	56.9	22	43.1
A2	ALL	12,777	6,415	50.2	6,362	49.8
	CBT	12,626	6,347	50.3	6,279	49.7
	PBT	151	68	45.0	83	55.0

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table ADD.5.6 All States Combined: Fall 2017 Spanish-Language Mathematics Test Takers by Grade and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
A1	ALL	130	69	53.1	61	46.9
	CBT	128	67	52.3	61	47.7
	PBT	2	2	100.0		
GO	ALL	46	24	52.2	22	47.8
	CBT	46	24	52.2	22	47.8
	PBT	-	-	-	-	-
A2	ALL	124	56	45.2	68	54.8
	CBT	124	56	45.2	68	54.8
	PBT	-	-	-	-	-

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table ADD.5.7 Demographic Information for Fall 2017 Grade 9 ELA/L, Overall and by State

Demographic	PARCC	BI	MD	NJ	NM
Econ Dis (%)	37.2	n/a	49.4	35.4	55.9
SWD (%)	18.4	n/a	13.0	19.5	n/r
EL (%)	1.7	n/a	n/r	1.1	8.2
Male (%)	48.9	n/a	40.7	49.4	47.8
Female (%)	51.1	n/a	59.3	50.6	52.2
AmInd/ANat (%)	n/r	n/a	n/a	n/r	n/r
Asian (%)	5.2	n/a	n/r	5.6	n/r
Black/AA (%)	20.9	n/a	18.6	22.2	n/r
Hisp/Lat (%)	21.1	n/a	12.1	19.1	63.3
Wh/Caus (%)	50.3	n/a	62.3	50.9	29.4
NtvHawaii/Pacific (%)	n/r	n/a	n/a	n/r	n/a
Two Or More (%)	2.0	n/a	n/r	1.8	n/r
Unknown (%)	n/r	n/a	n/a	n/r	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.8 Demographic Information for Fall 2017 Grade 10 ELA/L, Overall and by State

Demographic	PARCC	BI	MD	NJ	NM
Econ Dis (%)	35.3	n/a	34.1	39.4	35.7
SWD (%)	24.5	n/a	25.7	20.7	12.6
EL (%)	8.1	n/a	10.0	1.1	n/r
Male (%)	57.0	n/a	58.8	50.5	53.8
Female (%)	43.0	n/a	41.2	49.5	46.2
AmInd/ANat (%)	0.3	n/a	0.3	n/r	n/r
Asian (%)	3.5	n/a	2.6	6.9	n/r
Black/AA (%)	40.0	n/a	44.8	24.0	n/r
Hisp/Lat (%)	17.6	n/a	16.4	20.5	46.7
Wh/Caus (%)	35.6	n/a	32.6	46.7	38.7
NtvHawaii/Pacific (%)	0.1	n/a	n/r	n/r	n/a
Two Or More (%)	2.9	n/a	3.2	1.7	n/r
Unknown (%)	n/r	n/a	n/a	n/r	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.9 Demographic Information for Fall 2017 Grade 11 ELA/L, Overall and by State

Demographic	PARCC	BI	MD	NJ	NM
Econ Dis (%)	48.0	90.1	34.3	34.8	61.2
SWD (%)	20.9	n/r	19.5	19.2	22.6
EL (%)	9.9	46.2	4.8	1.0	16.6
Male (%)	57.0	52.7	58.5	52.3	58.8
Female (%)	43.0	47.3	41.5	47.7	41.2
AmInd/ANat (%)	7.0	100	n/r	n/r	12.7
Asian (%)	2.3	n/a	2.6	4.5	1.0
Black/AA (%)	19.4	n/a	46.9	25.5	2.9
Hisp/Lat (%)	39.8	n/a	11.8	18.3	65.4
Wh/Caus (%)	29.0	n/a	34.9	48.8	16.1
NtvHawaii/Pacific (%)	n/r	n/a	n/r	n/r	n/r
Two Or More (%)	1.9	n/a	3.6	2.3	1.0
Unknown (%)	0.5	n/a	n/a	n/r	0.9

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.10 Demographic Information for Fall 2017 Algebra I, Overall and by State

Demographic	PARCC	BI	MD	NJ	NM
Econ Dis (%)	35.0	n/a	33.7	37.2	53.4
SWD (%)	19.9	n/a	21.7	14.0	17.4
EL (%)	9.8	n/a	11.5	3.8	10.4
Male (%)	51.1	n/a	51.7	48.5	53.4
Female (%)	48.9	n/a	48.3	51.5	46.6
AmInd/ANat (%)	0.4	n/a	0.2	n/r	6.3
Asian (%)	4.0	n/a	3.3	6.9	n/r
Black/AA (%)	40.0	n/a	46.0	23.2	n/r
Hisp/Lat (%)	19.7	n/a	17.1	23.0	66.8
Wh/Caus (%)	32.8	n/a	30.0	43.7	23.0
NtvHawaii/Pacific (%)	0.2	n/a	0.2	n/r	n/r
Two Or More (%)	2.9	n/a	3.1	2.4	n/r
Unknown (%)	n/r	n/a	n/a	n/r	n/r

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.11 Demographic Information for Fall 2017 Geometry, Overall and by State

Demographic	PARCC	BI	MD	NJ	NM
Econ Dis (%)	41.8	100	48.7	37.1	50.9
SWD (%)	19.2	n/r	14.3	19.0	20.5
EL (%)	7.0	n/r	n/r	3.0	15.6
Male (%)	49.1	n/r	46.9	48.4	51.1
Female (%)	50.9	n/r	53.1	51.6	48.9
AmInd/ANat (%)	3.3	100	n/a	n/r	9.4
Asian (%)	5.5	n/a	n/r	7.8	n/r
Black/AA (%)	16.8	n/a	26.8	22.7	2.5
Hisp/Lat (%)	33.0	n/a	12.5	19.6	66.5
Wh/Caus (%)	38.7	n/a	53.1	47.4	17.8
NtvHawaii/Pacific (%)	n/r	n/a	n/a	n/r	n/r
Two Or More (%)	1.8	n/a	n/r	2.0	n/r
Unknown (%)	0.6	n/a	n/a	n/r	2.1

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.12 Demographic Information for Fall 2017 Algebra II, Overall and by State

Demographic	PARCC	BI	MD	NJ	NM
Econ Dis (%)	45.3	76.5	31.9	33.4	60.9
SWD (%)	12.4	n/r	11.1	14.7	11.8
EL (%)	8.1	63.0	1.9	2.8	14.6
Male (%)	49.8	55.6	51.3	48.8	49.4
Female (%)	50.2	44.4	48.7	51.2	50.6
AmInd/ANat (%)	7.7	98.8	n/r	n/r	15.9
Asian (%)	3.6	n/a	3.5	8.1	0.8
Black/AA (%)	18.0	n/a	39.0	21.9	2.6
Hisp/Lat (%)	35.7	n/a	11.3	20.1	61.4
Wh/Caus (%)	32.6	n/a	42.2	47.9	17.3
NtvHawaii/Pacific (%)	n/r	n/a	n/r	n/r	n/r
Two Or More (%)	1.8	n/r	3.7	1.5	0.8
Unknown (%)	0.5	n/a	n/a	n/r	1.0

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Addendum 8: Reliability

Table ADD.8.1 shows the total group level reliability estimates and raw score SEM for the fall 2017 forms. Tables ADD.8.2 – ADD.8.7 show the subgroup reliability estimates and raw score SEM. A minimum sample size of 100 per core form was required for calculating the reliability estimates for subgroups; therefore, the subgroup totals may not equal the total group sample size. Tables ADD.8.8 – ADD.8.10 provide the claim and subclaim reliability and raw score SEM estimates for the fall 2017 forms. The paper-based tests did not have sufficient sample sizes for reliability analyses.

Table ADD.8.1 Summary of Test Reliability Estimates for Fall 2017 Total Group

Content	Grade/ Course	Mode	Number of forms	Maximum Possible Score	Avg. Raw Score SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
ELA/L	9	CBT	2	121	6.03	0.94	160	0.90	4,287	0.94
ELA/L	10	CBT	2	121	5.71	0.94	203	0.89	13,646	0.94
ELA/L	11	CBT	2	121	5.44	0.94	120	0.87	8,456	0.94
Mathematics	A1	CBT	2	81	3.45	0.92	514	0.87	8,589	0.93
Mathematics	GO	CBT	2	81	3.30	0.94	683	0.93	3,767	0.95
Mathematics	A2	CBT	2	81	3.34	0.93	573	0.91	6,848	0.93

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II. ELA/L grades 9, 10, and 11, Algebra I, Geometry, and Algebra II had insufficient sample sizes for PBT.

Table ADD.8.2 Summary of Test Reliability Estimates for Fall 2017 Subgroups: Grade 9 ELA/L

	CBT						
	Max. Raw Score	**Avg. SEM**	**Avg. Relia-bility**	**Min. Sample Size**	**Min. Relia-bility**	**Max. Sample Size**	**Max. Relia-bility**
Gender							
Male	121	5.83	0.94	100	0.90	2,067	0.94
Female	121	6.23	0.93	2,220	0.93	2,220	0.93
Ethnicity							
White	121	6.15	0.93	2,204	0.93	2,204	0.93
African American	121	5.92	0.92	844	0.92	844	0.92
Asian/Pacific Islander	121	6.31	0.93	229	0.93	229	0.93
Am. Indian/Alaska Native							
Hispanic	121	5.93	0.92	902	0.92	902	0.92
Multiple							
Special Instructional Needs							
Economically Disadvantaged	121	5.88	0.93	1,552	0.93	1,552	0.93
Not Economically Disadvantaged	121	6.17	0.93	2,704	0.93	2,704	0.93
English Learner							
Non-English Learner	121	6.04	0.93	159	0.90	4,137	0.94
Students with Disabilities	121	5.31	0.92	160	0.90	649	0.93
Students w/o Disabilities	121	6.17	0.93	3,539	0.93	3,539	0.93
Students Taking Accommodated Forms							
A: ASL							
C: Closed Caption							
R: Screen Reader							
T: Text-to-Speech	121	4.85	0.90	159	0.90	159	0.90

Table ADD.8.3 Summary of Test Reliability Estimates for Fall 2017 Subgroups: Grade 10 ELA/L

	CBT						
	Max. Raw Score	**Avg. SEM**	**Avg. Relia-bility**	**Min. Sample Size**	**Min. Relia-bility**	**Max. Sample Size**	**Max. Relia-bility**
Gender							
Male	121	5.49	0.94	142	0.87	7,533	0.94
Female	121	5.98	0.94	6,113	0.94	6,113	0.94
Ethnicity							
White	121	6.08	0.94	5,581	0.94	5,581	0.94
African American	121	5.29	0.90	4,508	0.90	4,508	0.90
Asian/Pacific Islander	121	6.13	0.95	543	0.95	543	0.95
Am. Indian/Alaska Native							
Hispanic	121	5.47	0.93	2,528	0.93	2,528	0.93
Multiple	121	5.95	0.94	423	0.94	423	0.94
Special Instructional Needs							
Economically Disadvantaged	121	5.48	0.92	5,038	0.92	5,038	0.92
Not Economically Disadvantaged	121	5.84	0.95	105	0.93	8,564	0.95
English Learner	121	4.39	0.78	1,079	0.78	1,079	0.78
Non-English Learner	121	5.79	0.94	190	0.89	12,476	0.94
Students with Disabilities	121	5.11	0.91	203	0.89	2,888	0.91
Students w/o Disabilities	121	5.86	0.94	10,655	0.94	10,655	0.94
Students Taking Accommodated Forms							
A: ASL							
C: Closed Caption							
R: Screen Reader							
T: Text-to-Speech	121	4.51	0.83	192	0.83	192	0.83

Table ADD.8.4 Summary of Test Reliability Estimates for Fall 2017 Subgroups: Grade 11 ELA/L

	CBT						
	Max. Raw Score	Avg. SEM	Avg. Reliability	Min. Sample Size	Min. Reliability	Max. Sample Size	Max. Reliability
Gender							
Male	121	5.18	0.93	4,550	0.93	4,550	0.93
Female	121	5.73	0.94	3,906	0.94	3,906	0.94
Ethnicity							
White	121	5.72	0.94	3,328	0.94	3,328	0.94
African American	121	5.06	0.91	2,536	0.91	2,536	0.91
Asian/Pacific Islander	121	5.74	0.96	287	0.96	287	0.96
Am. Indian/Alaska Native	121	5.49	0.92	186	0.92	186	0.92
Hispanic	121	5.37	0.93	1,843	0.93	1,843	0.93
Multiple	121	5.41	0.93	223	0.93	223	0.93
Special Instructional Needs							
Economically Disadvantaged	121	5.33	0.93	2,937	0.93	2,937	0.93
Not Economically Disadvantaged	121	5.51	0.94	5,466	0.94	5,466	0.94
English Learner	121	4.51	0.80	352	0.80	352	0.80
Non-English Learner	121	5.46	0.94	116	0.87	8,022	0.94
Students with Disabilities	121	4.87	0.92	120	0.87	1,363	0.92
Students w/o Disabilities	121	5.53	0.94	6,979	0.94	6,979	0.94
Students Taking Accommodated Forms							
A: ASL							
C: Closed Caption							
R: Screen Reader							
T: Text-to-Speech	121	4.40	0.86	118	0.86	118	0.86

Table ADD.8.5 Summary of Test Reliability Estimates for Fall 2017 Subgroups: Algebra I

	CBT						
	Max. Raw Score	**Avg. SEM**	**Avg. Relia-bility**	**Min. Sample Size**	**Min. Relia-bility**	**Max. Sample Size**	**Max. Relia-bility**
Gender							
Male	81	3.44	0.93	286	0.87	4,268	0.93
Female	81	3.44	0.92	228	0.86	4,321	0.92
Ethnicity							
White	81	3.73	0.92	187	0.86	3,666	0.92
African American	81	2.94	0.88	126	0.84	2,699	0.88
Asian/Pacific Islander	81	4.10	0.94	380	0.94	380	0.94
Am. Indian/Alaska Native							
Hispanic	81	3.20	0.90	162	0.82	1,516	0.91
Multiple	81	3.66	0.94	265	0.94	265	0.94
Special Instructional Needs							
Economically Disadvantaged	81	3.10	0.89	246	0.83	3,127	0.89
Not Economically Disadvantaged	81	3.62	0.93	268	0.89	5,422	0.93
English Learner	81	2.77	0.87	103	0.82	540	0.88
Non-English Learner	81	3.48	0.92	411	0.87	7,976	0.93
Students with Disabilities	81	2.88	0.88	216	0.76	1,447	0.90
Students w/o Disabilities	81	3.54	0.92	298	0.87	7,033	0.93
Students Taking Accommodated Forms							
A: ASL							
C: Closed Caption							
R: Screen Reader							
T: Text-to-Speech	80	3.67	0.87	479	0.87	479	0.87
Students Taking Translated Forms							
Spanish-Language Form							

Table ADD.8.6 Summary of Test Reliability Estimates for Fall 2017 Subgroups: Geometry

	CBT						
	Max. Raw Score	**Avg. SEM**	**Avg. Relia-bility**	**Min. Sample Size**	**Min. Relia-bility**	**Max. Sample Size**	**Max. Relia-bility**
Gender							
Male	81	3.25	0.95	377	0.93	1,771	0.96
Female	81	3.35	0.93	306	0.92	1,996	0.94
Ethnicity							
White	81	3.48	0.93	386	0.92	1,678	0.94
African American	81	2.87	0.90	716	0.90	121	0.91
Asian/Pacific Islander	81	4.06	0.96	275	0.96	275	0.96
Am. Indian/Alaska Native							
Hispanic	81	2.90	0.91	131	0.88	958	0.92
Multiple							
Special Instructional Needs							
Economically Disadvantaged	81	2.92	0.91	225	0.90	1,387	0.91
Not Economically Disadvantaged	81	3.49	0.95	458	0.93	2,340	0.95
English Learner	81	2.31	0.50	115	0.50	115	0.50
Non-English Learner	81	3.33	0.94	626	0.93	3,573	0.95
Students with Disabilities	81	2.64	0.91	209	0.88	622	0.92
Students w/o Disabilities	81	3.43	0.94	474	0.92	3,049	0.95
Students Taking Accommodated Forms							
A: ASL							
C: Closed Caption							
R: Screen Reader							
T: Text-to-Speech	81	3.35	0.93	661	0.93	661	0.93
Students Taking Translated Forms							
Spanish-Language Form							

Table ADD.8.7 Summary of Test Reliability Estimates for Fall 2017 Subgroups: Algebra II

	CBT						
	Max. Raw Score	**Avg. SEM**	**Avg. Relia-bility**	**Min. Sample Size**	**Min. Relia-bility**	**Max. Sample Size**	**Max. Relia-bility**
Gender							
Male	81	3.30	0.94	277	0.93	3,395	0.94
Female	81	3.38	0.92	296	0.89	3,453	0.93
Ethnicity							
White	81	3.65	0.93	305	0.90	2,709	0.93
African American	81	2.87	0.81	1,898	0.81	100	0.86
Asian/Pacific Islander	81	3.79	0.95	387	0.95	387	0.95
Am. Indian/Alaska Native							
Hispanic	81	3.02	0.87	121	0.84	1,543	0.87
Multiple	81	3.41	0.93	167	0.93	167	0.93
Special Instructional Needs							
Economically Disadvantaged	81	3.02	0.87	2,323	0.87	200	0.89
Not Economically Disadvantaged	81	3.48	0.94	373	0.91	4,453	0.94
English Learner	81	2.60	0.76	170	0.75	100	0.79
Non-English Learner	81	3.37	0.93	473	0.91	6,569	0.93
Students with Disabilities	81	2.91	0.88	117	0.68	768	0.92
Students w/o Disabilities	81	3.39	0.93	456	0.91	5,952	0.93
Students Taking Accommodated Forms							
A: ASL							
C: Closed Caption							
R: Screen Reader							
T: Text-to-Speech	81	3.37	0.91	527	0.91	527	0.91
Students Taking Translated Forms							
Spanish-Language Form							

Table ADD.8.8 Fall 2017 Average ELA/L Reliability Estimates for Reading Total and Subscores

		Reading: Total		Reading: Literature		Reading: Information		Reading: Vocabulary	
Grade Level	**Mode**	**Max Possible Raw Score**	**Reliability**	**Max Possible Raw Score**	**Reliability**	**Max Possible Raw Score**	**Reliability**	**Max Possible Raw Score**	**Reliability**
9	**CBT**	76	0.90	28	0.75	36	0.82	12	0.60
10	**CBT**	76	0.91	26	0.81	30	0.80	20	0.66
11	**CBT**	76	0.91	24	0.80	30	0.81	22	0.68

Table ADD.8.9 Fall 2017 Average ELA/L Reliability Estimates for Writing Total and Subscores

		Writing: Total		Writing: Written Expression		Writing: Knowledge Language and Conventions	
Grade Level	**Mode**	**Max Possible Raw Score**	**Reliability**	**Max Possible Raw Score**	**Reliability**	**Max Possible Raw Score**	**Reliability**
9	**CBT**	45	0.88	36	0.88	9	0.90
10	**CBT**	45	0.89	36	0.89	9	0.90
11	**CBT**	45	0.88	36	0.87	9	0.88

Table ADD.8.10 Fall 2017 Average Mathematics Reliability Estimates for Total Test and Subscores

Grade Level	Mode	Major Content		Additional & Supporting Content		Mathematics Reasoning		Modeling Practice	
		Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability
A1	**CBT**	26	0.79	15	0.62	14	0.66	18	0.76
GO	**CBT**	30	0.87	19	0.78	14	0.77	18	0.68
A2	**CBT**	27	0.81	20	0.77	14	0.74	18	0.66

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Tables ADD.8.11 and ADD.8.12 provide information about the accuracy and the consistency of two classifications made on the basis of the scores on the fall block 2017 English language arts/literacy and mathematics assessments, respectively. The columns labeled “Exact level” provide the classification of the student into one of five achievement levels. The columns labeled “Level 4 or higher vs. 3 or lower” provide the classification of the student as being either in one of the upper two levels (Levels 4 and 5) or in one of the lower three levels (Levels 1, 2, and 3).

Tables ADD.8.13 to ADD.8.18 provide more detailed information about the accuracy and the consistency of the classification of students into proficiency levels for each fall block 2017 assessment. Each cell in the 5-by-5 table shows the estimated proportion of students who would be classified into a particular combination of proficiency levels. The sum of the five ***bold*** values on the diagonal should equal the exact level of decision accuracy or consistency presented in Tables ADD.8.11or ADD.8.12 for the corresponding assessment. For “Level 4 and higher vs. 3 and lower” found in Tables ADD.8.11 or ADD.8.12, the sum of the shaded values in Tables ADD.8.13 to ADD.8.18 should equal the level of decision accuracy or consistency for the corresponding assessment in ADD.8.11 or ADD.8.12. Note that the sums based on values may not match exactly to the values due to truncation and rounding.

Table ADD.8.11 Reliability of Classification: Summary for ELA/L Fall Block 2017

		Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
Grade Level	Testing Mode	Exact level	Level 4 or higher vs. 3 or lower	Exact level	Level 4 or higher vs. 3 or lower
9	CBT	0.78	0.93	0.70	0.90
	PBT	--	--	--	--
10	CBT	0.76	0.94	0.68	0.92
	PBT	--	--	--	--
11	CBT	0.79	0.95	0.71	0.93
	PBT	--	--	--	--

Table ADD.8.12 Reliability of Classification: Summary for Mathematics Fall Block 2017

		Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
Grade Level	Testing Mode	Exact Level	Level 4 or higher vs. 3 or lower	Exact Level	Level 4 or higher vs. 3 or lower
A1	CBT	0.77	0.94	0.68	0.92
	PBT	--	--	--	--
GO	CBT	0.81	0.95	0.73	0.93
	PBT	--	--	--	--
A2	CBT	0.80	0.96	0.73	0.94
	PBT	--	--	--	--

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II. "--" means insufficient sample size (< 100 students).

Table ADD.8.13 Reliability of Classification: Grade 9 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.10**	0.01	0.00	0.00	0.00	0.11
		700 – 724	0.02	**0.10**	0.03	0.00	0.00	0.15
		725 – 749	0.00	0.03	**0.19**	0.04	0.00	0.25
		750 – 809	0.00	0.00	0.04	**0.32**	0.03	0.39
		810 – 850	0.00	0.00	0.00	0.02	**0.07**	0.09
	Decision Consistency	650 – 699	**0.10**	0.02	0.00	0.00	0.00	0.12
		700 – 724	0.02	**0.09**	0.05	0.00	0.00	0.16
		725 – 749	0.00	0.03	**0.16**	0.05	0.00	0.24
		750 – 809	0.00	0.00	0.06	**0.29**	0.03	0.38
		810 – 850	0.00	0.00	0.00	0.04	**0.06**	0.10

Table ADD.8.14 Reliability of Classification: Grade 10 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.32**	0.04	0.00	0.00	0.00	0.37
		700 – 724	0.04	**0.15**	0.04	0.00	0.00	0.22
		725 – 749	0.00	0.04	**0.11**	0.03	0.00	0.17
		750 – 809	0.00	0.00	0.03	**0.13**	0.02	0.18
		810 – 850	0.00	0.00	0.00	0.01	**0.05**	0.06
	Decision Consistency	650 – 699	**0.31**	0.06	0.00	0.00	0.00	0.37
		700 – 724	0.04	**0.12**	0.05	0.00	0.00	0.21
		725 – 749	0.00	0.05	**0.08**	0.03	0.00	0.17
		750 – 809	0.00	0.00	0.04	**0.12**	0.02	0.18
		810 – 850	0.00	0.00	0.00	0.02	**0.05**	0.07

Table ADD.8.15 Reliability of Classification: Grade 11 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.31**	0.04	0.00	0.00	0.00	0.35
		700 – 724	0.04	**0.19**	0.04	0.00	0.00	0.27
		725 – 749	0.00	0.03	**0.14**	0.02	0.00	0.20
		750 – 809	0.00	0.00	0.02	**0.12**	0.01	0.15
		810 – 850	0.00	0.00	0.00	0.00	**0.03**	0.03
	Decision Consistency	650 – 699	**0.30**	0.05	0.00	0.00	0.00	0.35
		700 – 724	0.05	**0.16**	0.05	0.00	0.00	0.26
		725 – 749	0.00	0.05	**0.12**	0.03	0.00	0.20
		750 – 809	0.00	0.00	0.03	**0.11**	0.01	0.16
		810 – 850	0.00	0.00	0.00	0.01	**0.03**	0.04

Table ADD.8.16 Reliability of Classification: Algebra I

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.19**	0.04	0.00	0.00	0.00	0.23
		700 – 724	0.04	**0.26**	0.04	0.00	0.00	0.34
		725 – 749	0.00	0.04	**0.17**	0.03	0.00	0.24
		750 – 809	0.00	0.00	0.02	**0.15**	0.00	0.18
		810 – 850	0.00	0.00	0.00	0.00	**0.01**	0.01
	Decision Consistency	650 – 699	**0.18**	0.06	0.00	0.00	0.00	0.24
		700 – 724	0.05	**0.22**	0.05	0.00	0.00	0.33
		725 – 749	0.00	0.06	**0.14**	0.04	0.00	0.24
		750 – 809	0.00	0.00	0.04	**0.14**	0.00	0.19
		810 – 850	0.00	0.00	0.00	0.00	**0.01**	0.01

Table ADD.8.17 Reliability of Classification: Geometry

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.10**	0.03	0.00	0.00	0.00	0.13
		700 – 724	0.02	**0.34**	0.03	0.00	0.00	0.40
		725 – 749	0.00	0.05	**0.19**	0.03	0.00	0.27
		750 – 809	0.00	0.00	0.02	**0.15**	0.01	0.18
		810 – 850	0.00	0.00	0.00	0.00	**0.02**	0.02
	Decision Consistency	650 – 699	**0.10**	0.05	0.00	0.00	0.00	0.15
		700 – 724	0.03	**0.30**	0.05	0.00	0.00	0.37
		725 – 749	0.00	0.07	**0.17**	0.04	0.00	0.28
		750 – 809	0.00	0.00	0.03	**0.14**	0.01	0.18
		810 – 850	0.00	0.00	0.00	0.01	**0.02**	0.02

Table ADD.8.18 Reliability of Classification: Algebra II

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	**0.44**	0.05	0.00	0.00	0.00	0.49
		700 – 724	0.04	**0.17**	0.03	0.00	0.00	0.24
		725 – 749	0.00	0.03	**0.07**	0.02	0.00	0.13
		750 – 809	0.00	0.00	0.01	**0.11**	0.00	0.13
		810 – 850	0.00	0.00	0.00	0.00	**0.01**	0.01
	Decision Consistency	650 – 699	**0.42**	0.07	0.00	0.00	0.00	0.49
		700 – 724	0.05	**0.14**	0.03	0.00	0.00	0.22
		725 – 749	0.00	0.05	**0.06**	0.03	0.00	0.14
		750 – 809	0.00	0.00	0.02	**0.10**	0.00	0.13
		810 – 850	0.00	0.00	0.00	0.00	**0.01**	0.01

Addendum 9: Validity

The intercorrelations for the fall 2017 assessments are presented in Tables ADD.9.1 through ADD.9.3 for ELA/L grades 9, 10, and 11 and Tables ADD.9.4 though ADD.9.6 for the traditional mathematics courses (A1, GO, A2). Like the spring intercorrelations, the ELA/L all have moderate to high values with the writing subclaims being highly intercorrelated. The mathematics intercorrelations have moderate values. Tables ADD.9.7 through ADD.9.9 are the correlations between ELA/L and mathematics from the fall block.

Table ADD.9.1 Average Intercorrelations and Reliability between Grade 9 ELA/L Subclaims

	CBT						
	RD	RL	RI	RV	WR	WE	WKL
RD	0.90	4,474	4,474	4,474	4,474	4,474	4,474
RL	0.91	0.75	4,474	4,474	4,474	4,474	4,474
RI	0.95	0.77	0.82	4,474	4,474	4,474	4,474
RV	0.81	0.65	0.69	0.60	4,474	4,474	4,474
WR	0.75	0.71	0.72	0.53	0.88	4,474	4,474
WE	0.74	0.70	0.71	0.53	1.00	0.88	4,474
WKL	0.75	0.71	0.72	0.54	0.98	0.96	0.90

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Addendum 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.2 Average Intercorrelations and Reliability between Grade 10 ELA/L Subclaims

	CBT						
	RD	RL	RI	RV	WR	WE	WKL
RD	0.91	13,954	13,954	13,954	13,954	13,954	13,954
RL	0.93	0.81	13,954	13,954	13,954	13,954	13,954
RI	0.93	0.79	0.80	13,954	13,954	13,954	13,954
RV	0.86	0.71	0.70	0.66	13,954	13,954	13,954
WR	0.80	0.76	0.76	0.63	0.89	13,954	13,954
WE	0.79	0.75	0.76	0.62	1.00	0.89	13,954
WKL	0.79	0.75	0.76	0.63	0.98	0.97	0.90

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Addendum 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.3 Average Intercorrelations and Reliability between Grade 11 ELA/L Subclaims

	CBT						
	RD	RL	RI	RV	WR	WE	WKL
RD	0.91	8,710	8,710	8,710	8,710	8,710	8,710
RL	0.92	0.80	8,710	8,710	8,710	8,710	8,710
RI	0.94	0.81	0.81	8,710	8,710	8,710	8,710
RV	0.88	0.73	0.73	0.68	8,710	8,710	8,710
WR	0.79	0.77	0.75	0.62	0.88	8,710	8,710
WE	0.78	0.77	0.75	0.62	1.00	0.87	8,710
WKL	0.78	0.77	0.75	0.62	0.98	0.97	0.88

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Addendum 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.4 Average Intercorrelations and Reliability between Algebra I Subclaims

	CBT			
	MC	ASC	MR	MP
MC	0.79	10,157	10,157	10,157
ASC	0.72	0.62	10,157	10,157
MR	0.74	0.67	0.66	10,157
MP	0.76	0.66	0.70	0.76

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Addendum 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.5 Average Intercorrelations and Reliability between Geometry Subclaims

	CBT			
	MC	ASC	MR	MP
MC	0.87	5,072	5,072	5,072
ASC	0.83	0.78	5,072	5,072
MR	0.74	0.70	0.77	5,072
MP	0.73	0.69	0.79	0.68

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Addendum 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.6 Average Intercorrelations and Reliability between Algebra II Subclaims

	CBT			
	MC	ASC	MR	MP
MC	0.81	7,715	7,715	7,715
ASC	0.81	0.77	7,715	7,715
MR	0.75	0.75	0.74	7,715
MP	0.76	0.76	0.76	0.66

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Addendum 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.7 Average Correlations between ELA/L and Mathematics for High School

ELA/L	CBT		
	A1	GO	A2
9	0.70	0.73	
	(969)	(369)	
10	0.51	0.70	0.80
	(1,600)	(995)	(802)
11	0.35	0.52	0.53
	(194)	(1,442)	(4,397)

Note: ELA/L = English language arts/literacy, A1 = Algebra I, GO = Geometry, A2 = Algebra II. The correlations are provided with the sample sizes, below in parentheses.

Table ADD.9.8 Average Correlations between Reading and Mathematics for High School

RD	CBT		
	A1	GO	A2
9	0.71	0.74	
	(969)	(369)	
10	0.50	0.71	0.80
	(1,600)	(995)	(802)
11	0.36	0.52	0.54
	(194)	(1,442)	(4,397)

Note: RD = Reading, A1 = Algebra I, GO = Geometry, A2 = Algebra II. The correlations are provided with the sample sizes, below in parentheses.

Table ADD.9.9 Average Correlations between Writing and Mathematics for High School

WR	CBT		
	A1	GO	A2
9	0.58	0.59	
	(969)	(369)	
10	0.41	0.59	0.71
	(1,600)	(995)	(802)
11	0.23	0.40	0.39
	(194)	(1,442)	(4,397)

Note: WR = Writing, A1 = Algebra I, GO = Geometry, A2 = Algebra II. The average correlations are provided with the sample sizes, below in parentheses.

Addendum 12: Scale Scores

Table ADD.12.1 Fall 2017 Subgroup Performance for ELA/L Scale Scores: Grade 9

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**4,513**	**745.48**	**35.71**	**650**	**850**
Gender	Female	2,307	752.91	34.30	650	850
	Male	2,206	737.71	35.51	650	850
Ethnicity	American Indian/Alaska Native	n/r	n/r	n/r	n/r	n/r
	Asian	233	764.74	32.40	675	850
	Black or African American	941	729.45	33.12	650	830
	Hispanic/Latino	953	734.61	33.51	650	833
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	89	744.33	32.93	652	825
	White	2,272	754.96	34.08	650	850
Economic Status*	Economically Disadvantaged	1,681	732.54	34.54	650	850
	Not Economically Disadvantaged	2,801	753.23	34.22	650	850
English Learner Status	English Learner (EL)	78	695.26	27.85	650	768
	Non English Learner	4,359	746.32	35.36	650	850
Disabilities	Students with Disabilities (SWD)	830	715.04	33.36	650	817
	Students without Disabilities	3,582	752.47	32.71	650	850
Reading Score		**4,513**	**47.79**	**14.14**	**10**	**90**
Gender	Female	2,307	49.77	13.79	10	90
	Male	2,206	45.73	14.22	10	90
Ethnicity	American Indian/Alaska Native	n/r	n/r	n/r	n/r	n/r
	Asian	233	54.77	13.26	15	90
	Black or African American	941	41.54	12.90	10	84
	Hispanic/Latino	953	43.20	13.18	10	90
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	89	48.06	13.93	11	80
	White	2,272	51.65	13.53	10	90
Economic Status*	Economically Disadvantaged	1,681	42.71	13.55	10	90
	Not Economically Disadvantaged	2,801	50.86	13.62	10	90
English Learner Status	English Learner (EL)	78	27.62	10.71	10	59
	Non English Learner	4,359	48.21	13.99	10	90
Disabilities	Students with Disabilities (SWD)	830	36.66	13.39	10	79
	Students without Disabilities	3,582	50.42	13.12	10	90

Group Type	Group	*N*	Mean	*SD*	Min	Max
Writing Score		**4,513**	**32.46**	**11.51**	**10**	**60**
Gender	Female	2,307	35.59	10.01	10	60
	Male	2,206	29.20	12.05	10	60
Ethnicity	American Indian/Alaska Native	n/r	n/r	n/r	n/r	n/r
	Asian	233	38.16	9.28	10	60
	Black or African American	941	28.21	11.64	10	53
	Hispanic/Latino	953	30.03	11.44	10	53
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	89	32.38	9.64	10	50
	White	2,272	34.76	10.90	10	60
Economic Status*	Economically Disadvantaged	1,681	28.98	11.87	10	60
	Not Economically Disadvantaged	2,801	34.52	10.79	10	60
English Learner Status	English Learner (EL)	78	19.99	10.64	10	39
	Non English Learner	4,359	32.59	11.45	10	60
Disabilities	Students with Disabilities (SWD)	830	23.10	11.84	10	53
	Students without Disabilities	3,582	34.53	10.38	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.2 Fall 2017 Subgroup Performance for ELA/L Scale Scores: Grade 10

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**18,740**	**719.61**	**44.35**	**650**	**850**
Gender	Female	8,052	728.50	46.29	650	850
	Male	10,688	712.92	41.60	650	850
Ethnicity	American Indian/Alaska Native	60	718.73	37.43	650	808
	Asian	651	748.07	53.11	650	850
	Black or African American	7,500	703.01	33.23	650	847
	Hispanic/Latino	3,294	708.35	40.21	650	850
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	537	727.70	45.20	650	850
	White	6,675	740.38	46.26	650	850
Economic Status*	Economically Disadvantaged	6,607	707.95	37.94	650	850
	Not Economically Disadvantaged	12,089	725.95	46.29	650	850
English Learner Status	English Learner (EL)	1,525	683.43	25.69	650	776
	Non English Learner	17,124	722.74	44.27	650	850
Disabilities	Students with Disabilities (SWD)	4,591	697.77	35.26	650	847
	Students without Disabilities	14,046	726.65	44.76	650	850
Reading Score		**18,740**	**38.83**	**17.23**	**10**	**90**
Gender	Female	8,052	41.11	17.87	10	90
	Male	10,688	37.11	16.53	10	90
Ethnicity	American Indian/Alaska Native	60	39.02	14.74	10	70
	Asian	651	48.59	20.79	10	90
	Black or African American	7,500	33.03	13.33	10	90
	Hispanic/Latino	3,294	34.08	15.64	10	90
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	537	42.24	17.90	10	90
	White	6,675	46.44	18.05	10	90
Economic Status*	Economically Disadvantaged	6,607	34.23	14.75	10	90
	Not Economically Disadvantaged	12,089	41.33	17.97	10	90
English Learner Status	English Learner (EL)	1,525	24.62	9.99	10	63
	Non English Learner	17,124	40.07	17.19	10	90
Disabilities	Students with Disabilities (SWD)	4,591	30.76	14.19	10	90
	Students without Disabilities	14,046	41.44	17.35	10	90
Writing Score		**18,740**	**25.56**	**13.13**	**10**	**60**
Gender	Female	8,052	29.12	13.28	10	60
	Male	10,688	22.87	12.36	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	60	24.80	11.95	10	49
	Asian	651	33.85	14.29	10	60
	Black or African American	7,500	20.62	10.77	10	60
	Hispanic/Latino	3,294	23.38	12.17	10	60
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	537	27.39	13.02	10	60
	White	6,675	31.22	13.30	10	60
Economic Status*	Economically Disadvantaged	6,607	22.76	11.91	10	60
	Not Economically Disadvantaged	12,089	27.07	13.52	10	60
English Learner Status	English Learner (EL)	1,525	16.87	9.10	10	43
	Non English Learner	17,124	26.29	13.16	10	60
Disabilities	Students with Disabilities (SWD)	4,591	19.57	10.95	10	60
	Students without Disabilities	14,046	27.47	13.21	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.3 Fall 2017 Subgroup Performance for ELA/L Scale Scores: Grade 11

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**14,921**	**718.58**	**35.34**	**650**	**850**
Gender	Female	6,410	725.59	36.85	650	850
	Male	8,511	713.30	33.19	650	850
Ethnicity	American Indian/Alaska Native	1,046	715.29	28.52	650	845
	Asian	338	740.77	46.73	651	850
	Black or African American	2,901	711.03	32.56	650	845
	Hispanic/Latino	5,937	710.80	29.80	650	850
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	290	722.75	36.90	650	848
	White	4,327	732.97	39.01	650	850
Economic Status*	Economically Disadvantaged	7,165	711.83	31.29	650	850
	Not Economically Disadvantaged	7,679	724.77	37.67	650	850
English Learner Status	English Learner (EL)	1,477	699.18	22.68	650	793
	Non English Learner	13,337	720.57	35.80	650	850
Disabilities	Students with Disabilities (SWD)	3,114	701.87	29.61	650	840
	Students without Disabilities	11,667	722.79	35.37	650	850
Reading Score		**14,921**	**38.51**	**13.53**	**10**	**90**
Gender	Female	6,410	40.30	13.99	10	90
	Male	8,511	37.16	13.01	10	90
Ethnicity	American Indian/Alaska Native	1,046	35.76	10.42	10	81
	Asian	338	46.22	17.69	12	90
	Black or African American	2,901	35.66	12.40	10	86
	Hispanic/Latino	5,937	35.66	11.32	10	90
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	290	40.43	14.26	10	88
	White	4,327	44.20	15.11	10	90
Economic Status*	Economically Disadvantaged	7,165	35.83	11.80	10	90
	Not Economically Disadvantaged	7,679	40.97	14.52	10	90
English Learner Status	English Learner (EL)	1,477	30.78	8.30	10	77
	Non English Learner	13,337	39.32	13.72	10	90
Disabilities	Students with Disabilities (SWD)	3,114	32.30	11.40	10	90
	Students without Disabilities	11,667	40.09	13.56	10	90
Writing Score		**14,921**	**22.93**	**12.77**	**10**	**60**
Gender	Female	6,410	26.37	12.96	10	60
	Male	8,511	20.34	11.99	10	60

Group Type	Group	*N*	Mean	*SD*	Min	Max
Ethnicity	American Indian/Alaska Native	1,046	24.32	11.67	10	60
	Asian	338	30.18	14.73	10	60
	Black or African American	2,901	20.69	12.06	10	60
	Hispanic/Latino	5,937	20.43	11.64	10	60
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	290	23.89	12.92	10	60
	White	4,327	26.84	13.52	10	60
Economic Status*	Economically Disadvantaged	7,165	21.06	11.98	10	60
	Not Economically Disadvantaged	7,679	24.64	13.22	10	60
English Learner Status	English Learner (EL)	1,477	17.77	10.13	10	50
	Non English Learner	13,337	23.44	12.88	10	60
Disabilities	Students with Disabilities (SWD)	3,114	17.83	10.90	10	60
	Students without Disabilities	11,667	24.20	12.86	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.4 Fall 2017 Subgroup Performance for Mathematics Scale Scores: Algebra I

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**17,304**	**724.02**	**30.19**	**650**	**850**
Gender	Female	8,464	724.94	29.88	650	850
	Male	8,840	723.13	30.47	650	847
Ethnicity	American Indian/Alaska Native	65	713.11	26.43	669	776
	Asian	696	744.61	36.17	651	850
	Black or African American	6,917	713.14	23.30	650	830
	Hispanic/Latino	3,401	717.53	27.32	650	830
	Native Hawaiian or Pacific Islander	30	722.57	24.18	681	768
	Multiple Race Selected	504	730.37	32.90	650	850
	White	5,673	738.13	31.22	650	843
Economic Status*	Economically Disadvantaged	6,051	716.96	26.19	650	850
	Not Economically Disadvantaged	11,203	727.75	31.52	650	850
English Learner Status	English Learner (EL)	1,701	706.36	22.35	650	820
	Non English Learner	15,516	725.89	30.35	650	850
Disabilities	Students with Disabilities (SWD)	3,450	710.58	24.20	650	843
	Students without Disabilities	13,726	727.29	30.66	650	850
Language Form	Spanish	130	701.02	18.92	650	772

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.5 Fall 2017 Subgroup Performance for Mathematics Scale Scores: Geometry

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**6,306**	**725.51**	**27.10**	**650**	**850**
Gender	Female	3,211	726.70	26.01	650	850
	Male	3,095	724.27	28.14	650	850
Ethnicity	American Indian/Alaska Native	207	709.99	15.04	658	759
	Asian	347	759.17	29.69	685	850
	Black or African American	1,062	717.36	23.10	650	795
	Hispanic/Latino	2,083	714.56	21.97	650	792
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	115	731.64	25.74	658	788
	White	2,442	734.55	25.86	650	812
Economic Status*	Economically Disadvantaged	2,636	715.85	22.09	650	794
	Not Economically Disadvantaged	3,610	732.53	28.29	650	850
English Learner Status	English Learner (EL)	443	705.10	17.69	650	773
	Non English Learner	5,760	727.00	27.14	650	850
Disabilities	Students with Disabilities (SWD)	1,210	709.57	21.92	650	836
	Students without Disabilities	4,974	729.32	26.93	650	850
Language Form	Spanish	46	711.39	17.22	650	751

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.6 Fall 2017 Subgroup Performance for Mathematics Scale Scores: Algebra II

Group Type	Group	*N*	Mean	*SD*	Min	Max
Full Summative Score		**12,777**	**708.43**	**37.44**	**650**	**850**
Gender	Female	6,415	709.41	36.57	650	850
	Male	6,362	707.45	38.28	650	850
Ethnicity	American Indian/Alaska Native	985	693.81	24.17	650	804
	Asian	456	748.51	47.63	650	850
	Black or African American	2,306	695.00	28.19	650	815
	Hispanic/Latino	4,556	696.11	26.57	650	844
	Native Hawaiian or Pacific Islander	n/r	n/r	n/r	n/r	n/r
	Multiple Race Selected	235	717.09	39.11	650	834
	White	4,165	727.99	41.44	650	850
Economic Status*	Economically Disadvantaged	5,791	696.64	27.59	650	840
	Not Economically Disadvantaged	6,893	718.21	41.49	650	850
English Learner Status	English Learner (EL)	1,038	684.98	21.85	650	784
	Non English Learner	11,616	710.41	37.82	650	850
Disabilities	Students with Disabilities (SWD)	1,579	690.74	31.37	650	850
	Students without Disabilities	11,049	710.87	37.56	650	850
Language Form	Spanish	124	682.52	25.85	650	784

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.