New York State Testing Program 2019: English Language Arts and Mathematics Grades 3–8

Technical Report

Questar Assessment Inc.
2020

Table of Contents

List of Tables

List of Figures

Section 1: Introduction and Overview

1.1. Introduction

This technical report provides detailed information regarding the technical, statistical, and measurement attributes of the New York State Testing Program (NYSTP) for the Grades 3–8 English Language Arts (ELA) and Mathematics 2019 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, differential item functioning (DIF) studies, test administration, scoring, scaling, and student performance.

1.2. Test Purpose

The 2019 Grades 3–8 ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade-level New York State Learning Standards in ELA and Mathematics. The tests are designed to allow the classification of student proficiency into four performance levels (Level I, Level II, Level III, and Level IV). Likewise, the test provides opportunities for students at each of these performance levels to demonstrate their knowledge and skills in the Learning Standards. Details about the content standards for ELA and Mathematics are described in Section 2.4: Test Blueprints.

1.3. Expected Participants

Students in New York State public school Grades 3, 4, 5, 6, 7, and 8 (and ungraded students of equivalent chronological ages) are the expected participants for the Grades 3–8 NYSTP. Religious and independent schools may participate in the testing program, but their participation is not mandatory. In 2019, some religious and independent schools participated in the testing program across all grade levels. These schools were included in the data analyses. Public school students were required to take all State assessments administered at their grade level, except for a very small percentage of students with severe cognitive disabilities who took the New York State Alternate Assessment (NYSAA). For more detail on this exemption, please refer to the *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available online at http://www.p12.nysed.gov/assessment/sam/ei/eisam19.pdf.

1.4. Test Use and Decisions Based on Assessment

The NYSTP Grades 3–8 ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State Learning Standards in ELA and Mathematics, respectively, in order to determine whether schools, districts, and the State meet the required progress objectives specified in the New York State accountability system. Several types of scores are available from the Grades 3–8 ELA and Mathematics Tests, and they are discussed in this section.

1.4.1. *Scale Scores*

The scale scores are a quantification of the proficiency measured by the Grades 3–8 ELA and Mathematics Tests at each grade level. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades or across subjects. The scale scores are reported at the individual student level and can be aggregated. Detailed information on the derivation and properties of the scale scores including the range of scale scores for each subject and grade is provided in Section 6: IRT Calibration. The Grades 3–8 ELA and Mathematics

Tests' scale scores are the basis for placing students into performance levels, which are used to determine student progress within schools and districts; support registration of schools and districts; determine eligibility of students for additional educational services, and provide teachers with indicators of a student's need, or lack of need, for remediation in specific content-area knowledge.

1.4.2. *Statewide Percentile Ranks*

Students' scale scores are also presented as percentile ranks in order to indicate student performance relative to the entire testing population on a scale that may be more familiar than the operational test's scale. Such statistics are estimated based on the frequency distribution of the scale scores. More specifically, the percentile rank of a scale score indicates the percentage of students that obtained a scale score equal to or lower than the scale score in question.

1.4.3. *Performance Level Cut Scores and Classification*

Student performance is classified as Level I, Level II, Level III, or Level IV for the Grades 3–8 ELA and Mathematics Tests. The definitions of performance levels are as follows:

- **NYS Level I:** Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered insufficient for the expectations at this grade.

- **NYS Level II:** Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered partial but insufficient for the expectations at this grade.

- **NYS Level III:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered sufficient for the expectations at this grade.

- **NYS Level IV:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels I, II, III, and IV were originally established during the process of standard setting in summer 2013. In July 2018, Questar hosted a standards review meeting to revisit and update the established cut scores given a test design change and a reduced test length in 2018 from 2017. The original standard setting process is described in detail in Section 8 and Appendix P in the *2013 Technical Report* (NYSED, 2013). The *2018 Standards Review Report* is available in Appendix T of the *2018 Technical Report*.

1.4.4. *Subscores*

The Grades 3–8 ELA tests have two subscores: reading (which includes all multiple-choice items assessing both reading and language standards) and writing to sources (which includes all constructed-response items assessing reading, writing, and language standards). The Grades 3–8 Mathematics tests have three subscores that are the domain-level scores for items measuring the *Major Clusters* in each grade. The New York State Learning Standards are divided into *Major*, *Supporting*, and *Additional Clusters*. Standards within *Major Clusters* are the intended focus of instruction and assessment; these standards account for the majority of the Mathematics test items. The *Supporting* and *Additional Clusters* are Mathematics standards that both introduce and reinforce *Major Clusters*. Tables 1.1 and 1.2 present the reporting subscore categories and the point values that correspond to each on the 2019 tests. In 2019, subscores were reported in two ways:

1. A raw score (i.e., number of points earned) out of the total score on the test
2. The average score at the state level for each subscore category

Table 1.1. ELA Subscore Categories and Total Possible Score Points

	Total Subscore Points	
Grade	**Reading**	**Writing to Sources**
3	18	16
4	18	16
5	28	16
6	28	16
7	28	18
8	28	18

Table 1.2. Mathematics Subscore Categories and Total Possible Score Points

	Reporting Subscores and Total Subscore Points		
Grade	**Subscore 1**	**Subscore 2**	**Subscore 3**
3	Operations and Algebraic Thinking 19 (17–21)	Number and Operations—Fractions 8 (6–10)	Measurement and Data 11 (9–13)
4	Operations and Algebraic Thinking 9 (7–11)	Numbers and Operations in Base 10 10 (10–14)	Number and Operations—Fractions 12 (10–14)
5	Numbers and Operations in Base 10 13 (10–14)	Number and Operations—Fractions 18 (16–20)	Measurement and Data 12 (10–14)
6	Ratios and Proportional Relationships 13 (10–14)	The Number System 9 (9–13)	Expressions and Equations 23 (16–22)

Grade	Reporting Subscores and Total Subscore Points		
	Subscore 1	**Subscore 2**	**Subscore 3**
7	Ratios and Proportional Relationships 14 (12–16)	The Number System 9 (8–12)	Expressions and Equations 17 (13–19)
8	Expressions and Equations 18 (18–24)	Functions 13 (11–15)	Geometry 12 (10–14)

Note. Possible points are in parenthesis beneath the category; the number of items is outside of the parenthesis.

1.5. Testing Accommodations

In accordance with federal law under the Americans with Disabilities Act and the section Fairness in Testing and Test Use in the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 2014), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. The allowance is in accordance with a student's Individualized Education Program (IEP) or Section 504 Accommodation Plan (504 Plan). School principals are responsible for ensuring that proper accommodations are provided when necessary, and that staff providing accommodations are properly trained. Details on testing accommodations can be found in the 2019 School Administrator's Manual (SAM).

1.6. Test Transcriptions

For visually impaired students, large-type and Braille editions of the test books are provided. In most cases, the students dictate and/or record their responses, the teachers transcribe student responses to the multiple-choice items onto scannable answer sheets, and the teachers transcribe the responses to the constructed-response items onto the regular test books. Some of the students who use large-type editions will fill in the answer sheets by themselves. The large-type editions are created by Questar Assessment Inc. and printed by SeaChange Print Innovations. SeeWriteHear, LLC, produces the Braille editions. SeeWriteHear employs certified Library of Congress Braille transcribers and delivers Braille in accordance with the Braille Authority of North America (BANA) standards. Camera-ready versions of the regular test books are provided to the Braille vendor, which then produces the Braille editions. Proofs of the Braille editions are submitted to NYSED for review and approval prior to production.

1.7. Test Translations

The NYSTP Grades 3–8 Mathematics Tests are translated into five languages: Chinese (Traditional), Haitian-Creole, Korean, Russian, and Spanish. These tests are translated to provide students the opportunity to demonstrate mathematical proficiency independent of their command of the English language. Sample tests are available in each translated language at the following location: http://www.p12.nysed.gov/assessment/math/samplers/.

English Language Learner/Multilingual Learner (ELL/MLL) students taking the Grades 3–8 Mathematics Tests may be provided with an oral translation of the test when a written translation is not available in the student's native language. The following testing accommodations are also made available to ELLs: separate testing location, bilingual glossaries, simultaneous use of

English and alternative-language editions, oral translation for lower-incidence languages, and writing responses in the native language.

The NYSTP Grades 3–8 ELA Tests are not translated into any other language because they are assessments of proficiency in English language arts. The following testing accommodations are made available to ELL/MLLs taking the ELA Tests: separate testing location and bilingual glossaries.

Section 2: Test Design and Development

2.1. Test Descriptions

The 2019 Grades 3–8 ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) test items based on the New York State P–12 Learning Standards. The tests were administered in New York State classrooms during a three-day period for paper-based tests, and a six-day period for computer-based tests from April to May of 2019. Details on the administration and scoring of these tests can be found in Section 4: Test Administration and Scoring. Additional information can be found in the *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available at http://www.p12.nysed.gov/assessment/sam/ei/eisam19.pdf.

2.1.1. *ELA Tests*

The 2019 Grade 3–8 ELA Tests were designed to measure student literacy as defined by the New York State Learning Standards. The tests assessed Reading, Writing, and Language standards by using multiple-choice, short-response, and extended-response items. All items were based on close readings of informational, literary, or paired texts. All texts were drawn from authentic, grade-level works.

Multiple-choice items were designed to assess Reading and Language Standards. Multiple-choice items required students to analyze different aspects of a given text, including central idea, style elements, character and plot development, and vocabulary.

Short-response items were designed to assess Reading and Language Standards. These were single items in which students used textual evidence to support their answers to inferential questions. These items asked students to make an inference, state a position, or draw a conclusion based on their analysis of the passage and then provide two pieces of text-based evidence to support their answers. In responding to these items, students were expected to write in complete sentences. Appendix H provides the rubric for the short-response items.

Extended-response items were designed to assess Reading, Writing, and Language Standards, with a focus primarily on the Writing Standard. Extended-response items required comprehension and analysis of either an individual text (Grades 3–8) or paired texts (Grades 4–8). Paired texts required students to read and analyze two related texts. Paired texts were related by theme, genre, tone, time period, or other characteristics. Many extended-response items asked students to express a position and support it with text-based evidence. For paired texts, students were expected to synthesize ideas between and draw evidence from both texts. Extended-response items required students to demonstrate their ability to write a coherent essay, using textual evidence to support their ideas. Appendix I provides the rubric for the extended-response items.

2.1.2. *Mathematics Tests*

The 2019 Grade 3–8 Mathematics Tests were designed to measure student mathematic understanding as defined by the New York State Learning Standards. The tests required that students understand mathematics conceptually, use prerequisite skills with grade-level mathematical facts, decide which formulas and tools (e.g., protractors and rulers) to use, and solve mathematics problems rooted in the real world. The tests contained multiple-choice, short-

response (2-point), and extended-response (3-point) items. For multiple-choice items, students selected the correct response from four answer choices. For short- and extended-response items, students wrote an answer to an open-ended question. Some items required students to show their work or to explain, in words, how they arrived at their answers.

Mathematics multiple-choice items were used mainly to assess standard algorithms and conceptual standards. Multiple-choice items incorporated the New York State Learning Standards, some in real-world applications. Many multiple-choice items required students to complete multiple steps. Likewise, many of these items were linked to more than one standard, drawing on the simultaneous application of multiple skills and concepts.

Short-response items were used mainly to assess conceptual and application standards. The items required students to complete a task and show their work. Like multiple-choice items, short-response items often required multiple steps and the application of multiple mathematics skills, some in real-world applications. Appendix J provides the rubric for the Mathematics short-response items.

Extended-response items were used mainly to assess students' abilities to show their understanding of mathematical procedures, conceptual understanding, and application of those procedures and concepts. Extended-response items required students to complete two or more tasks, or a more extensive problem, and show their work. Some items also assessed student reasoning and the ability to critique the arguments of others. Appendix K provides the rubric for the Mathematics extended-response items.

2.2. Test Configuration

2.2.1. *Test Design*

The 2019 Grades 3–8 ELA Tests were composed of two sessions per grade and administered over two days. Each day consisted of one session. Session 1 contained literary and informational reading passages and MC items based on the passages. Session 2 contained reading passages with short-response items and an extended-response item based on those passages.

The 2019 Grades 3–8 Mathematics Tests were composed of two sessions per grade and administered over two days. Each day consisted of one session: Session 1 contained MC items. Session 2 contained MC items as well as short- and extended-response items.

The tables in Appendix A provide information on the numbers and types of items in each session for the Grades 3–8 ELA and Mathematics Tests, and the testing times.

2.2.2. *Embedded Field-Test Items*

In 2010, NYSED announced its commitment to embed multiple-choice items for field testing within the Spring 2012 Grades 3–8 ELA and Mathematics Operational Tests. This commitment continued for the Spring 2019 administrations of the tests. Embedding field-test items allows for a better representation of student responses and provides more reliable field-test data on which to build future operational tests. In other words, since the specific locations of the embedded field-test items were not disclosed and they look the same as operational test items, students were unable to differentiate field-test items from operational test items. Therefore, field-test data

derived from embedded items are free of the effects of differential student motivation that may characterize stand-alone field-test designs. Embedding field-test items also reduced the number of stand-alone field-test forms during Spring 2019, although it did not eliminate the need for them.

2.3. New York State Educators' Involvement in Test Development

New York State educators are actively involved in ELA and Mathematics test development. New York State educators provide critical input throughout all stages of the test development process, which include rangefinding, educator item review, operational forms construction, passage selection, item writing, and a *Final Eyes* meeting (a final review of the test materials prior to printing).

NYSED gathers a diverse group of educators to review all test materials, in order to create fair and valid tests. The participants are selected for each testing activity based on:

- Certification and appropriate grade-level experience
- Special population experience
- Geographical region
- Gender
- Ethnicity
- Type of school (urban, suburban, or rural)

The selected participants must be certified and have both teaching and testing experience. Most of the participants are classroom teachers. Specialists such as reading coaches, literacy coaches, and special education and bilingual instructors also participate. Some participants are also recommended by principals, professional organizations, Big Four Cities (i.e., Buffalo, Rochester, Syracuse, and Yonkers), and/or the Staff and Curriculum Development Network (SCDN). A file of participants is maintained and routinely updated with current participant information, as well as the addition of possible future participants as recruitment forms are received. The process of continually updating and adding to this file contributes to NYSED's ability to include many educators in the test development process. Every effort is made to have diverse groups of educators participate in each testing event.

Additionally, Content Advisory Panels (CAPs) meet quarterly to review, vet, and provide comments on curricular and assessment work. CAPs are content-area-specific advisory panels composed of between 15 and 20 New York State P–12 educators whose members are nominated by state professional organizations, institutes of higher education, and educator unions.

2.4. Test Blueprints

After careful consideration of test length and administration constraints (e.g., location of multiple-choice and constructed-response items within test sessions), the representation and distribution of content were determined.

The New York State Learning Standards for ELA are organized into four strands: Reading, Writing, Language, and Speaking/Listening. Due to administration constraints, Speaking/Listening was determined to be best assessed only in the classroom; therefore, the ELA

Tests assess three of the four strands: Reading, Writing, and Language. Content experts reviewed the Reading, Writing, and Language standards and recommended content coverage by standard and item type, based on the depth and breadth of each standard.

The New York State Learning Standards for Mathematics are divided into standards, clusters, and domains. Standards define what students should understand and be able to do and are further articulated into lettered components. Clusters are groups of related standards. Domains are larger groups of related clusters and standards. Content experts reviewed the Mathematics standards and recommended content coverage by standard and item type (MC or CR), based on the emphasis of the cluster (major, supporting, and additional), and the depth and breadth of each standard.

Tables B1 and B2 in Appendix B show the test blueprint and actual number of score points in the Grades 3–8 ELA and Mathematics Tests, respectively. The tables include the ranges of allowable points for each ELA strand, Mathematics domain, and the actual number of points on the 2019 operational tests.

2.5. Passage Selection and Item Criteria Documents

To guide test item development and to help ensure that New York State tests were measuring the Learning Standards for ELA and Mathematics with fidelity, criteria were established for selecting passages and writing test items, based on the consultation with the groups listed above.

Passage Review Criteria documents were created based on the passage selection guidelines and were used to evaluate each potential passage and determine whether it could be used to measure the New York State Learning Standards for ELA. The criteria documents were used to determine whether each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas. The full passage review criteria can be found here: https://www.engageny.org/resource/new-york-state-passage-selection-resources-for-grade-3-8-assessments.

Item Review Criteria for the Grade 3–8 ELA Tests were used to help ensure that each item was clear and fair, measured a specific standard or standards with fidelity, and conformed to the specifications for each item type. Each section of the criteria includes pertinent questions used to determine whether an item was of sufficient quality so that it could move forward in the development process. The first two of the *Item Review Criteria*, clarity and fairness, identify the basic components of quality items. The criteria for clarity are used to help ensure that students understand what is asked in each item and that the language choice in the item does not negatively affect a student's ability to perform the required task. For example, the criteria include checking to make sure that the vocabulary of test items is at grade level and that items avoid technical terms unrelated to the content. Likewise, the fairness criteria are used to ensure that items are unbiased, non-offensive, and not disadvantageous to any given subgroup. The criteria also address how each item measures a given standard or standards and articulates the aspects of each standard that the items need to address. Finally, the criteria establish key requirements for each item type (e.g., requiring that each two-point constructed-response item asks students to make a clear statement that can be supported with two independent text-based pieces of

evidence). The complete ELA criteria documents can be found here: http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests.

Item Review Criteria for the Grade 3–8 Mathematics Tests were used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to the New York State Learning Standards, and conformity to the expectations for specific item types and formats for each test item. Each section of the criteria includes pertinent questions that determine whether an item is of sufficient quality. The first two criteria, clarity and graphical appropriateness and fairness, identify the basic components of quality test items. The criteria for clarity and graphical appropriateness are used to help ensure that students understand what is asked in each item and that the language in the item does not adversely affect a student's ability to perform the required task. For example, the criteria include checking to make sure that the visual load for any item containing art is reasonable and that interpreting a graphic does not confuse the underlying construct. Likewise, the fairness criteria are used to evaluate whether or not items are unbiased, non-offensive, and not disadvantageous to any given subgroup. The criteria also require documentation of how each item measures the assigned Mathematics standard(s). Finally, the criteria address the specific demands for different item types and formats (making sure that each three-point constructed-response item involves a multi-step process and requires students to show work). The complete Mathematics criteria document can be found here: https://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests.

The *Multiple Representations for NYS Grade 3–8 Mathematics Tests* document was developed to ensure that the tests measured the deep conceptual understanding that the New York State Learning Standards demand, rather than focusing on predictable Mathematics items that require only algorithmic strategies to be solved correctly. *Multiple Representations* is a broad set of specifications that describes, refers to, and symbolizes the various, but not all, ways that Mathematics standards could be measured within the constraints of the NYSTP. The document specifies three overarching families: procedural skills, conceptual understanding, and application. It also includes information about how to identify standards that might be measured through the use of a particular representation. It identifies types of Mathematics skills (e.g., application of process and explanation of a principle) that are appropriate for assessing different representations. The full document can be found here: https://www.engageny.org/resource/multiple-representations-for-nys-grade-3-8-common-core-mathematics-tests.

2.5.1. *Principles of Universal Design*

To create tests as equitable as possible for students, principles of Universal Design were employed during the creation of the tests and test items. In a report published by the National Council on Educational Outcomes, "'Universally designed assessments' are designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment" (Thompson, S.J., Johnstone, C.J., & Thurlow, M.L. 2002). The report goes on to describe seven elements of a universally designed assessment. These elements are:

1. Inclusive assessment population
2. Precisely defined constructs

3. Accessible, unbiased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

In accordance with these elements, the Universal Design Item Checklist in Appendix D was developed for use during item development.

2.6. Passage Finding

The goal of passage finding is to obtain high-quality texts from which to generate Learning Standards-aligned test items. To do so, in the 2016–2017 development cycle, independent passage finders were recruited and trained, using passage selection resources such as the passage selection criteria. Passage finders were given assignments based on the test blueprint requirements. Passage finders submitted passages along with completed criteria documents and source information to ELA content specialists, who reviewed the passages against the agreed-upon criteria. Passages that did not meet the criteria were rejected, and passages that did meet the criteria were moved forward in the process, where the text from scanned copies of the original sources was entered into templates. Once in the templates, readability metrics were determined for each text. Passages were then proofread by copyeditors, fact checked by research librarians, reviewed for content issues by Science and Social Studies content specialists when necessary, and reviewed for Universal Design issues by specifically trained reviewers. After the passages went through these review steps, ELA content specialists posted the passages and completed criteria documents for NYSED's review and approval for moving forward in the process.

NYSED staff retrieved and reviewed the passages and criteria documents. If NYSED staff determined that a passage did not meet the criteria, the passage was rejected and the NYSED staff provided an explanation for rejection.

In addition to the content reviews performed by NYSED staff and its vendors, executives in both organizations also reviewed the passages. The executive review focused on bias and sensitivity issues particular to New York State. Passages that passed both content and executive reviews were moved forward for item development.

2.7. Item Development

Item development for the 2019 test forms was conducted during the 2017–2018 development cycle. The goal of item development is to develop a sufficient number of high-quality, Learning Standards-aligned items to populate the test forms. Using the criteria documents for both content areas and the multiple-perspective document for Mathematics, content leads trained item writers. The item writers had teaching or assessment experience in the content area for which they were writing items; experience in writing for large-scale, high-stakes assessments; and, at minimum, a bachelor's degree in either education and/or the content area for which they were assigned. The item writers were given specific assignments, based on the test blueprint. For ELA, the item writers were also provided with the completed passage criteria documents.

Item writers provided items to content specialists for review. At least two content specialists reviewed each item. After the content specialists were satisfied that all of the items met the

criteria, the items were reviewed by copyeditors. The Mathematics items were also reviewed by content specialists in Science, Social Studies, and by research librarians when necessary. The ELA and Mathematics content specialists evaluated the feedback from the different internal groups and edited the items accordingly. The items were then posted for NYSED's review and approval for moving forward in the process.

NYSED content experts retrieved and reviewed the items. If NYSED staff determined that an item did not meet the criteria, the NYSED staff provided an explanation for rejection or revision. If NYSED staff determined that an item met the criteria but could be improved with editing, the staff member recorded notes for the edits. Those notes were reviewed at face-to-face meetings at which content staff and NYSED staff reviewed and edited all of the items to ensure that they met the criteria. All passages and items accepted at that meeting were moved forward for the educator item review.

2.8. Educator Item Review

After being reviewed by NYSED, the items were presented to panels of New York State educators. Based on their expertise, educators were assigned to grade-level and content-specific groups where they reviewed the items. The reviews were facilitated by Questar content specialists and were attended by NYSED staff. For ELA, reviewers first read and then discussed the passages before reviewing items. For Mathematics and ELA, the educators used the following checklist to review each item.

ELA Checklist

Is the passage reading level appropriate for the grade?
Is the passage appropriate and fair for the grade?
Are the passage graphics accurate and appropriate?
Is the question aligned to the intended standard?
Is there one and only one key?
Are the distractors plausible?
Is the question clearly worded?
Is the item free of bias and sensitivity concerns?

Math Checklist

Is the question aligned to the intended standard?
Are the mathematics correct?
Is there one and only one key?
Are the distractors plausible?
Is the question clearly worded?
Is the item free of bias and sensitivity concerns?
Are the item graphics accurate and appropriate?

As the educators reviewed the items, they discussed their judgments about them. If the educators felt that an item did not align to the standards, did not meet quality standards, or was not fair, they made recommendations for editing the item. NYSED staff and Questar content specialists later reviewed the recommendations and made the appropriate edits.

2.9. Field-Testing

Once the items have been developed and thoroughly reviewed by a variety of stakeholders, they must then be field-tested. Field-testing items is a critically important step in the test development process, as it is only through the gathering of actual student response data that a variety of psychometric characteristics may be evaluated. Table 2.1 provides a summary of the unique items that passed the scrutiny of NYSED and Questar content specialists, as well as that of New York State educators, and were field-tested. More items were field-tested than were needed on the operational forms because that enabled tests to be constructed with items that include the best possible characteristics from both a content and psychometric perspective.

Table 2.1. Summary of Unique 2019 Field Test Items

	Unique ELA Items by Type		Unique Mathematics Items by Type	
Grade	**MC**	**CR**	**MC**	**CR**
3	130	54	84	30
4	131	45	98	30
5	142	36	98	30
6	128	43	110	30
7	130	43	98	30
8	134	43	98	30

Note. MC = multiple-choice. CR = constructed-response. All CR items were field-tested under stand-alone conditions, while nearly all MC items were administered under the embedded condition only. Twelve MC items were field tested for Math Grade 6 in the stand-alone condition with CR items.

Multiple-choice field-test items were administered as embedded field-test items within the 2019 operational test forms. A majority of MC items on the forms were embedded field-test items; stand-alone field-test items were mostly CR items. The use of embedded field-test items yields more reliable field-test data and has nearly eliminated the need for multiple-choice stand-alone field-testing. One additional round of field-testing was administered separately from the 2019 operational forms (i.e., as stand-alone tests) later in Spring 2019, which included CR items and a minimal number of MC items.

A variety of analyses were conducted in order to better understand how the items field-tested in 2019 may perform on future operational forms. All of the field-test data underwent a series of representativeness checks. Because only a small sample of schools participate for any given content area and grade for stand-alone field-testing, it was necessary to ensure that the stand-alone field-test samples were representative of the entire State population in terms of student achievement on prior years' tests, student gender, student ethnicity, and school Needs/Resource Capacity Category (NRC). Finally, a variety of psychometric analyses were conducted, including classical item analysis, inter-rater reliability for constructed-response items, differential item functioning (DIF), item response theory (IRT), item calibration, scaling, and fit evaluation. Many of these analyses are described at length in the Field Test Technical Report. However, inter-rater reliability analyses were not possible for the operational test, as only a single rater scored each constructed-response.

2.10. Rangefinding

Questar conducted rangefinding for items included on the 2019 test. Rangefinding occurs after constructed-response items have been field-tested. The purpose of rangefinding is to have New York State educators review student constructed-responses and arrive at consensus scores based on the standards established by NYSED and the scoring rubrics. The consensus scores become the basis for operational rating guides and scoring ancillaries. To arrive at consensus, committees of New York State educators review, discuss, and rate student responses to the constructed-response field-test items. NYSED content experts and Questar Scoring Directors oversaw this process. The first step in the rangefinding process was to have the educator committees review rubrics and a NYSED-approved grounding guide set, previously used for the 2018 field-test rangefinding sessions, to familiarize teachers with the application of NYSED standards and rubrics. The grounding guide sets contain student responses that illustrate the full range of scores on the rubric. The grounding guide sets are composed of student responses that had previously gone through the rangefinding process and been approved by NYSED and are used to guide the scoring of field-test and operational student responses. Referencing the previously approved guide set papers during the rangefinding sessions ensures consistency in the application of NYSED standards and rubrics from year-to-year.

After the committee reviewed the pre-approved grounding guide set, groups of committee members familiarized themselves with each item type, scoring a small number of responses representative of each of the different score points. After the group-scoring exercise, committee members independently scored other student responses. The committee then reviewed and discussed their results and determined consensus scores for the responses. The rangefinding results were used to build training materials for Questar scorers, who scored the field-test responses to constructed-response items.

2.11. Item Selection and Test Creation (Criteria and Process)

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered from April to May of 2019. The test items were selected from the pools of available ELA and Mathematics items. These items were field-tested either in embedded field-testing or stand-alone field-testing from 2016 through 2018.

The test construction process involved several iterative steps. Three criteria governed the item selection process:

- Meet the ELA and Mathematics content specifications provided by NYSED
- Select items with the best psychometric characteristics from the ELA and Mathematics item pools
- Combine psychometric characteristics of all srcfr/5
- -
- elected items with the intended psychometric goals for each entire form

Questar content specialists were provided the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines were based on the classical and IRT statistics associated with the test items.

Using the pool of field-tested items, Questar content specialists made preliminary selections for each grade and content area. The selections were then reviewed by the content leads for each content area, to make sure that the items conformed to the different criteria. If the content criteria were not met, new items were selected. After the content leads' review, the item selections were reviewed by Questar psychometricians. If items with undesirable statistics were selected, the psychometricians proposed items with more desirable statistics. The content specialists and their leads then reviewed those items. Once the Questar content teams and the psychometric teams were satisfied that the content and statistics of the selected items and the proposed whole forms met the requirements, the items were given to NYSED staff (including content and assessment experts) to review. Questar content specialists and psychometricians traveled to Albany, New York, in November 2018 to finalize item selection and test creation with NYSED staff (including content and assessment experts) and New York State educators.

2.12. Educator Form Construction

During an educator form construction meeting that took place from November 5–9, 2018 in Albany, New York, educators from around the State worked with NYSED and Questar to review the content of the proposed 2019 operational ELA passages, and ELA and Mathematics individual test items. They looked at how those items combine to create entire operational forms, and for quality and appropriateness using their subject matter expertise. The goal was to ensure that all test items and forms are defensible from content and psychometric perspectives. The outcome was test forms that meet psychometric parameters and contain items that meet content criteria.

On November 5, educators reviewed Grades 4, 6, and 8. Different educators reviewed Grades 3, 5, and 7 on the following day. Each grade and subject group had 5-6 educators. Since different groups of educators participated in the review of each subject and grade's test form, each morning began with training in each room. Once training was complete, participants began the form construction process by independently evaluating the items and passages (for ELA) against the criteria on the provided checklists. Each participant completed his or her own checklist and had access to Questar's Content Management System, which displayed the items corresponding to the order of items in the test.

- For ELA, the educators initially reviewed the first passage and a single item from the passage. Once they got used to the process, the educators reviewed the passages and the corresponding items. During this review, educators confirmed that there was only one correct answer for each multiple-choice item, and that the item was aligned to the standard that it purported to address.
- For Mathematics, the educators initially reviewed single items and discussed each item as a group. Once they got used to the process, the educators reviewed groups of items (e.g., 4 to 6 items, followed by discussion of each item). During this review, educators confirmed that there was only one correct answer for each multiple-choice item, and that the item was aligned to the standard that it purported to address.

In both ELA and Mathematics, the educators, in consultation with NYSED and Questar content experts, were permitted to recommend:

- revisions to the stated standard alignment;
- revisions to item sequencing to avoid cueing/clueing; and
- swapping any items and/or passages that they judged as having problems flagged by the above reviews.

Given other constraints, it was not always possible to make every change that educators recommended, but they were given the opportunity to voice any and all concerns that they had; NYSED made the final decision about any educator recommendations.

The facilitators then led a group discussion and helped the group reach consensus. Where time permitted, educators were presented with and approved the items that Questar and NYSED proposed for any necessary replacements. Following each session with educators, NYSED and Questar met to review the content and data of the proposed selections and explore alternate selections for consideration. NYSED then approved the item selections, including item positions within test sessions.

2.13. Test Form Production

Once the selection of items for the operational and embedded field-test positions was completed, Questar created test forms. The test forms were reviewed by Questar content specialists and were posted for NYSED to review. NYSED and Questar reviewed the forms to look for any errors in spelling, capitalization, punctuation, grammar, and formatting. They also confirmed that each multiple-choice item had a single correct answer.

2.14. Final Eyes Committees

After NYSED and Questar reviewed copies of the test forms, the test forms were reviewed by the Final Eyes committees. For each content area, the committee consisted of thirty New York State educators from around the State. During that review, the educators were charged with taking the test to make sure that each multiple-choice item had a single correct answer, and to look for errors in spelling, capitalization, punctuation, grammar, and formatting.

After the Final Eyes review and after NYSED approved edits made as a result of the review, the tests were then considered final and produced for the 2019 administration.

2.15. Proficiency and Performance Standards

In July 2018, a standards review meeting occurred in Albany where 56 New York State educators went through a rigorous process, guided by the best practices indicated by this intensely studied process, to recommend updated performance standards. These recommendations were presented to the Commissioner, who, in turn, adopted the recommended standards set forth by the committees. For additional details on the standards review process, see Appendix T of the *2018 Technical Report*.

Each grade level has four performance levels. Three cut points demarcate the performance levels needed to demonstrate each ascending level of performance. Section 6.7.1 contains the raw score-to-scale score, SEM conversion tables, and detailed information related to the performance standards.

Section 3: Validity

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests. Test validation is an ongoing process of gathering evidence from many sources to evaluate the soundness of the desired score interpretation or use. This evidence is acquired from studies of the content of the test and studies involving scores produced by the test. Additionally, reliability has to be considered before considerations of validity are made. A test cannot be valid if the test scores are not first reliable.

The *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 2014) addressed the concept of validity in testing, which refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Validity is the most important consideration in test evaluation. Test validation is the process of accumulating evidence to support any particular inference. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity refers to the degree to which evidence supports the inferences made from test scores.

3.1. Content Validity

Generally, achievement tests are used for student-level outcomes, either for making predictions about students or for describing students' performances (Mehrens and Lehmann, 1991). Tests are now also used for the purposes of accountability and adequate yearly progress (AYP). The NYSED uses various assessment data in reporting AYP. Specific to student-level outcomes, the NYSTP documents student performance in the area of Mathematics as defined by the New York State Mathematics Learning Standards and in the area of ELA as defined by the New York State ELA Learning Standards.

To allow test score interpretations appropriate for this purpose, the content of the test must be carefully matched to the specified standards. The 2014 AERA/APA/NCME standards state that content-related evidence of validity is a central concern during test development. Expert professional judgment should play an integral part in developing the definition of what is to be measured, such as describing the universe of the content, generating or selecting the content sample, and specifying the item format and scoring system.

Expert analysis of test content indicates the degree to which the content of a test covers the domain of content that the test is intended to measure. In the case of the NYSTP, the content is defined by detailed blueprints that describe New York State content standards and define the skills that must be measured to assess these content standards (see Tables B1 and B2 in Appendix B). The NYSTP test development process requires specific attention to content representation and the balance within each test form. New York State educators were involved in test construction in various development stages. For example, during the item review process, they reviewed field-test items for the alignment of the items with the Learning Standards. Educators also participated in a process of establishing scoring rubrics for constructed-response items during rangefinding. Section 2: Test Design and Development contains more information specific to the item review process.

3.2. Construct (Internal Structure) Validity

Construct validity (i.e., what scores mean and what kind of inferences they support) is often considered the most important type of test validity. Construct validity of the NYSTP Grades 3–8 ELA and Mathematics tests is supported by several types of evidence that can be obtained from the ELA and Mathematics test data.

3.2.1. *Internal Consistency*

Empirical studies of the internal structure of the test provide one type of evidence of construct validity. For example, high internal consistency constitutes evidence of validity. This is because high coefficients imply that the test items are measuring the same domain of skill and are reliable and consistent. Reliability coefficients of the tests for total populations and subgroups of students are presented in Section 7.1: Test Reliability. For the total population, the ELA reliability coefficients (Cronbach's alpha) ranged from 0.87 to 0.90. For all subgroups, the reliability coefficients were greater than or equal to 0.77. For the total population, the Mathematics reliability coefficients (Cronbach's alpha) ranged from 0.92 to 0.94. For all subgroups, the reliability coefficients were greater than or equal to 0.83. Overall, high internal consistency of the NYSTP Grades 3–8 ELA and Mathematics Tests provided sound evidence of construct validity.

3.2.2. *Unidimensionality*

Other validity evidence comes from analyses of the degree to which the test items conform to the requirements of the statistical models. These statistical models are used to scale and link the tests, as well as to generate student scores. The models require that the items fit the model well (item fit) and that the items in a test measure a single domain of skill (unidimensionality).

The first step is to assess the degree to which the items fit the IRT model. The item-model fit for the ELA and Mathematics tests was assessed using model-data fit plots, and the results are described in detail in Section 6: IRT Calibration. Most items demonstrated sound fit across grades and content areas, and only a few items were deemed to have less than ideal fit. This provides solid evidence for the appropriateness of the IRT models used to calibrate and scale the test data.

Additional evidence for the efficacy of the model involves demonstrating that the items on the New York State tests are related to each other, within their respective content areas. This relationship of the items within the ELA or Mathematics tests is the common proficiency acquired by students studying the content area. This "common proficiency," or, more formally, underlying construct, could be labeled as ELA proficiency (using the ELA scores) or Mathematics proficiency (using the mathematics scores), depending on the degree to which the ELA and Mathematics items are related.

Factor analysis of the test data is one way of modeling the common construct. This analysis may show that there is a single or main factor that can account for much of the variability between responses to test items. A large first component in factor analysis would provide evidence of the latent proficiency that students have in common regarding the particular items asked. A large main factor found from a factor analysis of an achievement test would suggest a primary

construct that may be related to what the items were designed to have in common (i.e., Mathematics proficiency or ELA proficiency).

To demonstrate the common factor underlying student responses to the ELA and Mathematics test items, principal component factor analyses were conducted on a correlation matrix of individual items for the ELA and Mathematics tests. Factoring a correlation (i.e., tetrachoric correlation) matrix rather than actual item response data is preferable when dichotomous variables are in the analyzed data set. Because the ELA and Mathematics tests contain both multiple-choice and constructed-response items, the matrices of *polychoric* correlations were used as input for the factor analyses, as polychoric correlations are appropriate with both multiple-choice and constructed-response data. The study was conducted on the New York State public, charter, and religious and independent school students for whom data were available. A large first principal component was evident in each analysis, demonstrating essential unidimensionality of the trait (i.e., proficiency) measured by each test. In other words, statistical evidence indicates that the ELA items are measuring one underlying construct, ELA proficiency, and that the Mathematic items are measuring one underlying construct, Mathematics proficiency.

The factor analyses conducted with the ELA and Mathematics data will show almost as many underlying constructs, or factors, as there are items on the test. Therefore, it is necessary to investigate the factor analysis results further to determine the number of "meaningful" factors. Specifically, more than one factor with an eigenvalue greater than 1.0 present in each dataset would suggest the presence of small additional factors (Kaiser, 1960). The magnitude of the ratio of the variance accounted for by the first factor compared to the remaining factors also provides evidence as to the number of meaningful factors (Cattell, 1966). In addition, the total amount of variance accounted for by the main factor was evaluated. According to M. Reckase (1979),

> . . . the 1PL and the 3PL models estimate different abilities when a test measures independent factors, but . . . both estimate the first principal component when it is large relative to the other factors. In this latter case, good ability estimates can be obtained from the models, even when the first factor accounts for less than 10 percent of the test variance, although item calibration results will be unstable. (p. 228)

Factor analyses related to the Grades 3–8 ELA and Mathematics Tests indicated that the ratio of the variance accounted for by the first factor to the remaining factors was sufficiently large to support the claim that the ELA and Mathematics tests were essentially unidimensional. The ELA-related ratios and the Mathematics-related ratios showed that the first eigenvalues were at least four times and six times as large as the second eigenvalues for all of the grades.

All of the Grades 3–8 ELA and Mathematics Tests exhibited first principal component accounting for more than 22% and 27% of the test variance, respectively. Tables 3.1 and 3.2 present the results of factor analyses, including eigenvalues greater than 1.0 and proportions of variance explained by the extracted factors, for ELA and Mathematics, respectively.

The evidence in Table 3.1 supports the claim that one single construct underlies the items/tasks in each ELA test and that scores from each test would represent performance primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors. Similarly, Table 3.2 supports the claim that a common construct underlies the

items/tasks in each Mathematics test and that scores from each test would represent performance primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors.

Table 3.1. ELA Tests Factor Analysis

Grade	Extracted Factor			
		Initial	Variance Accounted for	
	#	Eigenvalue	%	Cumulative %
3	1	**6.74**	**26.97**	**26.97**
	2	1.42	5.67	32.64
	3	1.18	4.72	37.36
4	1	**6.19**	**24.74**	**24.74**
	2	1.47	5.88	30.63
	3	1.18	4.70	35.33
5	1	**8.39**	**23.97**	**23.97**
	2	1.42	4.05	28.02
	3	1.10	3.13	31.15
6	1	**8.00**	**22.87**	**22.87**
	2	1.52	4.34	27.20
	3	1.06	3.02	30.22
	4	1.01	2.89	33.11
7	1	**8.82**	**24.49**	**24.49**
	2	1.61	4.49	28.97
	3	1.39	3.85	32.83
8	1	**8.71**	**24.19**	**24.19**
	2	1.83	5.07	29.26
	3	1.25	3.47	32.73
	4	1.01	2.80	35.53

Table 3.2. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
		Initial	Variance Accounted for	
	#	Eigenvalue	%	Cumulative %
3	1	**9.58**	**28.19**	**28.19**
	2	1.57	4.62	32.81
	3	1.01	2.96	35.77
4	1	**11.06**	**29.10**	**29.10**
	2	1.41	3.71	32.81
5	1	**11.25**	**29.60**	**29.60**
	2	1.53	4.03	33.64

Grade	Extracted Factor			
		Initial	Variance Accounted for	
	#	Eigenvalue	%	Cumulative %
5	3	1.09	2.86	36.50
	4	1.02	2.69	39.19
6	1	**12.21**	**31.31**	**31.31**
	2	1.20	3.08	34.39
	3	1.05	2.70	37.09
7	1	**12.69**	**30.94**	**30.94**
	2	1.37	3.34	34.28
	3	1.15	2.80	37.08
8	1	**11.17**	**27.25**	**27.25**
	2	1.26	3.08	30.33
	3	1.20	2.93	33.26
	4	1.02	2.49	35.74

As additional evidence for construct validity, the same factor analysis procedure was employed to assess the dimensionality of the Mathematics construct for selected subgroups of students in each grade: ELLs/MLLs, students with disabilities (SWD), and students using test accommodations (SUA), as well as ELL/MLL/SUA, and SWD/SUA. The ELL/MLL/SUA subgroup is defined as examinees who are ELLs/MLLs and who use at least one ELL/MLL-related accommodation. The SWD/SUA subgroup includes examinees who are classified as having disabilities and who use at least one disability-related accommodation. The results were comparable to the results obtained from the total population data. Evaluation of eigenvalue magnitude and proportions of variance explained by the main and secondary factors provide evidence of essential unidimensionality of the construct measured by the tests for the analyzed subgroups. Appendix L provides factor analysis results for ELL/MLL, SWD, SUA, ELL/MLL/SUA, and SWD/SUA classifications.

3.2.3. *Detection of Bias*

Minimizing item bias has the goal of minimizing construct-irrelevant variance and helps establish a strong validity argument for the tests. Specifically, bias occurs if items function differentially for key pairs of groups, which may, in turn, cause the test to be differentially valid for certain groups of test takers. The statistical means for flagging items that may exhibit bias is referred to as differential item functioning (DIF). These statistical procedures were designed to be conservative (i.e., they were designed to flag more items for DIF, rather than fewer). Therefore, it is rare in practice to observe a high-stakes test in which not a single item is flagged for DIF. Since these procedures tend to over-flag items, it is only through review of those flagged items by experts that the items flagged for DIF may be judged to have or be free of bias. If the test involves irrelevant skills or knowledge, the possibility of bias is increased. Thus, preserving content validity is essential.

The developers of the NYSTP tests gave careful attention to items of possible ethnic, gender, socioeconomic status (SES), and—only for the Mathematics tests—translation bias. All materials

were written and reviewed to conform to Questar's editorial policies and guidelines for equitable assessment, as well as NYSED's guidelines for item development. All materials were written to NYSED's specifications and carefully checked by groups of trained New York State educators during the item review process. These steps are essential in keeping bias to a minimum. However, current evidence suggests that expertise in this area is no substitute for data; reviewers are sometimes wrong about which items work to the disadvantage of a group, apparently because some of their ideas about how students will react to items may be faulty (Sandoval & Mille, 1979; Jensen, 1980). Thus, empirical studies were conducted.

Statistical methods were used to identify items exhibiting possible DIF. Although items flagged for DIF in the field-test stage were closely examined for content bias and avoided during the operational test construction, DIF analyses were conducted again on operational test data. Different methods were employed to evaluate the amount of DIF in all test items: constructed-response items were evaluated with standardized mean differences, and multiple-choice items were analyzed using Mantel-Haenszel methods (see Section 5: Operational Test Data Collection and Classical Analysis).

In each grade, for both ELA and Mathematics, few items were flagged for DIF. Moreover, the magnitude of DIF for the flagged items was typically small (for more details, see Appendix N). Multiple reviewers carefully reviewed items flagged for statistically significant DIF during the operational test item selection. All such items were deemed by the reviewers to be free of bias (i.e., judged not to adversely affect any demographic subgroup studied) and remained in the tests.

Section 4: Test Administration and Scoring

This section provides summaries of New York State test administration and scoring procedures. For further information, refer to the aforementioned *School Administrator's Manual* and the *New York State Scoring Leader Handbook (2019)* located here: http://www.p12.nysed.gov/assessment/sam/ei/scoring-leader-hb19p.pdf.

4.1. Test Administration

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered to students in a paper-based (PBT) and computer-based (CBT) testing mode in 2019. The PBT testing window was Tuesday, April 2–Thursday, April 4 for the Grades 3–8 ELA Tests and Wednesday, May 1–Friday, May 3 for the Grades 3–8 Mathematics Tests. The CBT testing window was Monday, April 1–Monday, April 8 for the Grades 3–8 ELA Tests and Tuesday, April 30–Tuesday, May 7 for the Grades 3–8 Mathematics Tests.

The makeup test administration windows allowed students who were ill or otherwise unable to test during the assigned window to take the tests. The makeup test administration window for PBT was Friday, April 5–Tuesday, April 9 for the Grades 3–8 ELA Tests and Monday, May 5–Wednesday, May 8 for the Grades 3–8 Mathematics Tests. The makeup test administration window for CBT was Thursday, April 4–Thursday, April 11 for the Grades 3–8 ELA Tests and Friday, May 5–Friday, May 10 for the Grades 3–8 Mathematics Tests.

4.2. Scoring Procedures of Operational Tests

Qualified teachers and administrators performed the scoring of the NYSTP 2019 Grades 3–8 ELA and Mathematics Tests at designated sites. The number of personnel at a given site varied, as districts have the option of regional, district-wide, or school-wide scoring (please refer to Section 4.3: Scoring Models for more details). Administrators were responsible for the oversight of scoring operations, including the preparation of the test site, the security of test materials, and the supervision of the scoring process. At each site, designated trainers taught scoring committee members the basic criteria for scoring each item and monitored the scoring sessions in the room. Facilitators or leaders, who also helped in monitoring the sessions and enforced scoring accuracy, assisted the trainers.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, a site coordinator conducted oversight. A scoring leader trained the scoring committee members and monitored the sessions, and a table facilitator assisted in monitoring the sessions. For each subject, the oversight was structured in the same way for district- and school-wide models. At the district-wide level, a school district administrator oversaw scoring. A district subject leader trained the scoring committee members and monitored the sessions, and a school subject leader assisted in monitoring the sessions. For school-wide scoring, oversight was provided by the principal; otherwise, titles for the school-wide model were the same as those for the district-wide model. The general title "scoring-committee members" included scorers at every site. Details on titles and responsibilities for each scoring model can be found on page 16 of the 2019 School Administrator's Manual.

The processes for PBT and CBT are the same excluding the following exceptions:

- For CBT, two schools within a district (Scoring Model 4) and one school (Scoring Model 5) are not permitted. Refer to page 15 of the *2019 Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* for descriptions of all of the scoring models.
- For CBT, scorers use the ScorePoint system to score responses.

4.3. Scoring Models

For the 2018–2019 school year, schools and school districts were able to score Grades 3–8 ELA and/or Mathematics Tests regionally, multi-district, district-wide, or school-wide, based on local need. Schools were required to enter one of the following scoring model codes on student answer sheets:

1. Regional scoring—The scorers for the school's test papers included either staff from three or more school districts or staff from all religious and independent schools in an affiliation group (religious and independent or charter schools may participate in regional scoring with public school districts, and may be counted as one district).
2. Schools from two districts—The scorers for the school's test papers included staff from two school districts, religious and independent schools, charter school districts, or a combination thereof.
3. Three or more schools within a district—The scorers for the school's test papers included staff from all schools administering this test in a district, provided that at least three schools are represented.
4. Two schools within a district—The scorers for the school's test papers included staff from all schools administering this test in a district, provided that two schools are represented (not available for CBT schools).
5. One school, only (local scoring)—The first readers for the school's test papers included staff from the only school in the district administering this test, staff from one charter school, or staff from one religious and independent school (not available for CBT schools).
6. Private contractor—Scored by a private contractor that does not belong to Boards of Cooperative Educational Services (BOCES).

Schools and districts were instructed to carefully analyze their individual needs and capacities to determine their appropriate scoring model. BOCES and the Staff and Curriculum Development Network (SCDN) provided districts with technical support and advice in making this decision.

4.4. Scoring of Constructed-Response Items

The key resources used to train scoring committee members on how to score student responses for constructed-response (CR) items were scoring guides. These guides were created by Questar from sets of actual field-test student responses that were consensus scored by NYSED and New York State teachers during Rangefinding sessions. Trainers used these materials to train scoring committee members on the criteria for scoring CR items and rubric application. Additionally, Scoring Leader Handbooks were distributed to provide guidelines, information, and procedures for both the Scorers and Scoring Site Coordinators to facilitate scoring.

The constructed response items were divided into three groups for scoring and three separate scoring committee members scored each constructed response item in the group they were assigned. After scoring was completed, the table facilitator or subject (ELA or Mathematics) leader conducted *read behinds* for the Scorers and items assigned to their scoring group.

4.5. Scorer Qualifications and Training

Qualified administrators and teachers conducted the scoring of the 2019 Grades 3–8 ELA and Mathematics Tests. Trainers used the scoring guides to train scoring-committee members on the criteria for scoring constructed-response items. Part of the training process was the administration of a consistency assurance set (CAS) that provided the State's scoring sites with information regarding strengths and weaknesses of their scorers. This tool allowed trainers to retrain their scorers, if necessary. The CAS also acknowledged those scorers who had grasped all aspects of the content area being scored and were well prepared to score student responses.

Regardless of the scoring model used, a minimum of three scorers is necessary to score each student's test. However, to comply with a State requirement, none of the scorers assigned to score a student's test responses may be that student's teacher. This policy is detailed in the *Scoring Leader Handbook* section "Assigning Scorer Numbers and Questions to PBT Scoring Committee Members" on page 20, found online at: http://www.p12.nysed.gov/assessment/sam/ei/scoring-leader-hb19p.pdf.

4.6. Quality Control Process

Test books and electronic responses were randomly distributed throughout each scoring room so that completed tests from each region, district, school, or class were evenly dispersed. Teams were divided into groups of three, in order to ensure that a variety of scorers graded each test. If a scorer and a facilitator could not reach a decision after reviewing the scoring guides, they called the Questar Scoring Helpline. The call center was established to help teachers and administrators during scoring. The helpline staff consisted of trained Questar personnel who answered questions by phone. When a member of the staff was unable to resolve an issue, it was referred to NYSED for a scoring decision. A quality check was also performed, in order to certify that all of the items were scored and that the scoring-committee members darkened each score on the answer document appropriately. The log of calls received by the scoring helpline was delivered to NYSED twice daily during the scoring window. To affirm that all schools across the state adhered to scoring guidelines and policies, approximately 5% of the schools' results are audited each year by an outside vendor.

Section 5: Operational Test Data Collection and Classical Analysis

5.1. Data Collection

Test data were collected in two phases. During Phase 1, a sample of approximately 95% of the student test records was received from the data warehouse and delivered to Questar, beginning at the end of May 2019. During Phase 2, "straggler files" were submitted to Questar in June 2019.

The "straggler files" contained fewer than about 5% of the total population cases, and were excluded from the classical, IRT, and reliability analyses (as described in Sections 5, 6, and 7, respectively) due to late submission. The analyses described in Section 8: Summary of Operational Test Results were based on the data collected from both Phase 1 and Phase 2. Data collected from both public schools and religious and independent schools were included in all data analyses.

5.2. Data Processing

Depending on the nature of the analysis, more student records were included in some analyses than in others. For example, all students with valid test scores were included in the analyses described in Section 8: Summary of Operational Test Results. For the analyses described in other sections, more stringent data cleaning procedures were applied (see details below).

Data processing here refers to the cleaning and screening procedures used to identify errors (such as out-of-range data), and the decisions made to exclude student cases or to suppress particular items in certain analyses. Questar's psychometric team performed data cleaning to the delivered data, and excluded some student cases, in order to obtain a sample of the utmost integrity. It should be noted that a student case being excluded from certain data analyses did not mean that the student record was invalidated. According to the NYSED's specific instructions, additional procedures were taken to correct or recover these students' records so that their test results were scored properly. As mentioned above, their records were included in later analyses (see Section 8).

The major groups of cases excluded from the data set (used for analyses in Sections 5, 6, and 7) were students with missing school type and those with at least one entirely missing test session. Other deleted cases included students with incorrect or incomplete grade information, duplicate record cases, and no-response record cases. The mathematical data cleaning procedure also excluded records with mismatched form language indicators for translated versions across the two test sessions for a given student.

5.2.1. *Sampling Down for Representativeness*

Historically, after data cleaning, the sample is reviewed for representativeness of the prior year's operational population in terms of key variables such as student gender, racial/ethnic identity, student disability status, ELL/MLL status, presence of test accommodation(s), and school Needs/Resource Capacity Category (NRC). At the recommendation of New York State's Assessment Technical Advisory Committee (TAC), Questar shifted the focus from sampling down according to demographic representativeness to instead focus on matching the prior year's population's distribution of ability. Questar and NYSED still reviewed the demographic patterns for 2019 relative to 2018, but they were not used directly in the sampling down analyses.

Comparison results between the final 2019 sample and 2018 operational population are further described in Section 6: IRT Calibration.

The numbers of cases considered for dropping because of sampling down varied across grades and subjects, but the process for all grades was consistent. The cleaned data file for a given subject and grade was the starting point. Questar reviewed the distribution of raw score proportion correct (RSPC) for the 2018 and 2019 operational forms. There were some minor differences in the 2018 and 2019 distributions of RSPC, but overall Questar, NYSED, and its TAC agreed that there was no evidence for a need to sample down in any subject or grade.

The data cleaning procedures and accompanying case counts are represented for ELA and Mathematics in Tables 5.1–5.6 and Tables 5.7–5.12, respectively.

Table 5.1. ELA Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	201,207
Wrong Subject	0	201,207
No Grade	0	201,207
Wrong Grade	77	201,130
Form Code Mismatch	958	200,172
Language or Mismatched Form	0	200,172
School Type	221	199,951
Missing Entire Session	22,353	177,598
Invalid Score	12	177,586
Not Tested Reason	0	177,586
Out-of-Range CR Scores	0	177,586
Duplicated Record	29	177,557
Test Mode Discrepancy	54	177,503

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.2. ELA Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	207,478
Wrong Subject	0	207,478
No Grade	0	207,478
Wrong Grade	77	207,401
Form Code Mismatch	1,003	206,398
Language or Mismatched Form	0	206,398
School Type	368	206,030
Missing Entire Session	25,624	180,406
Invalid Score	11	180,395
Not Tested Reason	0	180,395

Exclusion Rule	# Deleted	# Cases Remain
Out-of-Range CR Scores	0	180,395
Duplicated Record	33	180,362
Test Mode Discrepancy	53	180,309

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.3. ELA Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	207,447
Wrong Subject	0	207,447
No Grade	2	207,445
Wrong Grade	73	207,372
Form Code Mismatch	1,046	206,326
Language or Mismatched Form	0	206,326
School Type	509	205,817
Missing Entire Session	29,880	175,937
Invalid Score	16	175,921
Not Tested Reason	0	175,921
Out-of-Range CR Scores	0	175,921
Duplicated Record	43	175,878
Test Mode Discrepancy	495	175,383

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.4. ELA Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	212,803
Wrong Subject	0	212,803
No Grade	1	212,802
Wrong Grade	114	212,688
Form Code Mismatch	1,248	211,440
Language or Mismatched Form	0	211,440
School Type	1,110	210,330
Missing Entire Session	34,115	176,215
Invalid Score	17	176,198
Not Tested Reason	0	176,198
Out-of-Range CR Scores	0	176,198
Duplicated Record	15	176,183
Test Mode Discrepancy	1	176,182

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.5. ELA Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	202,191
Wrong Subject	0	202,191
No Grade	3	202,188
Wrong Grade	98	202,090
Form Code Mismatch	1,124	200,966
Language or Mismatched Form	0	200,966
School Type	835	200,131
Missing Entire Session	40,197	159,934
Invalid Score	11	159,923
Not Tested Reason	0	159,923
Out-of-Range CR Scores	0	159,923
Duplicated Record	11	159,912
Test Mode Discrepancy	0	159,912

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.6. ELA Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	203,111
Wrong Subject	0	203,111
No Grade	0	203,111
Wrong Grade	98	203,013
Form Code Mismatch	1,077	201,936
Language or Mismatched Form	0	201,936
School Type	1,329	200,607
Missing Entire Session	48,606	152,001
Invalid Score	16	151,985
Not Tested Reason	0	151,985
Out-of-Range CR Scores	0	151,985
Duplicated Record	19	151,966
Test Mode Discrepancy	0	151,966

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.7. Mathematics Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	206,317
Wrong Subject	0	206,317
No Grade	0	206,317
Wrong Grade	79	206,238

Exclusion Rule	# Deleted	# Cases Remain
Form Code Mismatch	2,617	203,621
Language or Mismatched Form	0	203,621
School Type	146	203,475
Missing Entire Session	22,031	181,444
Invalid Score	14	181,430
Not Tested Reason	0	181,430
Out-of-Range CR Scores	0	181,430
Duplicated Record	27	181,403
Test Mode Discrepancy	1	181,402

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.8. Mathematics Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	211,067
Wrong Subject	0	211,067
No Grade	1	211,066
Wrong Grade	90	210,976
Form Code Mismatch	2,868	208,108
Language or Mismatched Form	0	208,108
School Type	262	207,846
Missing Entire Session	25,473	182,373
Invalid Score	7	182,366
Not Tested Reason	0	182,366
Out-of-Range CR Scores	0	182,366
Duplicated Record	33	182,333
Test Mode Discrepancy	0	182,333

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.9. Mathematics Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	209,313
Wrong Subject	0	209,313
No Grade	0	209,313
Wrong Grade	85	209,228
Form Code Mismatch	7,597	201,631
Language or Mismatched Form	0	201,631
School Type	256	201,375
Missing Entire Session	30,375	171,000
Invalid Score	3	170,997

Exclusion Rule	# Deleted	# Cases Remain
Not Tested Reason	0	170,997
Out-of-Range CR Scores	0	170,997
Duplicated Record	39	170,958
Test Mode Discrepancy	0	170,958

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.10. Mathematics Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	213,355
Wrong Subject	0	213,355
No Grade	0	213,355
Wrong Grade	122	213,233
Form Code Mismatch	7,352	205,881
Language or Mismatched Form	0	205,881
School Type	214	205,667
Missing Entire Session	36,146	169,521
Invalid Score	81	169,440
Not Tested Reason	0	169,440
Out-of-Range CR Scores	0	169,440
Duplicated Record	13	169,427
Test Mode Discrepancy	0	169,427

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

Table 5.11. Mathematics Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	210,631
Wrong Subject	0	210,631
No Grade	0	210,631
Wrong Grade	102	210,529
Form Code Mismatch	39,739	170,790
Language or Mismatched Form	0	170,790
School Type	273	170,517
Missing Entire Session	10,578	159,939
Invalid Score	9	159,930
Not Tested Reason	0	159,930
Out-of-Range CR Scores	0	159,930
Duplicated Record	14	159,916
Test Mode Discrepancy	1	159,915

Table 5.12. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	159,292
Wrong Subject	0	159,292
No Grade	0	159,292
Wrong Grade	98	159,194
Form Code Mismatch	5,655	153,539
Language or Mismatched Form	0	153,539
School Type	278	153,261
Missing Entire Session	46,155	107,106
Invalid Score	3	107,103
Not Tested Reason	0	107,103
Out-of-Range CR Scores	0	107,103
Duplicated Record	14	107,089
Test Mode Discrepancy	0	107,089

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (i.e., refusal or absentee rates).

5.3. Classical Analysis and Calibration Sample Characteristics

The cleaned and sampled-down (if needed) data were used for classical analyses and calibration. The demographic characteristics of students in these data sets are presented in Tables 5.13–5.18 and Tables 5.19–5.24 for ELA and Mathematics, respectively. The Needs/Resource Capacity Category (NRC) is assigned at the district level and is an indicator of district and school socioeconomic status. The ethnicity and gender designations are based on student-level information.

Table 5.13. ELA Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	87,733	49.43
	Male	89,770	50.57
Ethnicity	Asian	17,600	9.92
	African American	30,276	17.07
	Hispanic	50,893	28.69
	American Indian	1,213	0.68
	Multiracial	5,514	3.11
	Pacific Islander	388	0.22
	White	71,495	40.31
NRC	New York	64,285	36.22
	Big 4 Cities	7,405	4.17
	Urban/Suburban	14,459	8.15
	High Needs Rural	9,685	5.46
	Average Needs	42,930	24.19

Demographic Category		N-Count	% of Total N-Count
NRC	Low Needs	19,085	10.75
	Charter School	12,483	7.03
	Religious and Independent	7,171	4.04
SWD	No	151,097	85.12
	Yes	26,406	14.88
SUA	No	154,155	86.85
	Yes	23,348	13.15
ELL/ MLL	No	158,475	89.28
	Yes	19,028	10.72
SWD/ SUA	No	156,785	88.33
	Yes	20,718	11.67
ELL/ MLL/ SUA	No	173,863	97.95
	Yes	3,640	2.05

*The total n-count was 177,503.

Table 5.14. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	88,946	49.33
	Male	91,363	50.67
Ethnicity	Asian	18,346	10.18
	African American	31,761	17.63
	Hispanic	51,532	28.60
	American Indian	1,210	0.67
	Multiracial	5,291	2.94
	Pacific Islander	435	0.24
	White	71,579	39.73
NRC	New York	66,425	36.84
	Big 4 Cities	7,755	4.30
	Urban/Suburban	14,124	7.83
	High Needs Rural	9,782	5.43
	Average Needs	41,818	23.19
	Low Needs	18,751	10.40
	Charter School	11,552	6.41
	Religious and Independent	10,102	5.60
SWD	No	152,152	84.38
	Yes	28,157	15.62

Demographic Category		N-Count	% of Total N-Count
SUA	No	153,485	85.12
	Yes	26,824	14.88
ELL/ MLL	No	162,833	90.31
	Yes	17,476	9.69
SWD/ SUA	No	156,743	86.93
	Yes	23,566	13.07
ELL/ MLL/ SUA	No	176,301	97.78
	Yes	4,008	2.22

*The total n-count was 180,309.

Table 5.15. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	86,147	49.12
	Male	89,236	50.88
Ethnicity	Asian	18,488	10.55
	African American	31,361	17.90
	Hispanic	49,603	28.31
	American Indian	1,175	0.67
	Multiracial	4,768	2.72
	Pacific Islander	456	0.26
	White	69,342	39.58
NRC	New York	65,311	37.24
	Big 4 Cities	7,597	4.33
	Urban/Suburban	13,673	7.80
	High Needs Rural	9,640	5.50
	Average Needs	40,861	23.30
	Low Needs	18,487	10.54
	Charter School	12,070	6.88
	Religious and Independent	7,744	4.42
SWD	No	147,134	83.89
	Yes	28,249	16.11
SUA	No	148,112	84.45
	Yes	27,271	15.55
ELL/ MLL	No	160,273	91.38
	Yes	15,110	8.62
SWD/ SUA	No	151,525	86.40

Demographic Category		N-Count	% of Total N-Count
SWD/ SUA	Yes	23,858	13.60
ELL/ MLL/ SUA	No	171,725	97.91
	Yes	3,658	2.09

*The total n-count was 175,383.

Table 5.16. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	86,677	49.20
	Male	89,505	50.80
Ethnicity	Asian	18,882	10.73
	African American	31,895	18.12
	Hispanic	49,776	28.28
	American Indian	1,255	0.71
	Multiracial	4,398	2.50
	Pacific Islander	527	0.30
	White	69,295	39.37
NRC	New York	65,716	37.30
	Big 4 Cities	7,424	4.21
	Urban/Suburban	12,659	7.19
	High Needs Rural	9,281	5.27
	Average Needs	38,543	21.88
	Low Needs	18,067	10.25
	Charter School	12,355	7.01
	Religious and Independent	12,137	6.89
SWD	No	148,381	84.22
	Yes	27,801	15.78
SUA	No	149,783	85.02
	Yes	26,399	14.98
ELL/ MLL	No	161,459	91.64
	Yes	14,723	8.36
SWD/ SUA	No	153,383	87.06
	Yes	22,799	12.94
ELL/ MLL/ SUA	No	172,699	98.02
	Yes	3,483	1.98

*The total n-count was 176,182.

Table 5.17. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	77,868	48.69
	Male	82,044	51.31
Ethnicity	Asian	18,014	11.27
	African American	27,460	17.19
	Hispanic	44,086	27.59
	American Indian	1,113	0.70
	Multiracial	3,636	2.28
	Pacific Islander	579	0.36
	White	64,890	40.61
NRC	New York	64,465	40.31
	Big 4 Cities	6,719	4.20
	Urban/Suburban	11,899	7.44
	High Needs Rural	9,025	5.64
	Average Needs	35,313	22.08
	Low Needs	17,858	11.17
	Charter School	4,326	2.71
	Religious and Independent	10,307	6.45
SWD	No	134,750	84.27
	Yes	25,162	15.73
SUA	No	135,552	84.77
	Yes	24,360	15.23
ELL/ MLL	No	147,841	92.45
	Yes	12,071	7.55
SWD/ SUA	No	139,133	87.01
	Yes	20,779	12.99
ELL/ MLL/ SUA	No	157,047	98.21
	Yes	2,865	1.79

*The total n-count was 159,912.

Table 5.18. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	73,609	48.44
	Male	78,357	51.56
Ethnicity	Asian	17,154	11.30
	African American	29,488	19.42

Demographic Category		N-Count	% of Total N-Count
Ethnicity	Hispanic	42,959	28.29
	American Indian	1,215	0.80
	Multiracial	2,884	1.90
	Pacific Islander	447	0.29
	White	57,722	38.01
NRC	New York	63,250	41.62
	Big 4 Cities	6,284	4.14
	Urban/Suburban	10,461	6.88
	High Needs Rural	8,213	5.40
	Average Needs	31,286	20.59
	Low Needs	15,688	10.32
	Charter School	9,988	6.57
	Religious and Independent	6,796	4.47
SWD	No	127,331	83.79
	Yes	24,635	16.21
SUA	No	128,405	84.50
	Yes	23,561	15.50
ELL/MLL	No	141,110	92.86
	Yes	10,856	7.14
SWD/ SUA	No	131,748	86.70
	Yes	20,218	13.30
ELL/ MLL/ SUA	No	149,508	98.38
	Yes	2,458	1.62

*The total n-count was 151,966.

Table 5.19. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	89,445	49.31
	Male	91,957	50.69
Ethnicity	Asian	17,969	9.92
	African American	30,035	16.58
	Hispanic	51,277	28.31
	American Indian	1,205	0.67
	Multiracial	5,477	3.02
	Pacific Islander	402	0.22
	White	74,787	41.28
NRC	New York	64,675	35.65

Demographic Category		N-Count	% of Total N-Count
NRC	Big 4 Cities	7,544	4.16
	Urban/Suburban	14,688	8.10
	High Needs Rural	9,846	5.43
	Average Needs	43,232	23.83
	Low Needs	19,303	10.64
	Charter School	12,037	6.64
	Religious and Independent	10,077	5.56
SWD	No	155,498	85.72
	Yes	25,904	14.28
SUA	No	159,605	87.98
	Yes	21,797	12.02
ELL/ MLL	No	159,555	87.96
	Yes	21,847	12.04
SWD/ SUA	No	162,099	89.36
	Yes	19,303	10.64
ELL/ MLL/ SUA	No	177,834	98.03
	Yes	3,568	1.97

*The total n-count was 181,402.

Table 5.20. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	89,939	49.33
	Male	92,394	50.67
Ethnicity	Asian	18,721	10.28
	African American	31,428	17.26
	Hispanic	51,770	28.43
	American Indian	1,209	0.66
	Multiracial	5,275	2.90
	Pacific Islander	440	0.24
	White	73,279	40.24
NRC	New York	66,458	36.45
	Big 4 Cities	7,844	4.30
	Urban/Suburban	14,218	7.80
	High Needs Rural	9,892	5.43
	Average Needs	42,304	23.20
	Low Needs	19,086	10.47
	Charter School	11,210	6.15

Demographic Category		N-Count	% of Total N-Count
NRC	Religious and Independent	11,321	6.21
SWD	No	155,211	85.13
	Yes	27,122	14.87
SUA	No	157,828	86.56
	Yes	24,505	13.44
ELL/ MLL	No	162,569	89.16
	Yes	19,764	10.84
SWD/ SUA	No	160,699	88.13
	Yes	21,634	11.87
ELL/ MLL/ SUA	No	178,485	97.89
	Yes	3,848	2.11

*The total n-count was 182,333.

Table 5.21. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	83,825	49.03
	Male	87,133	50.97
Ethnicity	Asian	18,888	11.06
	African American	30,274	17.73
	Hispanic	48,722	28.54
	American Indian	1,172	0.69
	Multiracial	4,500	2.64
	Pacific Islander	463	0.27
	White	66,702	39.07
NRC	New York	65,842	38.51
	Big 4 Cities	7,051	4.12
	Urban/Suburban	12,678	7.42
	High Needs Rural	8,917	5.22
	Average Needs	39,166	22.91
	Low Needs	18,301	10.70
	Charter School	11,632	6.80
	Religious and Independent	7,371	4.31
SWD	No	148,284	86.74
	Yes	22,674	13.26
SUA	No	150,782	88.20
	Yes	20,176	11.80

Demographic Category		N-Count	% of Total N-Count
ELL/ MLL	No	154,464	90.35
	Yes	16,494	9.65
SWD/ SUA	No	153,490	89.78
	Yes	17,468	10.22
ELL/ MLL/ SUA	No	167,940	98.23
	Yes	3,018	1.77

*The total n-count was 170,958.

Table 5.22. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	83,258	49.14
	Male	86,169	50.86
Ethnicity	Asian	19,153	11.32
	African American	30,429	17.98
	Hispanic	48,292	28.54
	American Indian	1,195	0.71
	Multiracial	4,127	2.44
	Pacific Islander	522	0.31
	White	65,502	38.71
NRC	New York	65,114	38.43
	Big 4 Cities	6,834	4.03
	Urban/Suburban	11,814	6.97
	High Needs Rural	8,628	5.09
	Average Needs	36,606	21.61
	Low Needs	17,830	10.52
	Charter School	11,823	6.98
	Religious and Independent	10,778	6.36
SWD	No	147,180	86.87
	Yes	22,247	13.13
SUA	No	150,209	88.66
	Yes	19,218	11.34
ELL/ MLL	No	154,295	91.07
	Yes	15,132	8.93
SWD/ SUA	No	152,855	90.22
	Yes	16,572	9.78

Demographic Category		N-Count	% of Total N-Count
ELL/ MLL/ SUA	No	166,610	98.34
	Yes	2,817	1.66

*The total n-count was 169,427.

Table 5.23. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,603	49.15
	Male	81,312	50.85
Ethnicity	Asian	17,973	11.25
	African American	28,755	18.00
	Hispanic	45,020	28.18
	American Indian	1,093	0.68
	Multiracial	3,458	2.16
	Pacific Islander	589	0.37
	White	62,843	39.34
NRC	New York	63,197	39.52
	Big 4 Cities	6,113	3.82
	Urban/Suburban	10,944	6.84
	High Needs Rural	8,204	5.13
	Average Needs	33,003	20.64
	Low Needs	17,170	10.74
	Charter School	10,440	6.53
	Religious and Independent	10,844	6.78
SWD	No	138,954	86.89
	Yes	20,961	13.11
SUA	No	141,294	88.36
	Yes	18,621	11.64
ELL/ MLL	No	146,705	91.74
	Yes	13,210	8.26
SWD/ SUA	No	144,056	90.08
	Yes	15,859	9.92
ELL/ MLL/ SUA	No	157,586	98.54
	Yes	2,329	1.46

*The total n-count was 159,915.

Table 5.24. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	50,691	47.34
	Male	56,398	52.66
Ethnicity	Asian	10,401	9.73
	African American	22,170	20.74
	Hispanic	32,923	30.8
	American Indian	881	0.82
	Multiracial	1,918	1.79
	Pacific Islander	337	0.32
	White	38,276	35.8
NRC	New York	47,991	44.81
	Big 4 Cities	4,993	4.66
	Urban/Suburban	7,265	6.78
	High Needs Rural	6,221	5.81
	Average Needs	20,471	19.12
	Low Needs	8,292	7.74
	Charter School	6,605	6.17
	Religious and Independent	5,251	4.9
SWD	No	89,583	83.65
	Yes	17,506	16.35
SUA	No	91,653	85.59
	Yes	15,436	14.41
ELL/ MLL	No	97,008	90.59
	Yes	10,081	9.41
SWD/ SUA	No	93,923	87.71
	Yes	13,166	12.29
ELL/ MLL/ SUA	No	105,400	98.42
	Yes	1,689	1.58

*The total n-count was 107,089.

5.4. Classical Data Analysis

Classical data analysis of the NYSTP Grades 3–8 ELA and Mathematics Tests consists of several important elements. One element is the analysis of item-level statistical information about student performance. It is important to verify that the items and test forms function as intended. If any serious error were to occur with an item, errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution) during item analysis. Analyses of test-level data comprise the second element of classical data analysis. These include examination of the raw score (RS) statistics (mean and standard deviation or "SD") and test

reliability measures Cronbach's alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). Additionally, classical DIF analysis is conducted at this stage. DIF analysis includes computation of standardized mean differences and Mantel-Haenszel statistics for New York State items to identify potential item bias. All classical data analysis results contribute information on the validity and reliability of the tests (see also Section 3: Validity, and Section 7: Reliability and Standard Error of Measurement).

5.4.1. *Item Difficulty and Point-Biserial Correlation Coefficients*

Item difficulty is classically measured by the *p*-value statistic. It assesses the proportion of students who responded correctly to each MC item or the average proportion of the maximum score that students earned on each CR item. It is important to have a good range of *p*-values to increase test reliability and to avoid floor or ceiling effects. *P*-values represent the overall degree of difficulty, but do not account for demonstrated student performance on other test items. Usually, *p*-value information is coupled with point-biserial (pbis) statistics, to verify that items are functioning as intended. In Appendix M, Tables M1–M12 illustrate classical test statistics for all items on each grade-level test. Appendix F provides general psychometric guidelines for operational item selection.

Item difficulties (*p*-values) ranged from 0.23 to 0.91 for the ELA tests and 0.32 to 0.93 on the Mathematics tests. These statistics are provided in Appendix M, Tables M1–M12, along with other classical test statistics.

Point-biserial statistics are used to examine item-test correlations, or item discrimination, for MC items. The pbis correlation for the key (i.e., the correct answer) is a measure of internal consistency, while pbis for specific response options aid in flagging possible alternate keys; each is a correlation that ranges between +/–1. It is the correlation of students' responses to an item relative to their performance on the rest of the test and, unless otherwise noted, this discussion will be limited to the point-biserial of the correct response with the remainder of the test.

Point-biserial correlations from the operational analyses are presented in Appendix M, Tables M1–M12. The column labeled "Pbis Key" contains the point-biserial correlation associated with the correct response. The guideline for building the NYSTP Grades 3–8 ELA and Mathematics Tests was that the point-biserial correlation for the key for MC items should be equal to or greater than 0.20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. The few exceptions to this guideline were due to content considerations that required the inclusion of particular items. Decisions to use such items were made very carefully, and no item with a negative point-biserial correlation was allowed on the test.

Point-biserials for correct answer options on the ELA tests ranged from 0.10 to 0.72, as shown in Appendix M, Tables M1–M6. For Grade 3, the item pbis values ranged from 0.28 to 0.64, with a mean of 0.45. For Grade 4, the item pbis values ranged from 0.25 to 0.64, with a mean of 0.42. For Grade 5, the item pbis values ranged from 0.18 to 0.61, with a mean of 0.43. For Grade 6, the item pbis values ranged from 0.10 to 0.64, with a mean of 0.42. For Grade 7, the item pbis values ranged from 0.12 to 0.72, with a mean of 0.43. For Grade 8, the item pbis values ranged from 0.22 to 0.68, with a mean of 0.43.

Point-biserials for correct answer options on the Mathematics tests ranged from 0.22 to 0.77, as shown in Appendix M, Tables M7–M12. For Grade 3, the item pbis values ranged from 0.29 to 0.69, with a mean of 0.48. For Grade 4, the item pbis values ranged from 0.22 to 0.67, with a mean of 0.49. For Grade 5, the item pbis values ranged from 0.27 to 0.74, with a mean of 0.50. For Grade 6, the item pbis values ranged from 0.33 to 0.72, with a mean of 0.52. For Grade 7, the item pbis values ranged from 0.28 to 0.77, with a mean of 0.51. For Grade 8, the item pbis values ranged from 0.31 to 0.72, with a mean of 0.48.

5.4.2. *Omit Rates*

Omit rates (i.e., percentage of students not answering a given item) are routinely checked, based on test data, after each administration. Tables M1–M12 in Appendix M show the omit rates for items on the Grades 3–8 ELA and Mathematics Tests, respectively. The industry standard general rule of thumb is that omit rates for multiple-choice items should be less than 5%. Omit rates across multiple-choice and constructed-response items on the Grades 3–8 ELA and Mathematics Tests typically ranged from 0% to 3%, remaining within the acceptable range for large-scale achievement tests. As may be expected, omit rates tended to increase for items at the end of the test sessions.

5.4.3. *Differential Item Functioning (DIF)*

Classical differential item functioning (DIF) analyses are statistical methods for identifying items that are estimated to have functioned differently for one group (i.e., the "focal" group) as compared with another group (i.e., the "reference" group). In other words, DIF analysis only *flags* items that may later be judged by content experts to exhibit bias, rather than directly detecting bias. First, the psychometric phenomenon of DIF was extensively investigated and experts' judgments of bias collected when items were field-tested, which reduced the likelihood of including any differentially functioning items on the operational forms. Turning to the analysis of this year's operational data, as discussed in Section 3.2.3. Detection of Bias, items flagged for DIF do not necessarily indicate item bias. For example, DIF may be attributed to true group differences on the content measured by the item or Type I error, which refers to statistically flagging items that have no true DIF. Operational items flagged for DIF are given additional scrutiny by content specialists, above and beyond the existing rounds of reviews by New York State educators, and those content specialists make the final judgment as to whether or not an item is biased for or against the focal group.

DIF was evaluated using two methods, both of which involve checks on statistical and practical significance. First, the Mantel-Haenszel (MH) method is employed for MC items. This non-parametric DIF method partitions the sample of examinees into categories based on total raw test scores. It then compares the log-odds ratio of keyed responses for the focal and reference groups. In terms of statistical significance, the Mantel-Haenszel method has a critical value of 6.63 (degrees of freedom = 1 for MC items; alpha = 0.01) and as far as practical significance is concerned, it is compared to its corresponding delta-value. Delta-values are a commonly used metric in testing that indicates the magnitude of DIF. Typically, delta-values above 1.50 are considered indicative of moderate DIF that should be examined more closely (Zwick, Donoghue, and Grima, 1993). Second, the standardized mean difference (SMD) was computed for CR items. The SMD statistic (Dorans, Schmitt, and Bleistein, 1992) compares the mean scores of reference and focal groups, after adjusting for proficiency differences. The SMD was also evaluated for statistical significance and, in terms of practical significance, a moderate amount of

DIF, for or against the focal group, is represented by an SMD with an absolute value between 0.10 and 0.19, inclusive; a large amount of DIF is represented by an SMD with an absolute value of 0.20 or greater.

Classical DIF analyses were conducted on subgroups of the Needs/Resource Capacity Category (focal group: High Needs; reference group: Low Needs), gender (focal group: Female; reference group: Male), ethnicity (focal groups: African American, Hispanic, and Asian; reference group: White), ELL/MLLs (focal group: ELL/MLLs; reference group: Non-ELLs), and mode (focal group: PBT students; reference group: CBT students). The DIF analyses were conducted using all cases from the clean data sets. Table 5.25 and Table 5.26 show the numbers of cases for the subgroups for ELA and Mathematics, respectively.

Table 5.25. ELA Classical DIF Sample N-Counts

	Ethnicity				Gender		Needs/Resource Capacity Category		English Language Learners/ Multilingual Learners		Mode	
Grade	African American	Hispanic/ Latino	Asian American	White	Female	Male	High	Low	ELL/MLL	Non-ELL/MLL	CBT	PBT
3	30,276	50,893	17,600	71,495	87,733	89,770	95,834	62,015	19,028	158,475	27,115	150,388
4	31,761	51,532	18,346	71,579	88,946	91,363	98,086	60,569	17,476	162,833	28,299	152,010
5	31,361	49,603	18,488	69,342	86,147	89,236	96,221	59,348	15,110	160,273	33,839	141,544
6	31,895	49,776	18,882	69,295	86,677	89,505	95,080	56,610	14,723	161,459	34,548	141,634
7	27,460	44,086	18,014	64,890	77,868	82,044	92,108	53,171	12,071	147,841	31,360	128,552
8	29,488	42,959	17,154	57,722	73,609	78,357	88,208	46,974	10,856	141,110	25,878	126,088

Table 5.26. Mathematics Classical DIF Sample N-Counts

	Ethnicity				Gender		Needs/Resource Capacity Category		English Language Learners/ Multilingual Learners		Mode	
Grade	African American	Hispanic/ Latino	Asian American	White	Female	Male	High	Low	ELL/MLL	Non-ELL/MLL	CBT	PBT
3	30,035	51,277	17,969	74,787	89,445	91,957	96,753	62,535	21,847	159,555	22,128	159,274
4	31,428	51,770	18,721	73,279	89,939	92,394	98,412	61,390	19,764	162,569	21,198	161,135
5	30,274	48,722	18,888	66,702	83,825	87,133	94,488	57,467	16,494	154,464	21,975	148,983
6	30,429	48,292	19,153	65,502	83,258	86,169	92,390	54,436	15,132	154,295	24,435	144,992
7	28,755	45,020	17,973	62,843	78,603	81,312	88,458	50,173	13,210	146,705	23,151	136,764
8	22,170	32,923	10,401	38,276	50,691	56,398	66,470	28,763	10,081	97,008	13,994	93,095

Table 5.27 (ELA) and Table 5.28 (Mathematics) present the number of items flagged for DIF by either of the classical methods described earlier. Appendix N provides a detailed list of items flagged by either one or both of these classical DIF methods, including DIF direction and associated DIF statistics.

Table 5.27. ELA Items Flagged for DIF

Grade	Flagged Items
3	9
4	6
5	4
6	8
7	9
8	13

Table 5.28. Mathematics Items Flagged for DIF

Grade	Flagged Items
3	6
4	5
5	4
6	6
7	3
8	3

As discussed in Section 3: Validity, items showing statistically significant DIF (flagged as described above for MH statistics on MC items and SMD statistics for CR items) do not necessarily pose bias. The items flagged with DIF were examined further by the content experts; no signs of potential content-based issues were discovered. The items are possibly functioning differently statistically.

Section 6: IRT Calibration

6.1. IRT Models and Rationale for Use

IRT allows for comparisons between items and scale scores, even those from different test forms, by using a common scale for all items and examinees (i.e., as if there were a hypothetical test that contained items from all forms). The three-parameter logistic (3PL) model (Lord and Novick, 1968; Lord, 1980) was used to analyze item responses on the MC items. For analysis of the CR items, the two-parameter partial credit (2PPC) model (Muraki, 1992; Yen, 1993) was used.

IRT is a statistical methodology that takes into account the fact that not all test items are alike and that not all test items provide the same amount of information in determining how much a student knows or can do. Computer programs that implement IRT models use actual student data to estimate the characteristics of the items on a test, called "parameters." The parameter estimation process is called "item calibration."

IRT models typically vary according to the number of parameters estimated. For the New York State tests, three parameters are estimated: the discrimination parameter, the difficulty parameter(s), and, for MC items, the guessing parameter. The discrimination parameter is an index of how well an item differentiates between high-performing and low-performing students. An item that cannot be answered correctly by low-performing students but can be answered correctly by high-performing students will have a high-discrimination value. The difficulty parameter is an index of how easy or difficult an item is. The higher the difficulty parameter is, the harder the item is. The guessing parameter is the probability that a student with very low proficiency will answer the item correctly.

Because the characteristics of MC and CR items are different, two IRT models were used in item calibration. The three-parameter logistic (3PL) model was used in the analysis of MC items. In this model, the probability that a student with proficiency θ responds correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]},$$

where

a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response from a very low-scoring student.

For analysis of the CR items, the 2PPC model was used. The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with proficiency θ having a score (k - 1) at the kth level of the jth item is:

$$P_{jk}(\theta) = P(x_j = k-1 \mid \theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \; k = 1 \ldots m_j,$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk},$$

and

k is the item response category ($k = 1, 2, \ldots. m_j$).

The m_j denotes the number of score categories for the jth item, and, typically, the highest score category is assigned (m_j - 1) score points. For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k-1),$$

and

$$C_{jk} = -\sum_{i=0}^{k-1}\gamma_{ji},$$

where

$$\gamma_{j0} = 0,$$

and

α_j and γ_{ji} are the free parameters to be estimated from the data.

Each item has (m_j - 1) independent γ_{ji} parameters and one α_j parameter; a total of m_j parameters are estimated for each CR item.

6.2. Calibration Sample

The cleaned data were used for calibration of the NYSTP 2019 Grades 3–8 ELA and Mathematics Tests. It should be noted that the sample sizes were adequate, as the calibration was performed using nearly all of the New York State public and non-public school student population data in each tested grade. As shown in Tables 6.1–6.3 and Tables 6.4–6.6 for ELA and Mathematics, respectively, the 2019 operational test samples were generally comparable to 2018 populations in terms of NRC, student race and ethnicity, proportions of ELL/MLL students, proportions of students with disabilities, and proportions of students using testing accommodations.

Table 6.1. ELA Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Gender	Female	49.34	49.43	49.30	49.33
	Male	50.66	50.57	50.70	50.67
Ethnicity	Asian	9.80	9.92	10.04	10.18
	African American	17.00	17.07	17.55	17.63
	Hispanic	28.31	28.69	28.17	28.60

Demographic Category		Grade 3		Grade 4	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Ethnicity	American Indian	0.68	0.68	0.66	0.67
	Multiracial	3.04	3.11	2.85	2.94
	Pacific Islander	0.22	0.22	0.24	0.24
	White	40.94	40.31	40.49	39.73
NRC	New York	35.54	36.22	36.00	36.84
	Big 4 Cities	4.12	4.17	4.22	4.30
	Urban/Suburban	8.00	8.15	7.66	7.83
	High Needs Rural	5.44	5.46	5.40	5.43
	Average Needs	23.83	24.19	22.75	23.19
	Low Needs	10.54	10.75	10.14	10.40
	Charter School	6.91	7.03	6.36	6.41
	Religious and Independent	5.60	4.04	7.47	5.60
SWD	No	85.85	85.12	85.53	84.38
	Yes	14.15	14.88	14.47	15.62
SUA	No	92.35	86.85	91.80	85.12
	Yes	7.65	13.15	8.20	14.88
ELL/ MLL	No	89.24	89.28	90.23	90.31
	Yes	10.76	10.72	9.77	9.69
SWD/ SUA	No	94.79	88.33	94.83	86.93
	Yes	5.21	11.67	5.17	13.07
ELL/ MLL/ SUA	No	99.29	97.95	99.33	97.78
	Yes	0.71	2.05	0.67	2.22

Table 6.2. ELA Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Gender	Female	49.06	49.12	49.13	49.20
	Male	50.94	50.88	50.87	50.80
Ethnicity	Asian	10.52	10.55	10.86	10.73
	African American	17.95	17.90	18.36	18.12
	Hispanic	28.03	28.31	28.23	28.28
	American Indian	0.68	0.67	0.70	0.71
	Multiracial	2.65	2.72	2.44	2.50
	Pacific Islander	0.27	0.26	0.30	0.30

Demographic Category		Grade 5		Grade 6	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Ethnicity	White	39.91	39.58	39.12	39.37
NRC	New York	36.81	37.24	37.07	37.30
	Big 4 Cities	4.26	4.33	4.21	4.21
	Urban/Suburban	7.70	7.80	7.16	7.19
	High Needs Rural	5.50	5.50	5.32	5.27
	Average Needs	22.96	23.30	21.79	21.88
	Low Needs	10.32	10.54	10.18	10.25
	Charter School	6.84	6.88	7.02	7.01
	Religious and Independent	5.60	4.42	7.24	6.89
SWD	No	85.24	83.89	85.40	84.22
	Yes	14.76	16.11	14.60	15.78
SUA	No	91.04	84.45	91.00	85.02
	Yes	8.96	15.55	9.00	14.98
ELL/ MLL	No	91.49	91.38	91.96	91.64
	Yes	8.51	8.62	8.04	8.36
SWD/ SUA	No	94.47	86.40	94.43	87.06
	Yes	5.53	13.60	5.57	12.94
ELL/ MLL/ SUA	No	99.31	97.91	99.30	98.02
	Yes	0.69	2.09	0.70	1.98

Table 6.3. ELA Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Gender	Female	48.82	48.69	48.32	48.44
	Male	51.18	51.31	51.68	51.56
Ethnicity	Asian	10.93	11.27	11.24	11.30
	African American	18.72	17.19	19.46	19.42
	Hispanic	27.91	27.59	28.01	28.29
	American Indian	0.70	0.70	0.80	0.80
	Multiracial	2.19	2.28	1.88	1.90
	Pacific Islander	0.36	0.36	0.30	0.29
	White	39.20	40.61	38.32	38.01
NRC	New York	38.31	40.31	41.02	41.62
	Big 4 Cities	4.02	4.20	4.13	4.14

Demographic Category		Grade 7		Grade 8	
		2018 Population	2019 Sample	2018 Population	2019 Sample
NRC	Urban/Suburban	7.12	7.44	6.88	6.88
	High Needs Rural	5.40	5.64	5.36	5.40
	Average Needs	21.06	22.08	20.38	20.59
	Low Needs	10.61	11.17	10.17	10.32
	Charter School	6.60	2.71	6.51	6.57
	Religious and Independent	6.87	6.45	5.53	4.47
SWD	No	85.29	84.27	84.77	83.79
	Yes	14.71	15.73	15.23	16.21
SUA	No	91.26	84.77	91.27	84.50
	Yes	8.74	15.23	8.73	15.50
ELL/ MLL	No	92.78	92.45	93.02	92.86
	Yes	7.22	7.55	6.98	7.14
SWD/ SUA	No	94.72	87.01	94.54	86.70
	Yes	5.28	12.99	5.46	13.30
ELL/ MLL/ SUA	No	99.42	98.21	99.44	98.38
	Yes	0.58	1.79	0.56	1.62

Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Gender	Female	49.09	49.31	49.02	49.33
	Male	50.91	50.69	50.98	50.67
Ethnicity	Asian	9.85	9.92	10.07	10.28
	African American	16.72	16.58	17.21	17.26
	Hispanic	28.48	28.31	28.26	28.43
	American Indian	0.67	0.67	0.66	0.66
	Multiracial	3.00	3.02	2.85	2.90
	Pacific Islander	0.22	0.22	0.24	0.24
	White	41.07	41.28	40.71	40.24
NRC	New York	35.89	35.65	36.29	36.45
	Big 4 Cities	4.12	4.16	4.22	4.30
	Urban/Suburban	8.02	8.10	7.64	7.80
	High Needs Rural	5.38	5.43	5.31	5.43
	Average Needs	23.57	23.83	22.64	23.20

Demographic Category		Grade 3		Grade 4	
		2018 Population	2019 Sample	2018 Population	2019 Sample
NRC	Low Needs	10.51	10.64	10.19	10.47
	Charter School	6.84	6.64	6.27	6.15
	Religious and Independent	5.66	5.56	7.44	6.21
SWD	No	84.61	85.72	84.22	85.13
	Yes	15.39	14.28	15.78	14.87
SUA	No	91.93	87.98	90.45	86.56
	Yes	8.07	12.02	9.55	13.44
ELL/ MLL	No	87.90	87.96	88.84	89.16
	Yes	12.10	12.04	11.16	10.84
SWD/ SUA	No	93.14	89.36	91.90	88.13
	Yes	6.86	10.64	8.10	11.87
ELL/ MLL/ SUA	No	99.03	98.03	98.80	97.89
	Yes	0.97	1.97	1.20	2.11

Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Gender	Female	48.73	49.03	48.85	49.14
	Male	51.27	50.97	51.15	50.86
Ethnicity	Asian	10.54	11.06	10.85	11.32
	African American	17.63	17.73	17.97	17.98
	Hispanic	28.18	28.54	28.31	28.54
	American Indian	0.67	0.69	0.70	0.71
	Multiracial	2.62	2.64	2.42	2.44
	Pacific Islander	0.27	0.27	0.30	0.31
	White	40.08	39.07	39.45	38.71
NRC	New York	37.20	38.51	37.39	38.43
	Big 4 Cities	4.28	4.12	4.19	4.03
	Urban/Suburban	7.65	7.42	7.18	6.97
	High Needs Rural	5.42	5.22	5.24	5.09
	Average Needs	22.67	22.91	21.32	21.61
	Low Needs	10.28	10.70	10.13	10.52
	Charter School	6.77	6.80	6.97	6.98

Demographic Category		Grade 5		Grade 6	
		2018 Population	2019 Sample	2018 Population	2019 Sample
NRC	Religious and Independent	5.74	4.31	7.56	6.36
SWD	No	83.74	86.74	84.05	86.87
	Yes	16.26	13.26	15.95	13.13
SUA	No	91.56	88.20	92.83	88.66
	Yes	8.44	11.80	7.17	11.34
ELL/ MLL	No	90.01	90.35	90.54	91.07
	Yes	9.99	9.65	9.46	8.93
SWD/ SUA	No	93.14	89.78	94.29	90.22
	Yes	6.86	10.22	5.71	9.78
ELL/ MLL/ SUA	No	99.19	98.23	99.42	98.34
	Yes	0.81	1.77	0.58	1.66

Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2018 Population	2019 Sample	2018 Population	2019 Sample
Gender	Female	48.66	49.15	47.33	47.34
	Male	51.34	50.85	52.67	52.66
Ethnicity	Asian	10.84	11.25	9.16	9.73
	African American	18.24	18.00	20.43	20.74
	Hispanic	28.15	28.18	30.07	30.80
	American Indian	0.69	0.68	0.82	0.82
	Multiracial	2.15	2.16	1.79	1.79
	Pacific Islander	0.36	0.37	0.30	0.32
	White	39.57	39.34	37.43	35.80
NRC	New York	38.88	39.52	42.76	44.81
	Big 4 Cities	3.99	3.82	4.73	4.66
	Urban/Suburban	7.07	6.84	6.99	6.78
	High Needs Rural	5.31	5.13	5.96	5.81
	Average Needs	20.59	20.64	18.83	19.12
	Low Needs	10.38	10.74	7.39	7.74
NRC	Charter School	6.61	6.53	6.09	6.17
	Religious and Independent	7.17	6.78	7.26	4.90

Demographic Category		Grade 7		Grade 8	
		2018 Population	2019 Sample	2018 Population	2019 Sample
SWD	No	84.00	86.89	80.70	83.65
	Yes	16.00	13.11	19.30	16.35
SUA	No	91.96	88.36	90.56	85.59
	Yes	8.04	11.64	9.44	14.41
ELL/ MLL	No	91.33	91.74	89.87	90.59
	Yes	8.67	8.26	10.13	9.41
SWD/ SUA	No	93.56	90.08	92.52	87.71
	Yes	6.44	9.92	7.48	12.29
ELL/ MLL/ SUA	No	99.32	98.54	99.35	98.42
	Yes	0.68	1.46	0.65	1.58

6.2.1. *Calibration Process*

The item parameters were estimated using Scientific Software International (SSI) Inc.'s IRTPRO Version 2.1 (Cai, Thissen, & du Toit, 2011) package. MC and CR items were calibrated simultaneously, using marginal maximum likelihood procedures.

The calibration of NYSTP 2019 Grades 3–8 ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were on the original theta scale, and all of the items were well within the prescribed parameter ranges. For both the Grades 3–8 ELA and Mathematics Tests, all calibration estimation results were reasonable. Tables 6.7 and 6.8 present the summaries of the calibration results for ELA and Mathematics, respectively. Additional details, including individual item parameter estimates, may be found in Appendix O, in Tables O13–O24. The parameter estimates are expressed on the theta metric and are defined below:

- MC items:
  - *a*-parameter is a discrimination parameter
  - *b*-parameter is a difficulty parameter
  - *c*-parameter is a guessing parameter

- CR items:
  - *alpha* is a discrimination parameter
  - *step* is a difficulty parameter for category m_j

As described above in Section 6.1, m_j denotes the number of score categories for the jth item, and, typically, the highest score category is assigned (m_j - 1) score points. For the 2PPC model, there are m_j - 1 independent steps and one alpha, for a total of m_j independent parameters estimated for each item, while there is one *a*-parameter and one *b*-parameter per item in the 3PL model.

Table 6.7. ELA Calibration Results

Grade	Item-Level				Student-Level		
	Range of *a*-Parameters		Range of *b*- Parameters		N-Count	Theta Est.*	
						Mean	SD
3	0.495	1.159	-1.809	1.219	177,503	0.00	0.92
4	0.387	1.247	-1.967	1.228	180,309	0.00	0.92
5	0.369	1.740	-1.495	1.913	175,383	0.00	0.92
6	0.127	1.532	-1.722	1.260	176,182	0.00	0.92
7	0.527	1.740	-1.810	1.748	159,912	0.00	0.94
8	0.537	1.514	-1.586	1.149	151,966	-0.01	0.93

*Maximum *a posteriori* (MAP) theta estimates.

Table 6.8. Mathematics Calibration Results

Grade	Item-Level				Student-Level		
	Range of *a*-Parameters		Range of *b*- Parameters		N-Count	Theta Est.*	
						Mean	SD
3	0.554	1.556	-2.352	0.909	181,402	0.01	0.92
4	0.374	2.127	-1.632	0.96	182,333	0.01	0.91
5	0.387	1.883	-1.348	0.779	170,958	0.01	0.91
6	0.698	1.873	-0.628	1.066	169,427	0.03	0.89
7	0.481	2.513	-1.667	0.950	159,915	0.02	0.90
8	0.485	2.062	-0.874	1.022	107,089	0.04	0.89

*Maximum *a posteriori* (MAP) theta estimates.

6.3. Item-Model Fit

Standards for Educational and Psychological Testing (AERA, APA, and NCME, 2014) suggests documenting evidence of model fit when model-based methods such as IRT are used to estimate item parameters in test development. The standard process of assessing the fit of an item under unidimensional IRT models involves steps such as: (a) defining a number of examinee groups ("buckets") and then (b) making an informed judgment by comparing the observed and model-predicted proportion-correct scores for the item of the examinees in different "buckets" (Sinharay, 2006). To make the aforementioned judgment on each item, Hambleton and Swaminathan (1985) have recommended the use of graphical plots comparing the estimated/predicted item response function to the empirical student response data for an item.

To visually examine the model-data fit for each item, fit plots were produced and closely examined for all operational items. Except for one item in grade 7 ELA, all items showed adequate model-data fit. One operational item in grade 7 ELA had a negative slope parameter estimate and a poor fit in the original calibration with no interventions. After adjusting the prior of slope parameter, the estimates and model fit were greatly improved, and the parameter estimates were within the reasonable range. Figure 6.1 below presents the fit plots before and after intervention for this item. The Questar Content team also reviewed this item and confirmed

that the key is correct despite a popular distractor. After reviewing the empirical and theoretical item characteristic curves (ICCs) for all items, Questar did not find evidence that any items should be dropped from operational scoring.

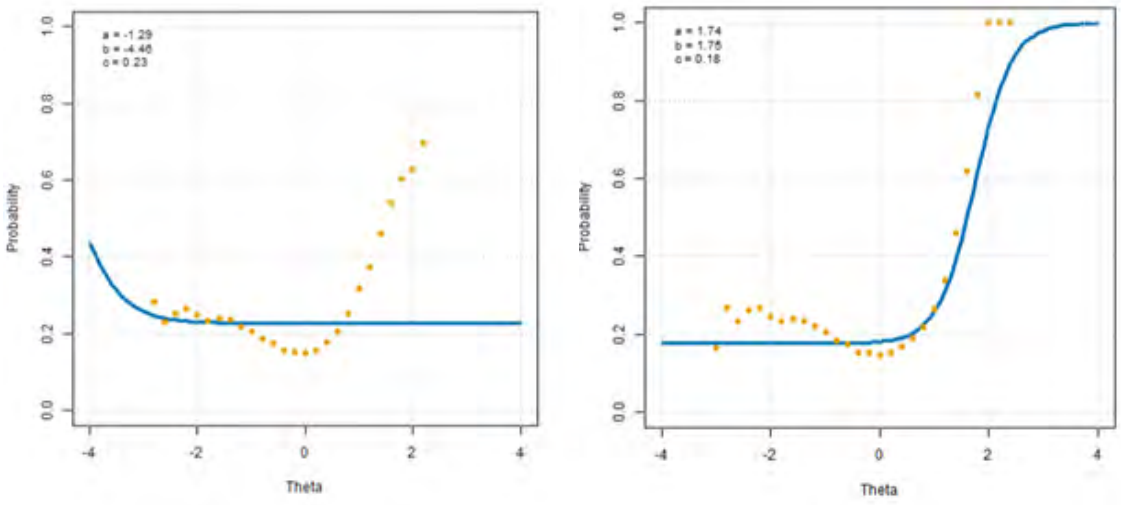

Figure 6.1. Item Model Fit Plots Before and After Intervention

As seem in the example above, any item flagged with extreme item parameters or significant mis-fit was reviewed by both content and psychometric teams. Interventions were applied as needed to improve the parameter estimates and model fit. The fact that the majority of the items in the NYSTP 2019 Grades 3–8 ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models.

6.4. Local Independence

In using IRT models, one of the assumptions made is that the items are locally independent; that a student's response to one item is not dependent upon his or her response to another item. In other words, when a student's proficiency is accounted for, his or her response to each item is statistically independent.

One way to measure the statistical independence of items within a test is via the Q_3 statistic (Yen, 1984). This statistic was obtained by correlating differences between students' observed and expected responses for pairs of items after taking into account overall test performance. The Q_3 statistic for binary items was computed as

$$d_{ij} \equiv u_{ij} - P_j\left(\hat{\theta}_i\right)$$

where $\hat{\theta}_i$ is the estimated trait value (i.e., proficiency) for the *i*th examinee; u_{ij} is the observed probability for the *i*th examinee to get the *j*th item correct and P_j is estimated probability for the *i*th examinee to get the *j*th item correct, and

$$Q_{3jj'} = r\left(d_j, d_{j'}\right).$$

The generalization to items with multiple response categories uses

$$d_{ij} \equiv x_{ij} - E_{ij},$$

where

$$E_{ij} \equiv E\left(x\middle|\hat{\theta}_i\right) = \sum_{k=1}^{m_j} kP_{jk}\left(\hat{\theta}_i\right).$$

If a substantial number of items in the test demonstrate local dependence, these items may need to be calibrated separately. All pairs of items with Q_3 values greater than 0.20 were classified as significant for local dependency. The maximum value for this index is 1.00. When item pairs are flagged by Q_3, the content of the flagged items is examined to identify possible sources of the local dependence. The primary concern about locally dependent items is that they contribute less psychometric information about examinee proficiency than do locally independent items, and therefore inflate score reliability estimates. After reviewing the results and the content of the pairs of items, there was not sufficient evidence to warrant further concern or action regarding the IRT calibration.

6.5. Equating and Scaling

A new reporting scale was established following the Standards Review meeting in summer 2018. The reporting scale was developed to quantify the information captured by the assessment. Because the theta score scale used in the psychometric modeling and the IRT calibration do not appeal to the public, the reporting scale was developed to interpret changes, make comparisons, facilitate inferences, and inform educational decisions.

The purpose of equating was to place the 2019 item parameters and proficiency estimates on the same scale as those in 2018. The following steps constitute the linking process for each subject and grade:

1. All operational items were calibrated in IRTPRO.

2. The 2019 item parameter estimates for all anchor items enabled the establishment of the equating relationship via a test characteristic curve (TCC) method (Stocking and Lord, 1983; implemented in STUIRT, Kim, & Kolen, 2004) to the 2018 theta scale, using the established 2018 item parameter estimates for those same items. Tables 6.9 and 6.10 present the resulting equating coefficients. The following parameters were equated using the formula below:

$$a_i^E = a_i^C \Big/ M_1^E,$$

$$b_i^E = M_1^E \cdot b_i^C + M_2^E \text{, and}$$

$$d_{ij}^E = d_{ij}^C + \left[\left(a_i^C / M_1^E\right)\right] \cdot M_2^E \text{,}$$

where

M_1^E is defined as the multiplicative adjustment for linking and M_2^E is the additive adjustment for linking. The superscript "*E*" denotes linked item parameter estimates, while the superscript "*C*" denotes calibrated item parameter estimates.

Table 6.9. ELA Equating Coefficients

Grade	M_1^E	M_2^E
3	0.9768	-0.0548
4	0.9691	-0.0966
5	1.0740	-0.0644
6	1.1233	-0.1106
7	1.0614	-0.0391
8	1.0023	-0.0183

Table 6.10. Mathematics Equating Coefficients

Grade	M_1^E	M_2^E
3	0.9826	-0.0115
4	1.0242	0.0093
5	1.0222	0.0303
6	1.0069	0.0382
7	0.9734	0.0848
8	1.0247	0.0129

Also, a scaling process was used to determine the transformation from the theta scale to the reporting scale. The following analysis steps were involved in the scaling process:

1. All operational items in the 2019 Grades 3–8 ELA and Mathematics Tests were calibrated using IRT models and transformed onto the same scale as previous year.
2. The raw-to-theta score conversion tables were built up using the test characteristic curve (TCC) approach, based on which each student receives a theta score estimate corresponding to their raw score.
3. For raw scores below the chance level or near the perfect score, the following adjustment and interpolation was conducted to derive the adjusted theta scores:

- At the lower end of the scale, for any theta estimates that were lower than -2.5, 0.25 was subtracted from the preceding adjusted theta value that was within the range.
- At the higher end of the scale, for any theta estimates that were higher than 3.0, 0.25 was added to the previous theta value that was within the range.
- See the table below for an example in the lower end of the scale.

Raw score	Theta	Adjusted theta
7	-3.66491	-3.07129
8	-3.03055	-2.82129
9	-2.62458	-2.57129
10	-2.32129	-2.32129

4. The M_1^S and M_2^S were applied to derive the scale score of each student from their theta score estimate as follows:

$$ScaleScore = (M_1^S \cdot \theta) + M_2^S ,$$

Table 6.11. Operational Scaling Coefficients

Grade	Slope (M_1^S)	Intercept (M_2^S)
ELA		
3	18.310914	600.340994
4	18.276716	600.101132
5	18.212931	600.127742
6	18.309278	600.006654
7	18.318571	600.223246
8	18.308395	600.129092
Mathematics		
3	18.635919	600.082128
4	18.485491	600.009369
5	18.404109	600.040856
6	18.191784	600.432302
7	18.559827	600.499091
8	18.115200	600.640639

6.6. Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2019 TCCs were generated using final item parameters for all reporting test items administered in Spring 2019. TCCs are the summation of all the item characteristic curves (ICCs) for items that contribute to the scale score. Conditional standard error of measurement

(CSEM) curves graphically show the amount of measurement error at different performance levels. The TCCs and CSEM curves are presented in Figures 6.2–6.25.

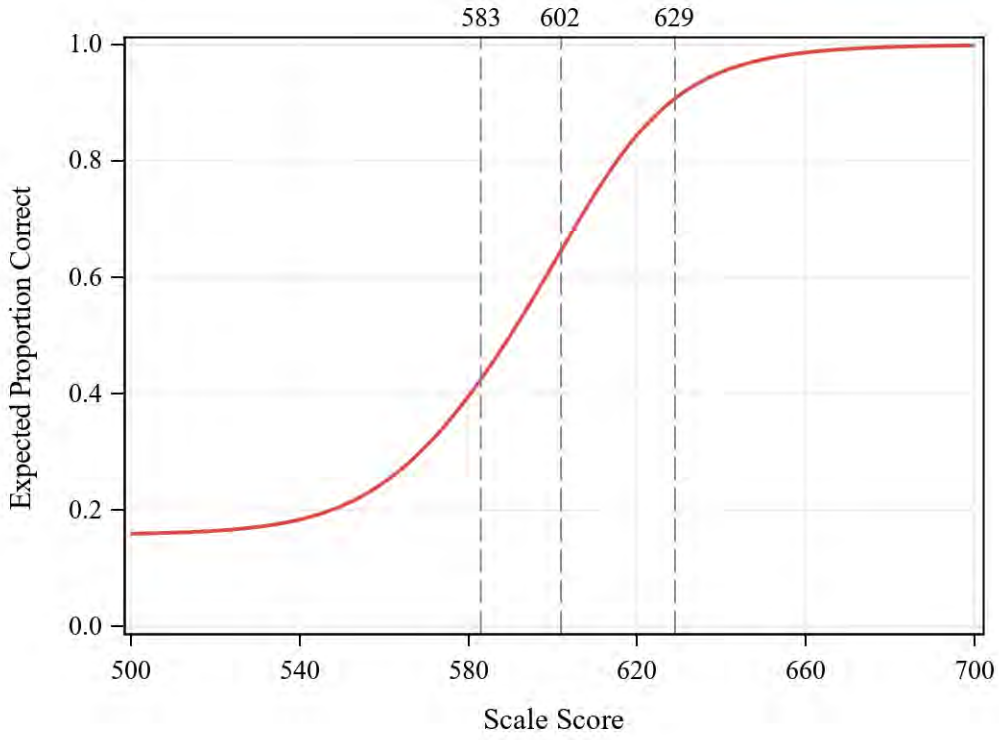

Figure 6.2. ELA Grade 3 TCC

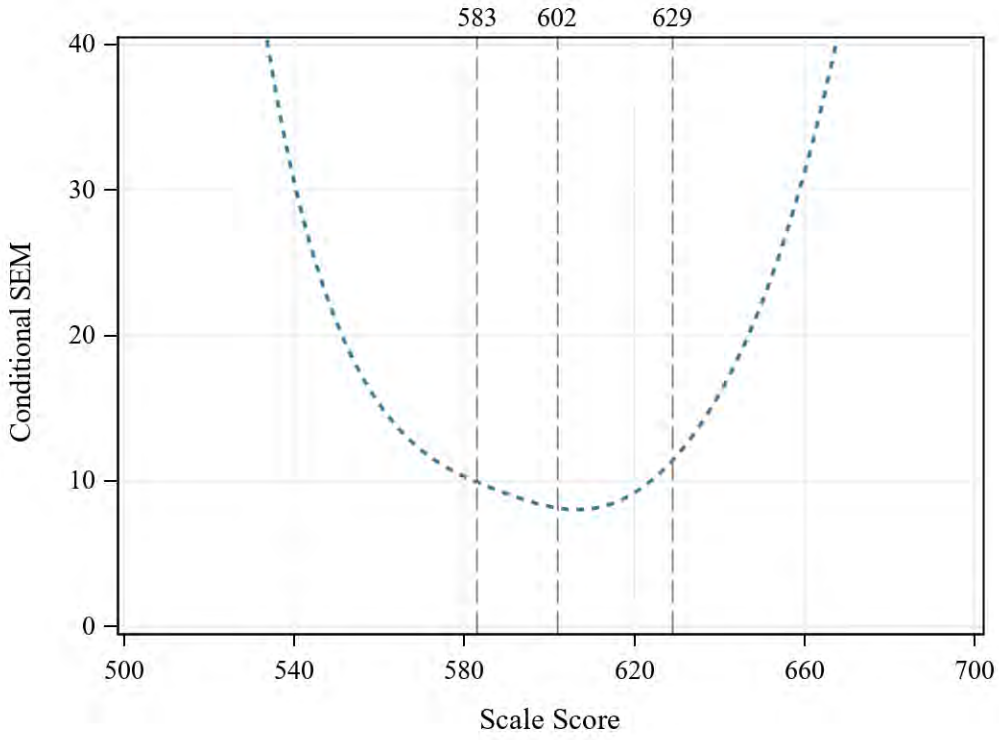

Figure 6.3. ELA Grade 3 CSEM Curve

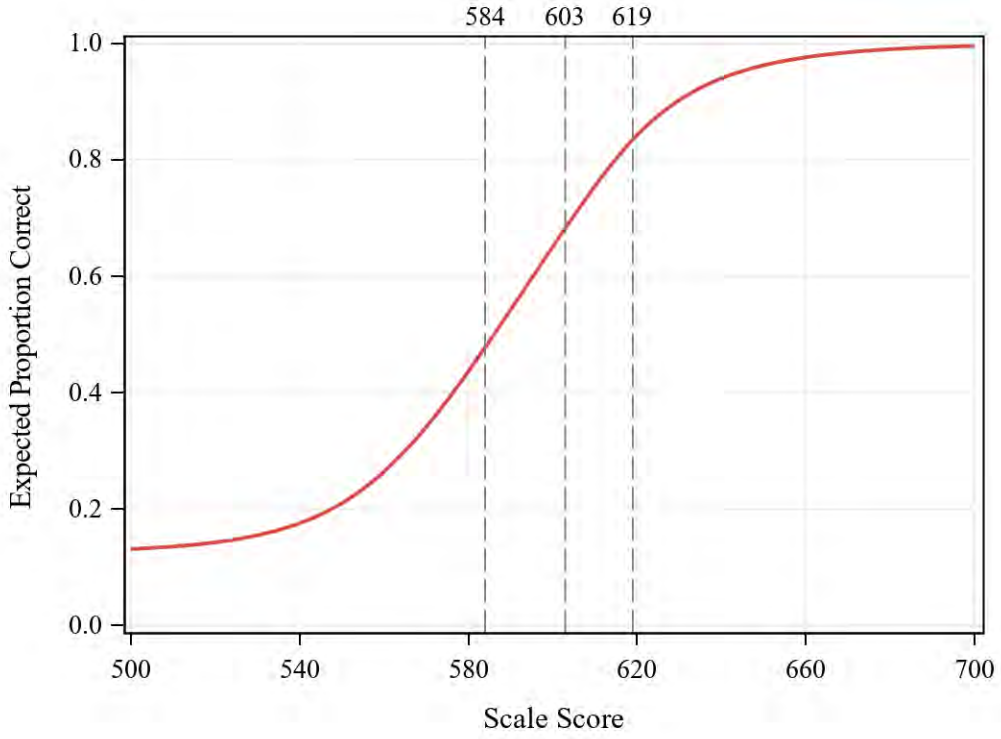

Figure 6.4. ELA Grade 4 TCC

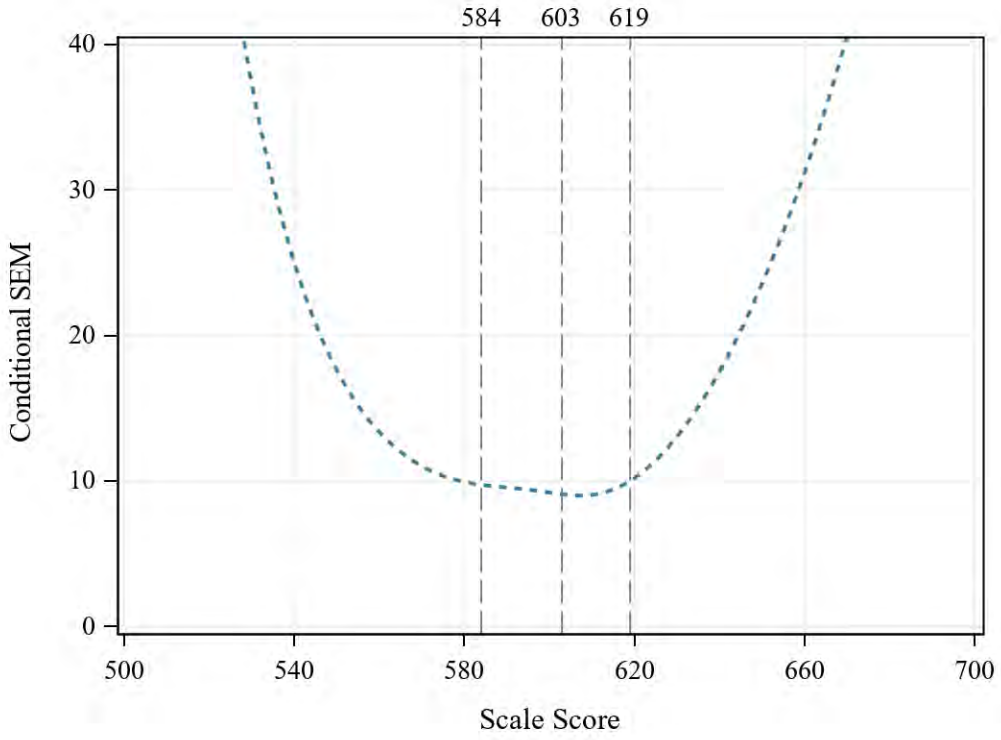

Figure 6.5. ELA Grade 4 CSEM Curve

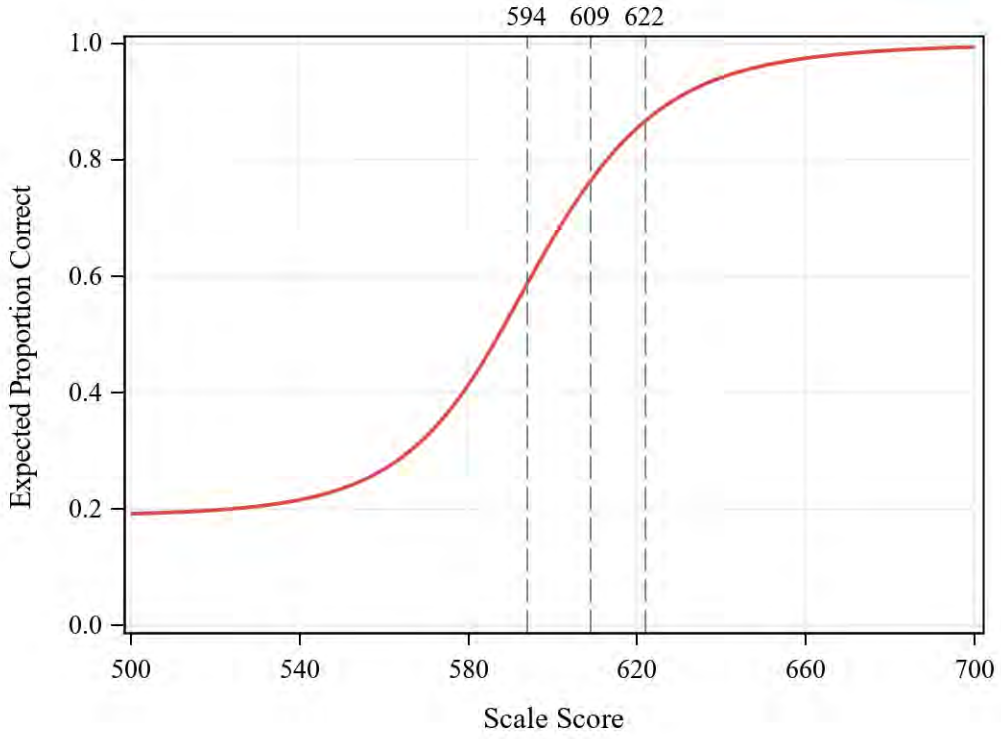

Figure 6.6. ELA Grade 5 TCC

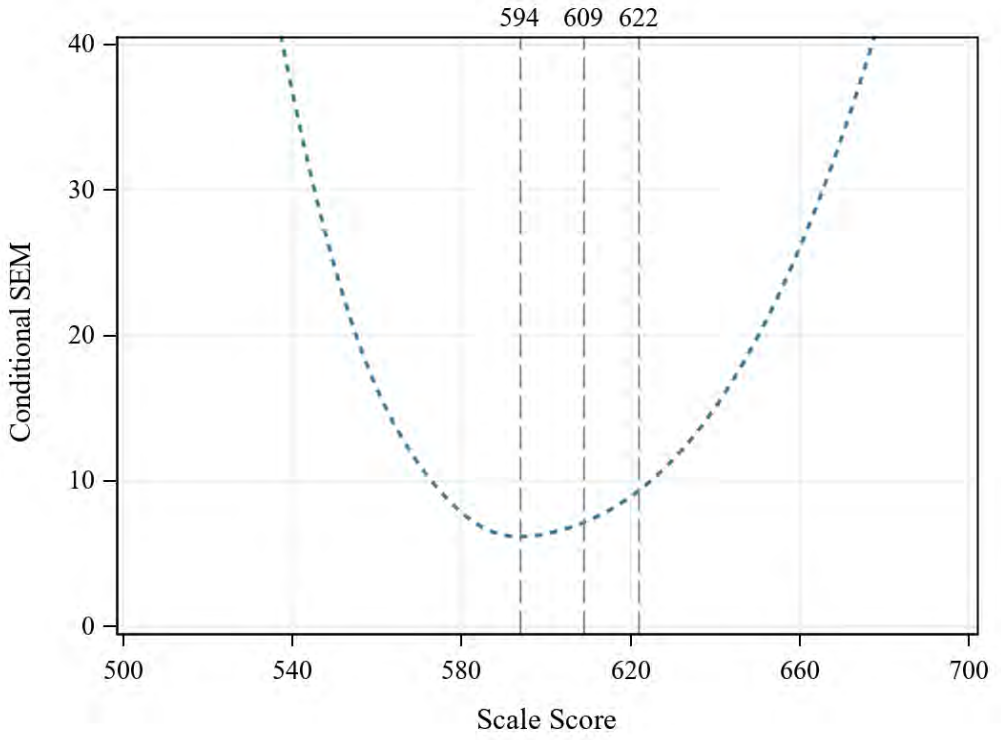

Figure 6.7. ELA Grade 5 CSEM Curve

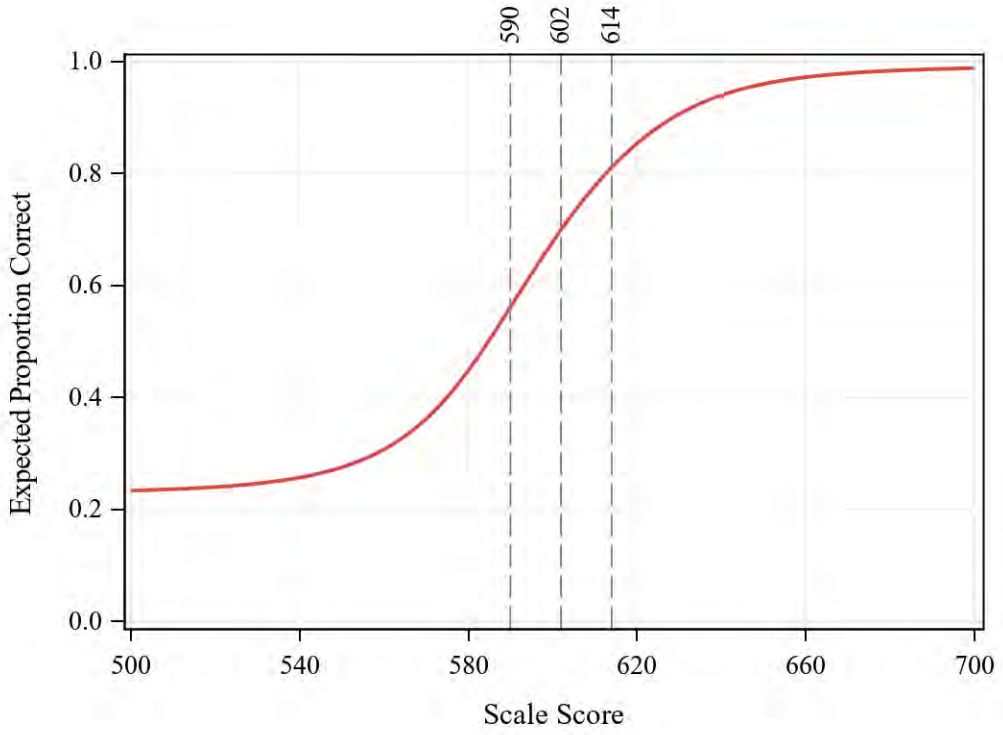

Figure 6.8. ELA Grade 6 TCC

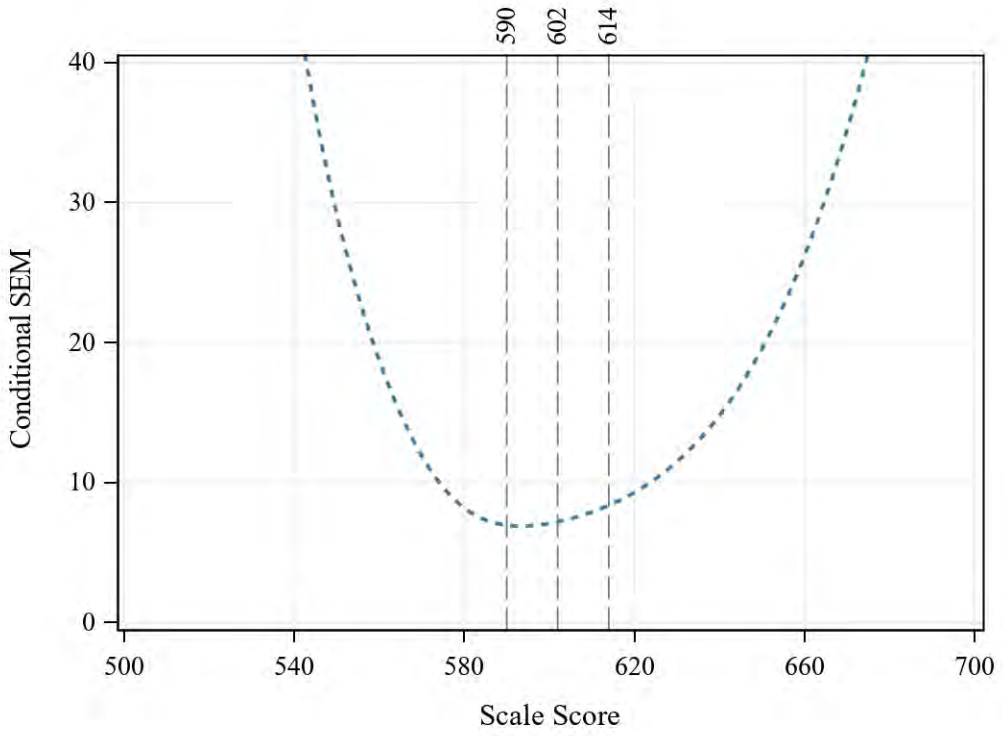

Figure 6.9. ELA Grade 6 CSEM Curve

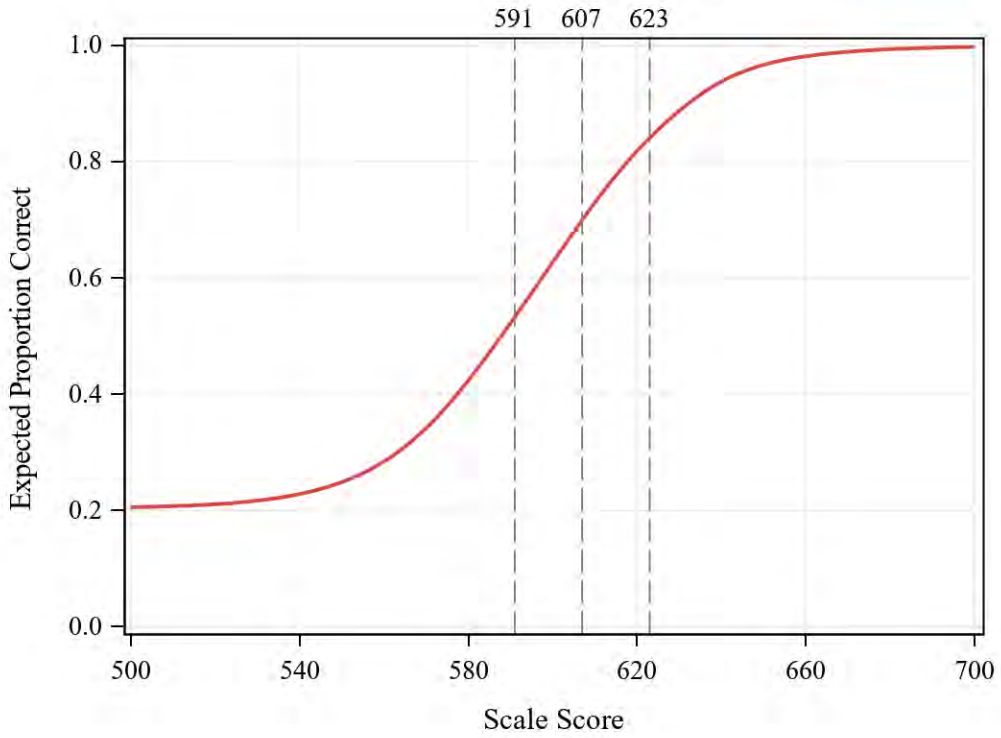

Figure 6.10. ELA Grade 7 TCC

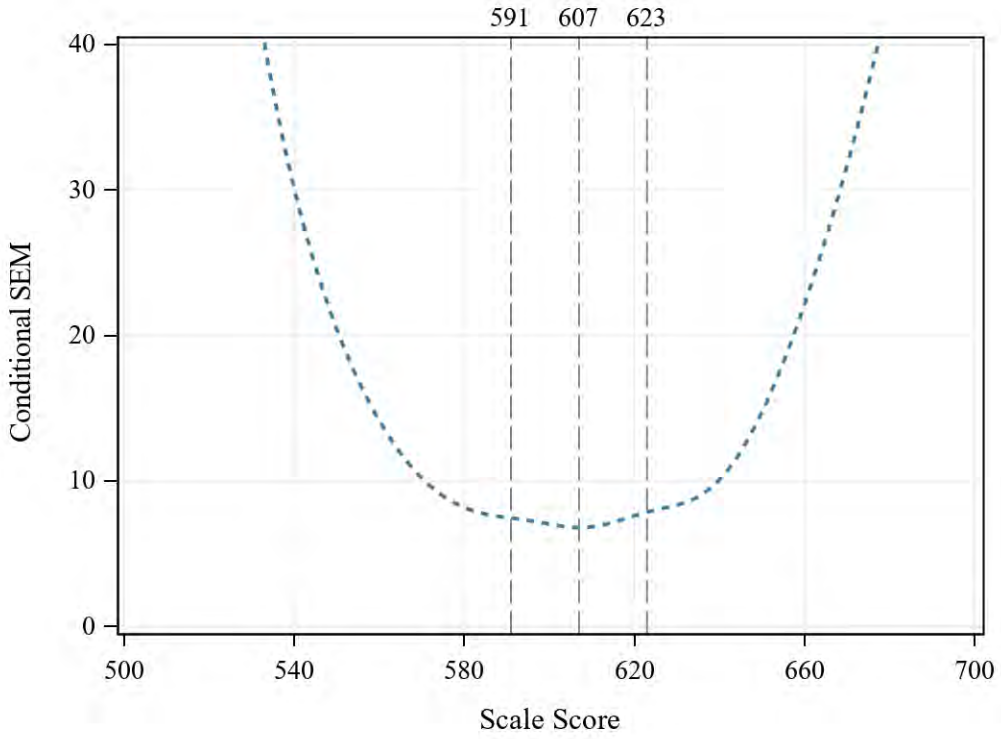

Figure 6.11. ELA Grade 7 CSEM Curve

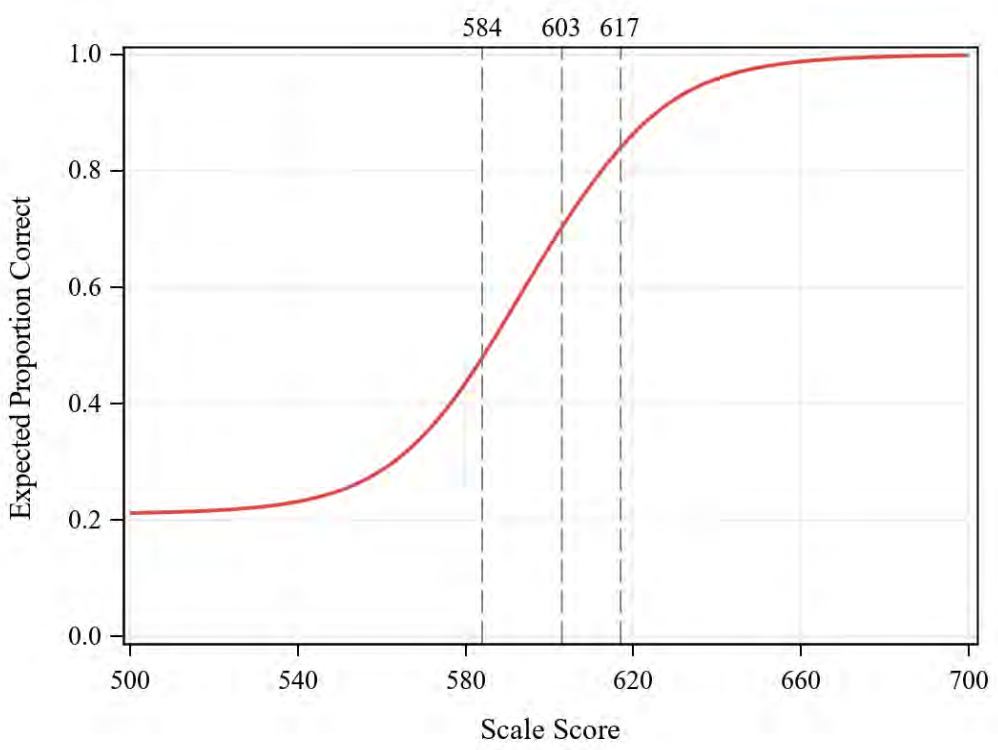

Figure 6.12. ELA Grade 8 TCC

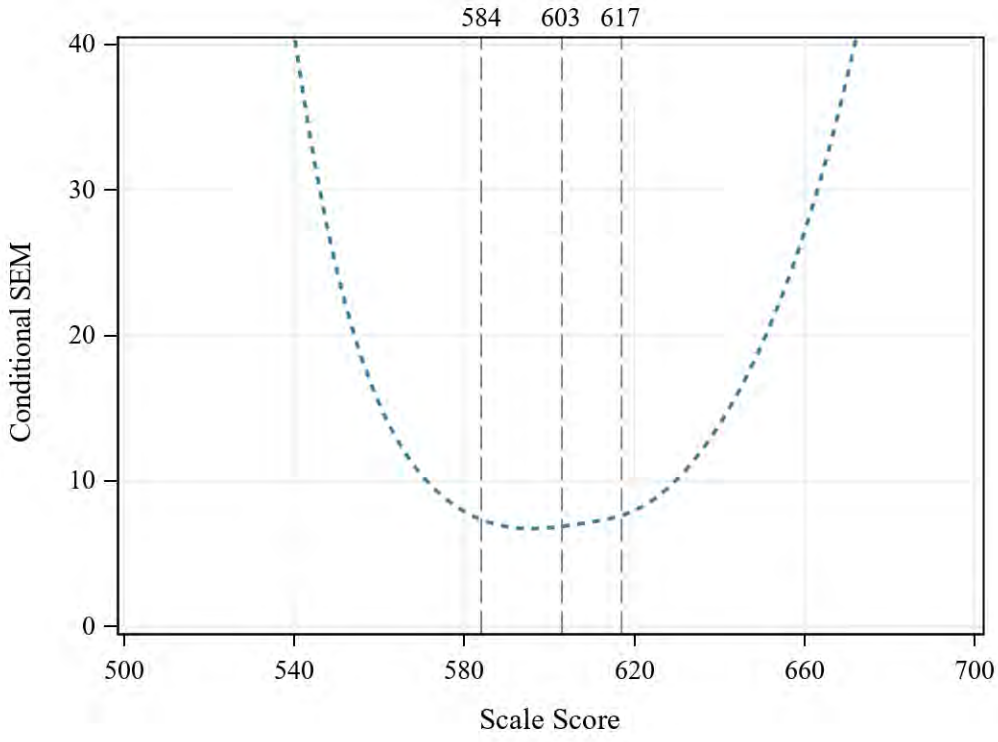

Figure 6.13. ELA Grade 8 CSEM Curve

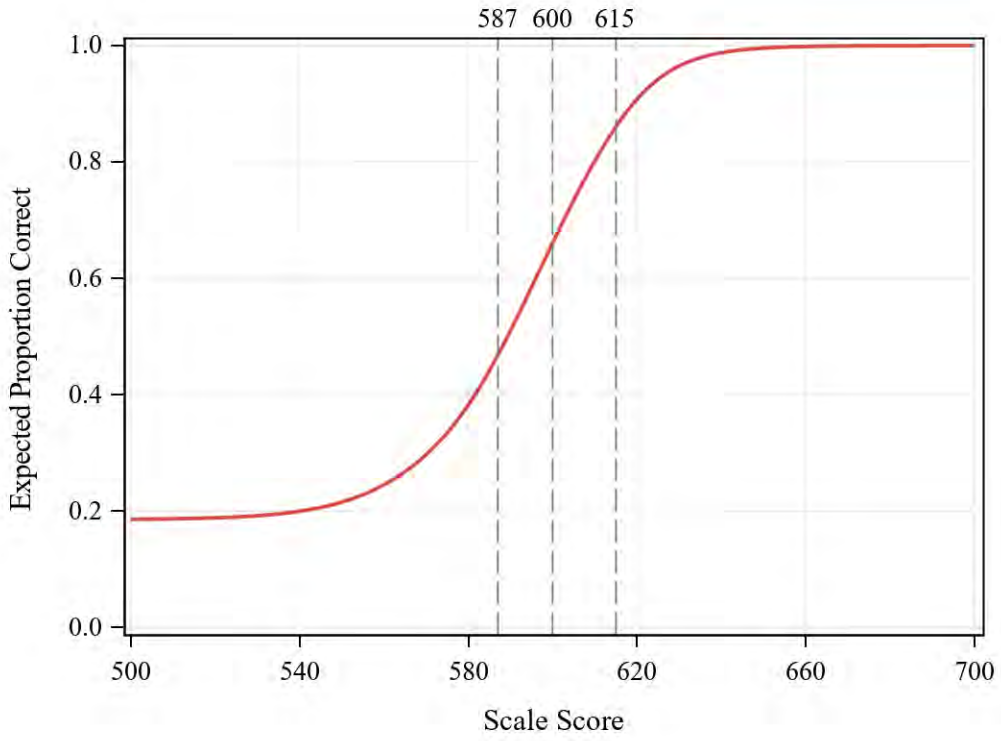

Figure 6.14. Mathematics Grade 3 TCC

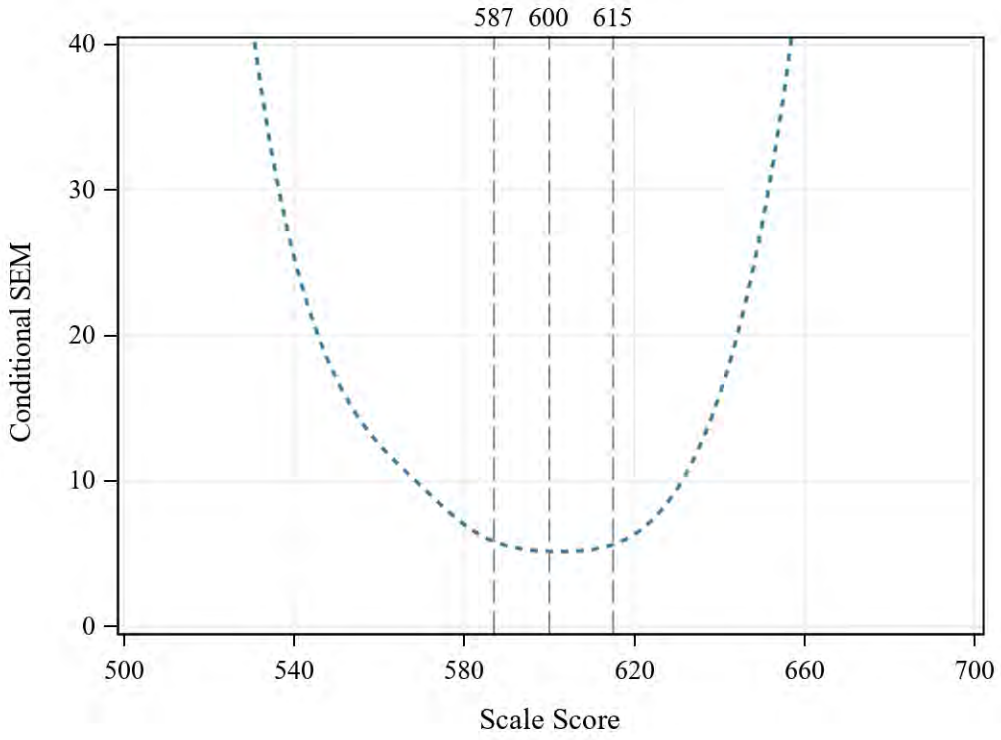

Figure 6.15. Mathematics Grade 3 CSEM Curve

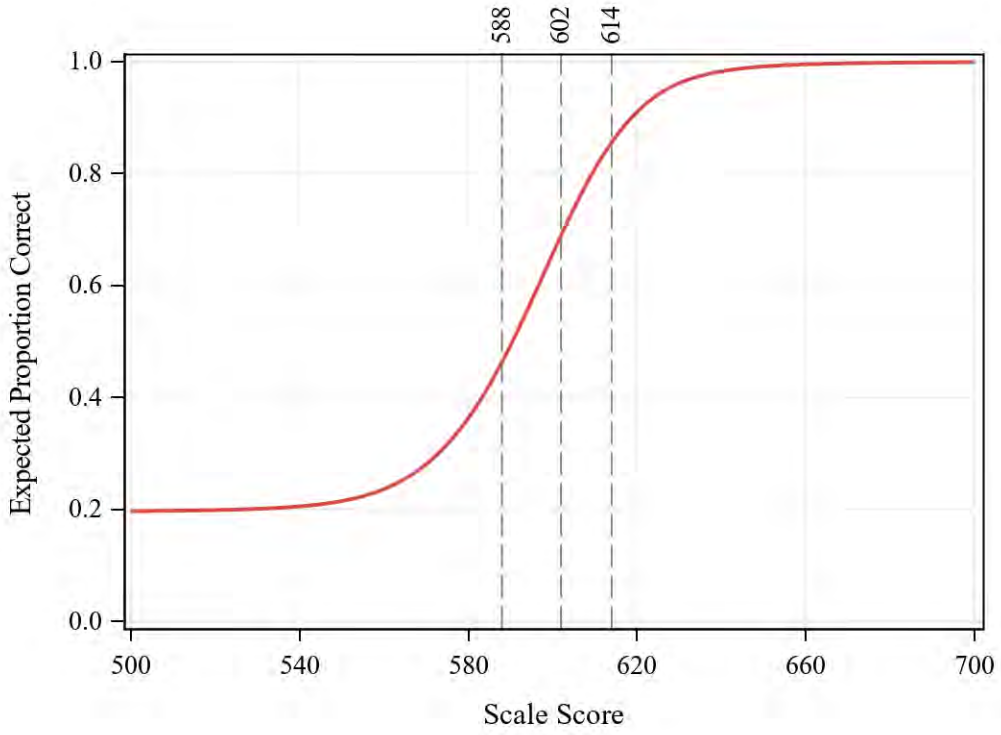

Figure 6.16. Mathematics Grade 4 TCC

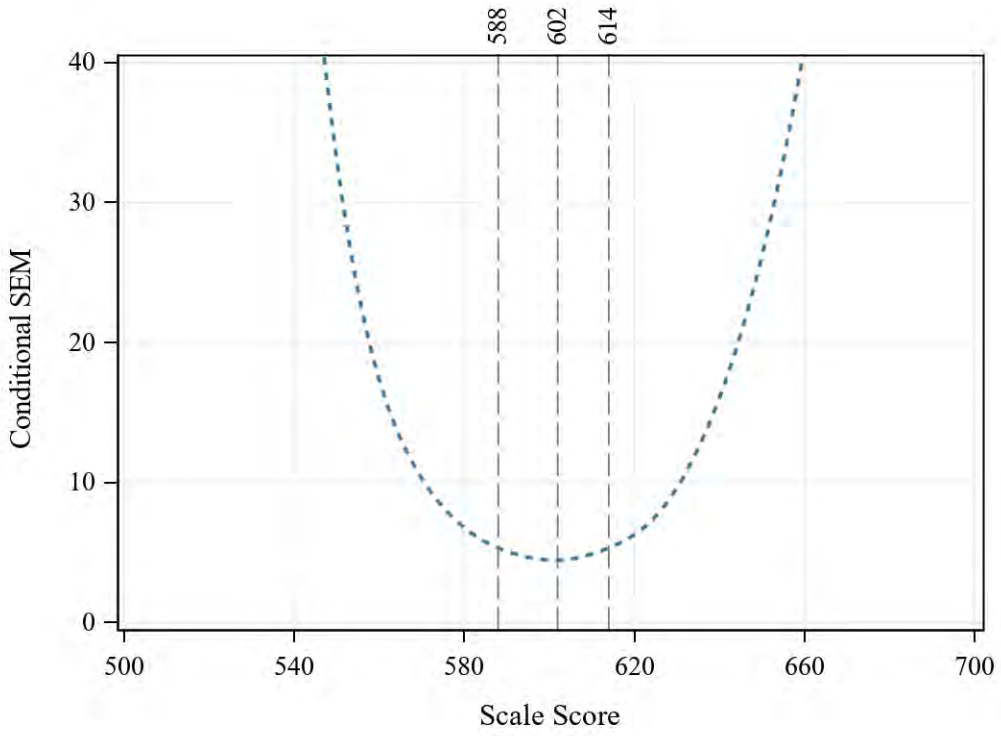

Figure 6.17. Mathematics Grade 4 CSEM Curve

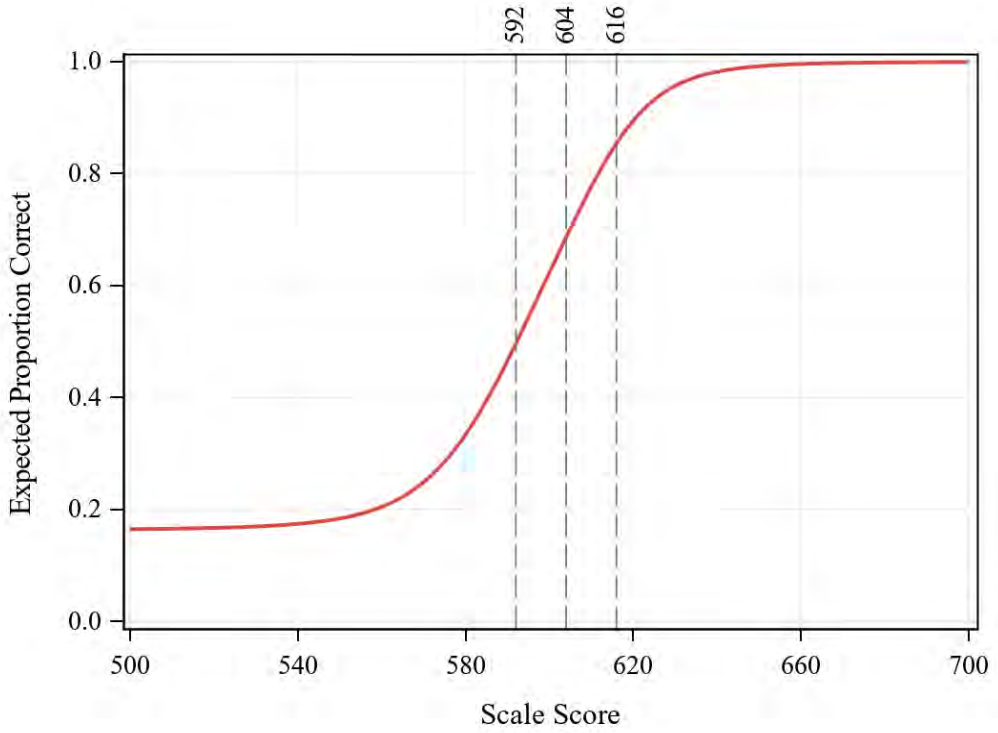

Figure 6.18. Mathematics Grade 5 TCC

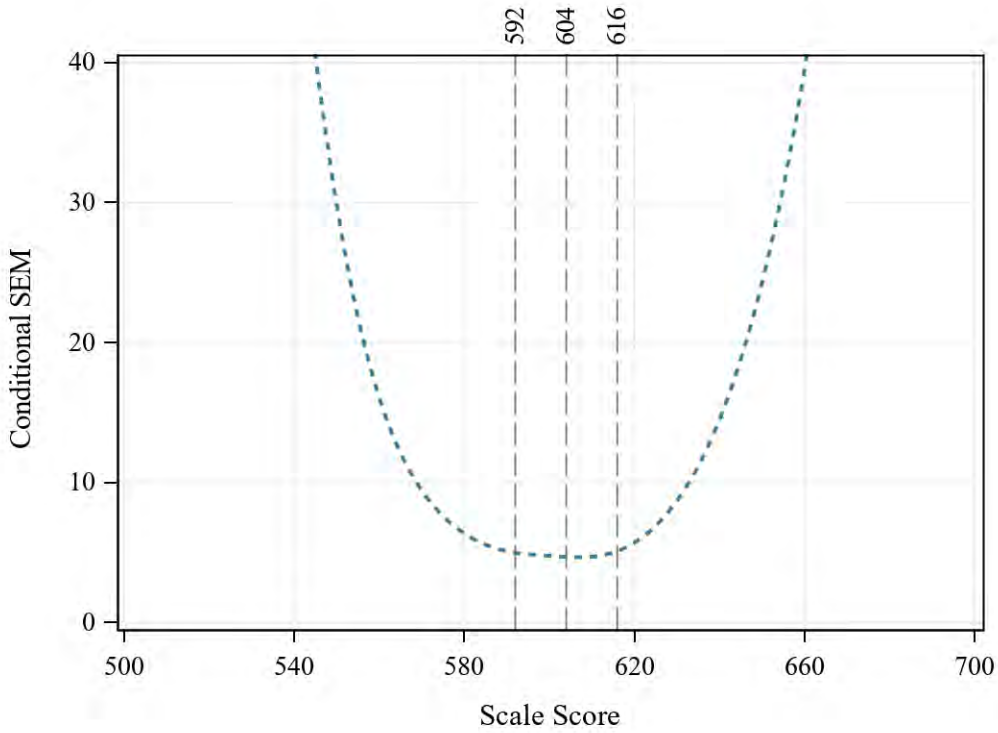

Figure 6.19. Mathematics Grade 5 CSEM Curve

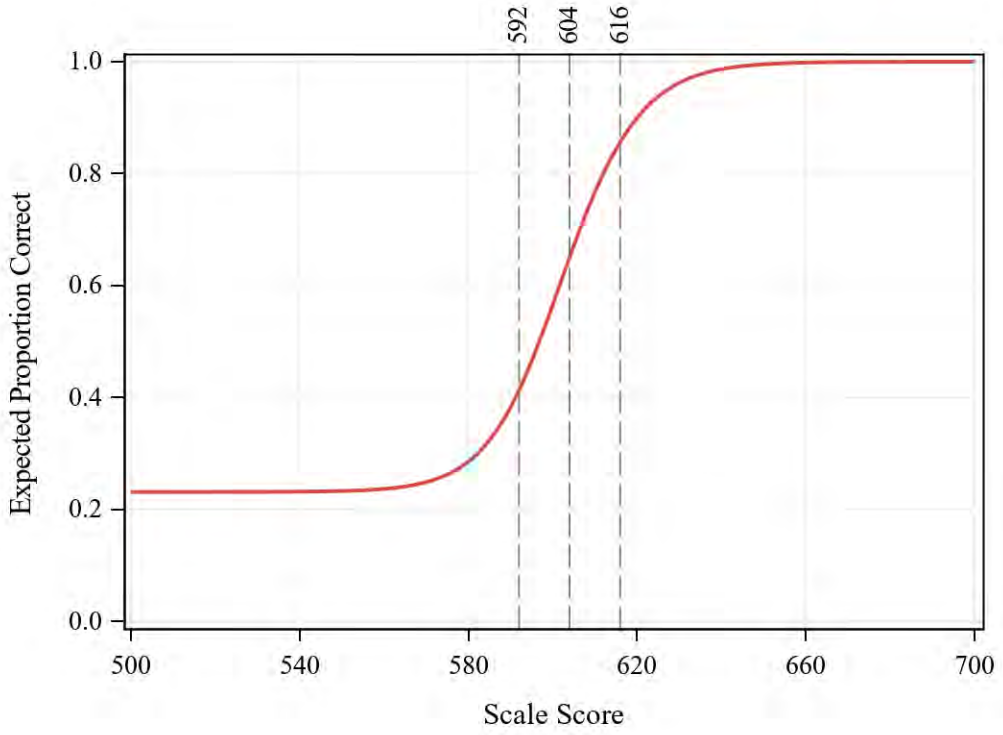

Figure 6.20. Mathematics Grade 6 TCC

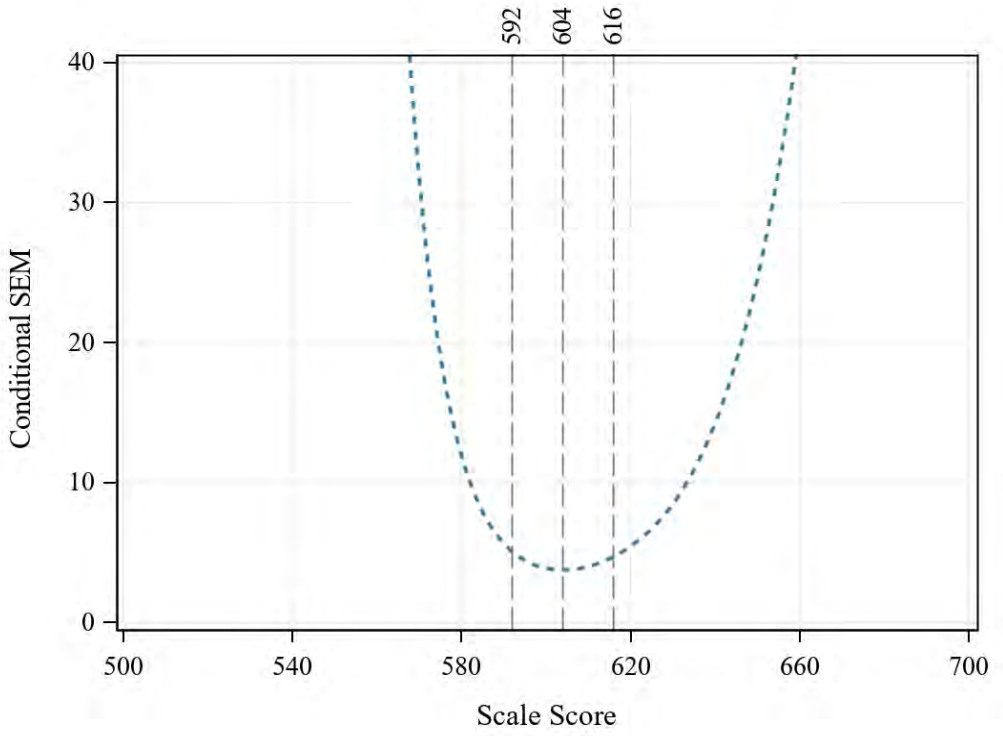

Figure 6.21. Mathematics Grade 6 CSEM Curve

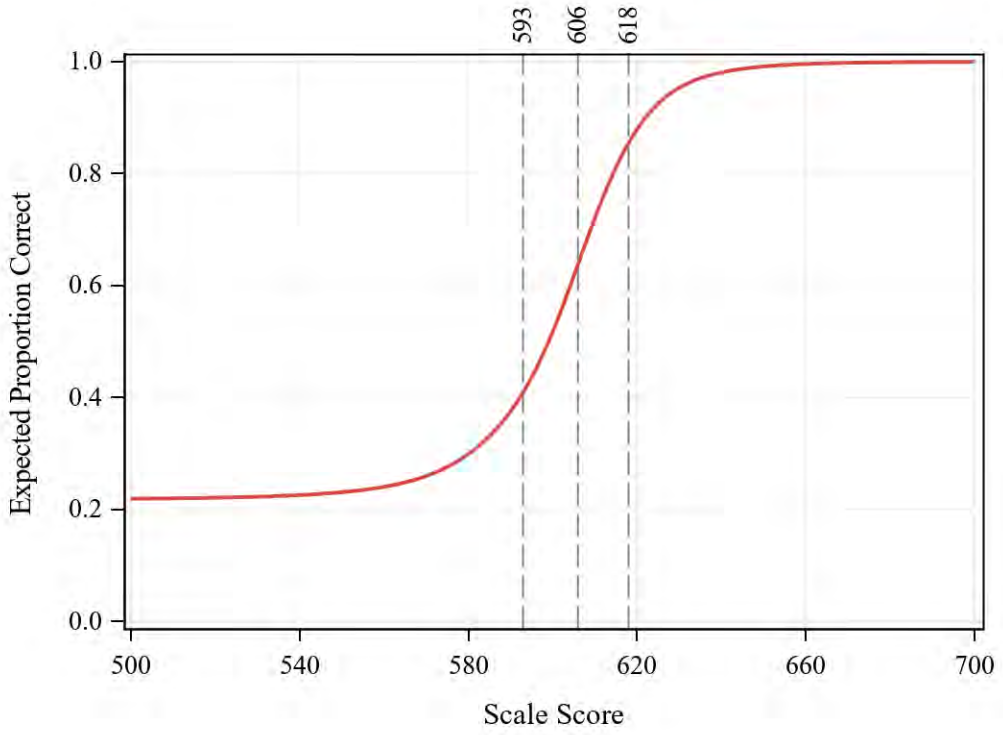

Figure 6.22. Mathematics Grade 7 TCC

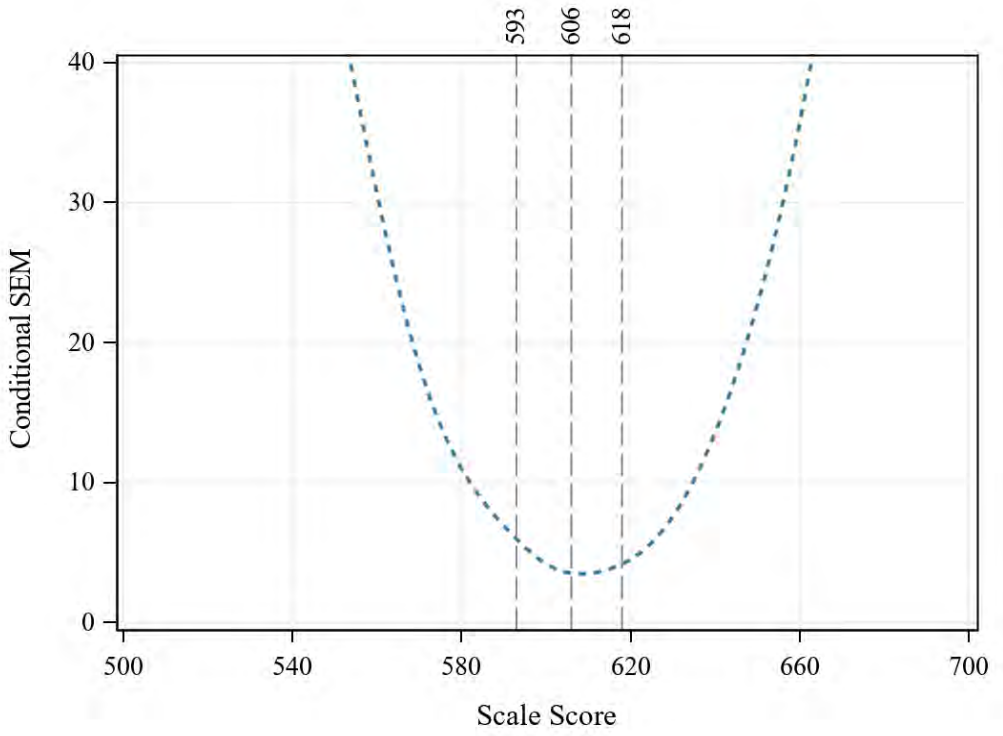

Figure 6.23. Mathematics Grade 7 CSEM Curve

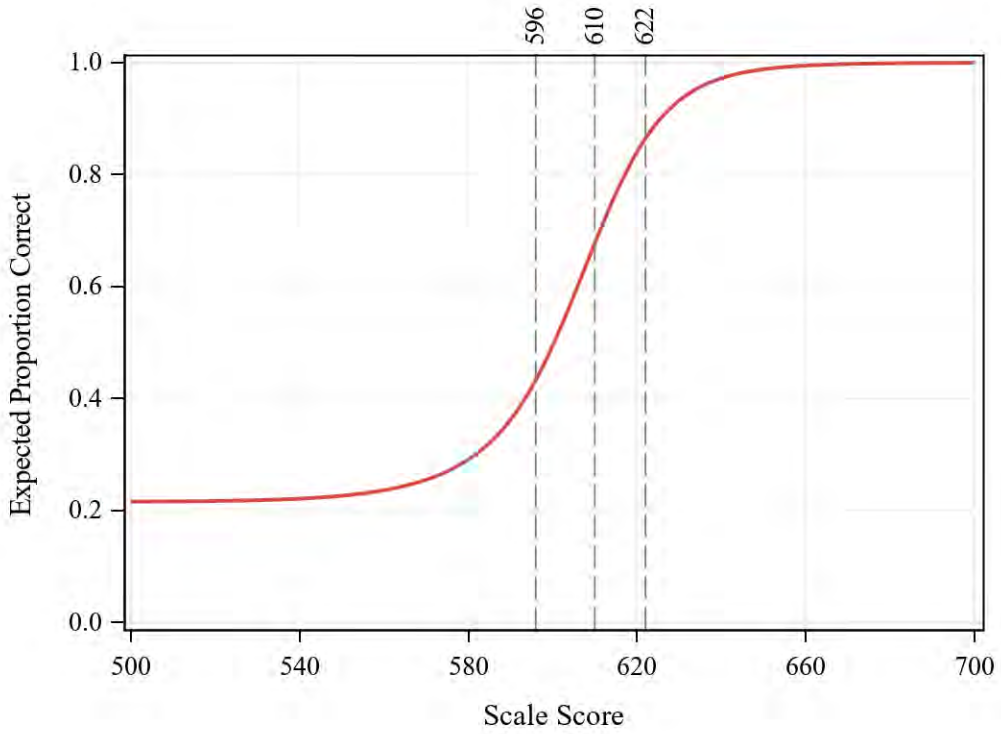

Figure 6.24. Mathematics Grade 8 TCC

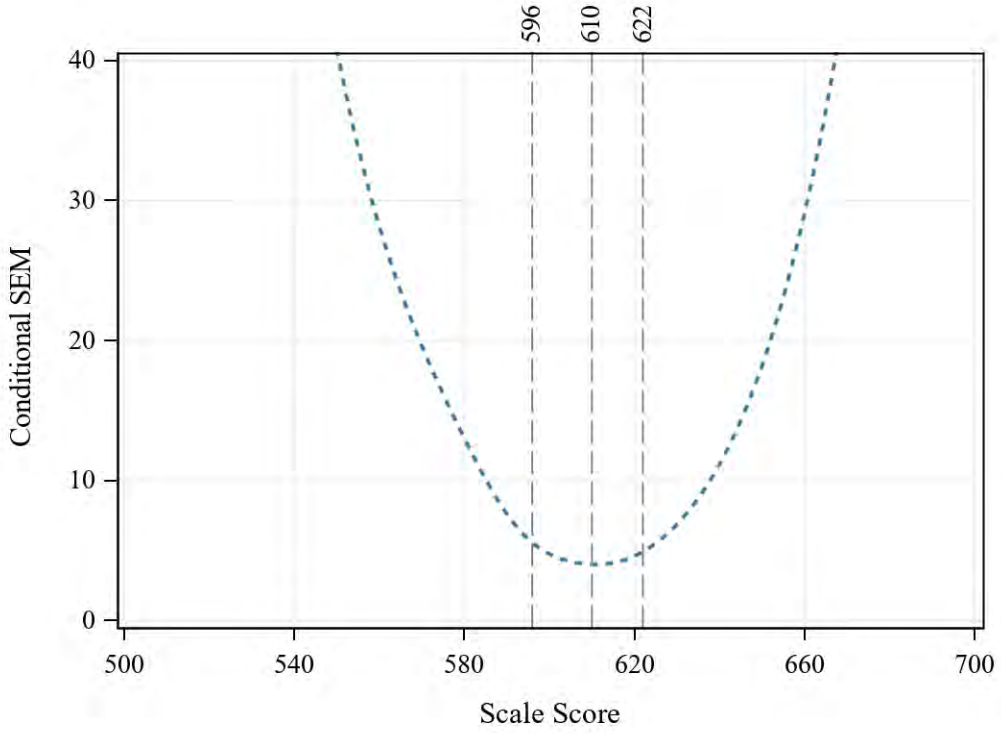

Figure 6.25. Mathematics Grade 8 CSEM Curve

6.7. Scoring Procedure

New York State student examinations were scored using the number correct (NC) scoring method. This method considers how many score points a student obtained on a test in determining his or her scale score. That is, two students with the same number of score points on the test will receive the same scale score, regardless of which items they answered correctly. In this method, the number correct (or raw) score on the test is converted to a scale score by means of a conversion table. This traditional scoring method is often preferred for its conceptual simplicity and familiarity.

The final item parameters were used to calculate the raw-score-to-theta tables, using a TCC method (see the details provided below). The obtained scaling transformation intercept and slope (M_1^S and M_2^S) were then applied to the theta values to produce raw score-to-scale score-conversion tables for the Grades 3–8 ELA and Mathematics Tests.

An inverse TCC method was employed using POLYEQUATE (Kolen & Cui, 2004). The inverse of the TCC procedure produces trait values (i.e., proficiency) based on unweighted raw scores. These estimates show negligible statistical bias (defined in statistics as the difference between an estimator's expected value and the true value of the parameter being estimated) for tests with maximum possible raw scores of at least 30 points. All NYSTP ELA and Mathematics tests have a maximum raw score higher than 30 points. In the inverse TCC method, a student's trait (i.e., proficiency) estimate is taken to be the trait value that has an expected raw score equal to the student's observed raw score. It was found that, for tests containing only MC items, the inverse of the TCC is an excellent first-order approximation of the number of correct maximum likelihood estimates (MLE) showing negligible bias for tests of at least 30 items. For tests with a mixture of MC and CR items, the MLE and TCC estimates are even more similar (Yen, 1984).

The inverse of the TCC method relies on the following equation:

$$\sum_{i=1}^{n} v_i x_i = \sum_{i=1}^{n} v_i E(X_i \left| \tilde{\theta} \right.),$$

where:

x_i is a student's observed raw score on item *i,*
v_i is a non-optimal weight specified in a scoring process ($v_i = 1$ if no weights are specified), and
$\tilde{\theta}$ is a trait estimate.

Potential differences in test form difficulty at different performance levels are accounted for in the resulting raw score-to-scale score conversion tables, so that students of the same proficiency are expected to obtain the same scale score, regardless of which form they took.

6.7.1. *Raw Score-to-Scale Score and SEM Conversion Tables*

The scale score is the basic score for the NYSTP. Raw score-to-scale score (RSSS) conversion tables based on the total number correct are presented in Appendix Q, Tables Q1–Q12.

The standard error (SE) of a scale score indicates the precision with which the proficiency is estimated, and it inversely is related to the amount of information provided by the test at each performance level. The SE is estimated as follows:

$$SE\left(\hat{\theta}\right) = \frac{1}{\sqrt{I(\theta)}},$$

where

$SE(\hat{\theta})$ is the standard error of the scale score (theta).

$I\hat{\theta}$ is the amount of information provided by the test at a given performance level.

The information is estimated based on thetas in the scale score metric; therefore, the SE is also expressed in the scale score metric. The SE value varies across performance levels and is the highest at the extreme ends of the scale where the amount of test information is typically the lowest. The final element of the raw-score-to-scale-score tables is the application of the performance level cut scores.

New scale score cuts were set in the summer of 2018 through standard review and can be applied to the 2019 scale scores because the 2019 scale scores are on the same scale of the 2018 scale scores after being transformed through the equating and scaling procedures described in section 6.5. See Section 8 and Appendix T of the *2018 Technical Report* for more information on the standards review process.

Table 6.12 and Table 6.13 present scale score ranges associated with each performance level for ELA and Mathematics, respectively.

Table 6.12. ELA Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	532-582	583-601	602-628	629-654
4	528-583	584-602	603-618	619-656
5	513-593	594-608	609-621	622-658
6	502-589	590-601	602-613	614-656
7	510-590	591-606	607-622	623-657
8	507-583	584-602	603-616	617-651

Table 6.13. Mathematics Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	524-586	587-599	600-614	615-645
4	521-587	588-601	602-613	614-648
5	527-591	592-603	604-615	616-649
6	529-591	592-603	604-615	616-649
7	520-592	593-605	606-617	618-643
8	520-595	596-609	610-621	622-653

A mode comparability study was completed to identify whether or not there were any differences in student performance that could be attributed to the mode of test administration (i.e. PBT versus CBT). The main inference to be drawn from the mode comparability study is whether scores that arise from students testing on paper or on computer are interchangeable. A propensity score matching approach was conducted to generate the PBT and CBT samples that were comparable on covariates that may affect student performance, aside from the test mode itself (e.g., gender, school-type, previous performance). The difference in students' test scores were computed between the matched PBT and CBT samples to evaluate test-level mode comparability, and mode adjustments were made accordingly. Please see Appendix R (the mode comparability report) and Appendix S (the NYSED memorandum on the mode comparability results) for more details.

Section 7: Reliability and Standard Error of Measurement

This section presents specific information on various test reliability statistics and standard error of measurement (SEM), as well as the results from a study of performance level classification accuracy and consistency. The data set for these studies includes all tested New York State students who received valid scores.

7.1. Test Reliability

Test reliability is directly related to score stability and standard error and, as such, is an essential element of fairness and validity. Test reliability can be directly measured with an alpha statistic, or the alpha statistic can be used to derive the SEM. For the Grades 3–8 ELA and Mathematics Tests, Questar calculated two types of reliability statistics: Cronbach's alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). These two measures are appropriate for assessment of a test's internal consistency when a single test is administered to a group of examinees on one occasion. The reliability of the test is then estimated by considering how well the items that reflect the same construct yield similar results (or how consistent the results are for different items that reflect the same construct measured by the test). Both Cronbach's alpha and Feldt-Raju coefficient measures are appropriate for tests of multiple-item formats (MC and CR items).

7.1.1. *Test Statistics and Reliability for Total Test*

Table 7.1 and Table 7.3 present the test statistics including raw-score (RS) means and raw-score standard deviations (SDs) for ELA and Mathematics, respectively. These statistics give the necessary context for Table 7.2 and Table 7.4, which present the case counts (N-Count), number of test items (# Items), Cronbach's alpha and associated SEM, and Feldt-Raju coefficient and associated SEM obtained for the total ELA and Mathematics tests. Reliability coefficients provide measures of internal consistency that range from zero to one. High reliability indicates that scores are consistent and not unduly influenced by random error. Overall test reliability is a very good indication of each test's internal consistency.

Grades 3–8 ELA reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.87 to 0.91. Grades 3–8 Mathematics reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.92 to 0.95. The reliabilities are similar across grades and slightly higher for the Mathematics tests than for the ELA tests. All reliabilities were at least 0.87 across all grades and both subjects, which is a good indication that the NYSTP Grades 3–8 ELA and Mathematics Tests are acceptably reliable.

Table 7.1. ELA Test Form Statistics

	Item-Level			Student-Level			
	p-value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.60	0.36	0.87	177,503	34	19.56	7.09
4	0.62	0.36	0.91	180,309	34	20.50	6.74
5	0.63	0.34	0.87	175,383	44	27.12	8.95
6	0.65	0.41	0.83	176,182	44	28.51	8.59
7	0.64	0.23	0.87	159,912	46	29.76	9.18

	Item-Level			Student-Level			
	p-value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
8	0.68	0.38	0.88	151,966	46	31.43	8.98

Table 7.2. ELA Test Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	177,503	25	34	0.88	2.43	0.89	2.32
4	180,309	25	34	0.87	2.47	0.88	2.35
5	175,383	35	44	0.90	2.84	0.91	2.75
6	176,182	35	44	0.89	2.81	0.90	2.72
7	159,912	36	46	0.90	2.85	0.91	2.71
8	151,966	36	46	0.90	2.78	0.91	2.67

Table 7.3. Mathematics Test Form Statistics

	Item-Level			Student-Level			
	p-value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.63	0.32	0.93	181,402	42	25.64	9.97
4	0.63	0.4	0.92	182,333	46	28.33	11.06
5	0.61	0.35	0.86	170,958	46	27.43	11.49
6	0.57	0.21	0.76	169,427	48	26.12	12.30
7	0.58	0.36	0.84	159,915	50	28.34	13.39
8	0.53	0.26	0.70	107,089	50	25.46	12.53

Table 7.4. Mathematics Test Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	181,402	34	42	0.92	2.90	0.92	2.74
4	182,333	38	46	0.93	2.97	0.93	2.84
5	170,958	38	46	0.93	3.05	0.94	2.90
6	169,427	39	48	0.94	3.04	0.94	2.91
7	159,915	41	50	0.94	3.28	0.95	3.04
8	107,089	41	50	0.93	3.33	0.94	3.18

7.1.2. *Reliability of MC Items*

In addition to overall test reliability, Cronbach's alpha and Feldt-Raju coefficient were computed separately for MC and CR item sets. It is important to recognize that reliability is directly affected by test length; therefore, reliability estimates for tests by item type will always be lower than reliability estimates for the overall test form. Table 7.5 and Table 7.6 present reliabilities for the subsets of MC items.

Table 7.5. ELA MC Item Reliability and Standard Error of Measurement

			Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Est.	SEM	Est.	SEM
3	177,503	18	0.80	1.77	0.80	1.76
4	180,309	18	0.78	1.78	0.78	1.77
5	175,383	28	0.86	2.22	0.86	2.21
6	176,182	28	0.85	2.24	0.85	2.23
7	159,912	28	0.84	2.20	0.84	2.19
8	151,966	28	0.86	2.19	0.86	2.18

Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement

			Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Est.	SEM	Est.	SEM
3	181,402	27	0.89	2.04	0.89	2.03
4	182,333	31	0.90	2.24	0.91	2.22
5	170,958	31	0.91	2.26	0.91	2.26
6	169,427	31	0.92	2.30	0.92	2.29
7	159,915	33	0.91	2.40	0.91	2.39
8	107,089	33	0.90	2.51	0.90	2.50

7.1.3. *Reliability of CR Items*

Reliability coefficients were also computed for the subsets of CR items. The results are presented in Table 7.7 and Table 7.8.

Table 7.7. ELA CR Item Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	177,503	7	16	0.86	1.43	0.86	1.4
4	180,309	7	16	0.83	1.51	0.85	1.44
5	175,383	7	16	0.82	1.58	0.83	1.53
6	176,182	7	16	0.82	1.51	0.84	1.44
7	159,912	8	18	0.88	1.53	0.89	1.45

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
8	151,966	8	18	0.87	1.45	0.89	1.38

Note. Results should be interpreted with caution because the number of items is low.

Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
3	181,402	7	15	0.82	1.87	0.82	1.84
4	182,333	7	15	0.83	1.76	0.83	1.75
5	170,958	7	15	0.83	1.84	0.83	1.82
6	169,427	8	17	0.86	1.81	0.87	1.77
7	159,915	8	17	0.89	1.91	0.90	1.85
8	107,089	8	17	0.86	1.95	0.86	1.94

Note. Results should be interpreted with caution because the number of items is low.

7.1.4. *Test Reliability for Subgroups*

In this section, reliability coefficients that were estimated for the population and subgroups are presented. The reporting subgroups include the following: gender, ethnicity, NRC, ELL/MLL, all SWD, all SUA, SWD/SUA (includes examinees who are classified as having a disability and who use at least one disability-related accommodation), and ELL/MLLs using accommodations specific to their ELL/MLL status (ELL/MLL/SUA). Accommodations available to students include the following: Flexibility in Scheduling/Timing, Flexibility in Setting, Method of Presentation (excluding Braille), Method of Response, Braille and Large-type, and others. Accommodations available to ELL/MLLs are Separate Location, and Bilingual Dictionaries and Glossaries.

As shown in Tables 7.9–7.14 and Tables 7.15–7.20 for ELA and Mathematics, respectively, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Cronbach's alpha reliability coefficients were all at least 0.77. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least 0.78. These indicate a very good test internal consistency (reliability) for analyzed subgroups of examinees.

Table 7.9. ELA Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	177,490	0.88	2.43	0.89	2.32
Gender	Female	87,728	0.88	2.42	0.89	2.32
	Male	89,762	0.89	2.42	0.90	2.32
Ethnicity	Asian	17,599	0.88	2.33	0.89	2.24
	African American	30,270	0.89	2.46	0.90	2.35
	Hispanic	50,889	0.88	2.46	0.89	2.36
	American Indian	1,213	0.88	2.44	0.89	2.34
	Multiracial	5,514	0.89	2.39	0.90	2.28
	Pacific Islander	388	0.88	2.39	0.89	2.30
	White	71,493	0.88	2.39	0.89	2.29
NRC	New York	64,279	0.89	2.42	0.90	2.30
	Big 4 Cities	7,405	0.88	2.48	0.89	2.37
	Urban/Suburban	14,458	0.87	2.45	0.88	2.37
	Rural	9,685	0.87	2.40	0.87	2.33
	Average Needs	42,929	0.86	2.38	0.87	2.31
	Low Needs	19,085	0.85	2.30	0.86	2.23
	Charter School	12,481	0.86	2.39	0.87	2.31
	Religious and Independent	7,168	0.90	2.47	0.91	2.34
SWD	All Codes	26,405	0.86	2.47	0.87	2.37
SUA	All Codes	23,348	0.85	2.46	0.86	2.37
ELL/MLL	ELL/MLL=Y	19,027	0.84	2.52	0.85	2.41
SWD/SUA	SWD & SUA codes	20,718	0.84	2.45	0.85	2.37
ELL/MLL/SUA	SUA & ELL/MLL codes	3,640	0.80	2.46	0.81	2.38

Table 7.10. ELA Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	180,298	0.87	2.47	0.88	2.35
Gender	Female	88,939	0.86	2.46	0.87	2.34
	Male	91,359	0.87	2.47	0.88	2.36
Ethnicity	Asian	18,345	0.86	2.35	0.87	2.24
	African American	31,760	0.86	2.51	0.88	2.40
	Hispanic	51,525	0.86	2.50	0.87	2.39
	American Indian	1,210	0.86	2.48	0.87	2.37
	Multiracial	5,291	0.87	2.44	0.88	2.32
	Pacific Islander	435	0.86	2.48	0.87	2.36

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Ethnicity	White	71,577	0.86	2.43	0.87	2.32
NRC	New York	66,417	0.88	2.45	0.89	2.33
	Big 4 Cities	7,755	0.87	2.53	0.88	2.42
	Urban/Suburban	14,124	0.85	2.50	0.86	2.42
	Rural	9,782	0.85	2.46	0.86	2.39
	Average Needs	41,818	0.84	2.44	0.86	2.35
	Low Needs	18,750	0.82	2.35	0.83	2.26
	Charter School	11,550	0.83	2.44	0.85	2.35
	Religious and Independent	10,102	0.88	2.48	0.89	2.35
SWD	All Codes	28,154	0.85	2.49	0.86	2.41
SUA	All Codes	26,821	0.84	2.49	0.85	2.41
ELL/MLL	ELL/MLL=Y	17,474	0.81	2.52	0.82	2.43
SWD/SUA	SWD & SUA codes	23,563	0.83	2.48	0.84	2.41
ELL/MLL/SUA	SUA & ELL/MLL codes	4,008	0.77	2.47	0.78	2.41

Table 7.11. ELA Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	175,361	0.90	2.84	0.91	2.75
Gender	Female	86,137	0.89	2.82	0.90	2.73
	Male	89,224	0.90	2.85	0.91	2.77
Ethnicity	Asian	18,482	0.89	2.68	0.90	2.58
	African American	31,357	0.90	2.91	0.90	2.83
	Hispanic	49,597	0.89	2.90	0.90	2.82
	American Indian	1,175	0.89	2.90	0.89	2.81
	Multiracial	4,768	0.91	2.81	0.91	2.71
	Pacific Islander	456	0.89	2.79	0.90	2.71
	White	69,337	0.90	2.79	0.90	2.70
NRC	New York	65,302	0.90	2.85	0.91	2.74
	Big 4 Cities	7,597	0.90	2.94	0.91	2.85
	Urban/Suburban	13,672	0.89	2.90	0.90	2.83
	Rural	9,640	0.88	2.87	0.89	2.81
	Average Needs	40,861	0.89	2.81	0.89	2.74
	Low Needs	18,483	0.87	2.65	0.87	2.58
	Charter School	12,070	0.88	2.80	0.88	2.74
	Religious and Independent	7,736	0.92	2.86	0.92	2.74
SWD	All Codes	28,245	0.87	2.94	0.88	2.88

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SUA	All Codes	27,268	0.87	2.94	0.88	2.87
ELL/MLL	ELL/MLL=Y	15,106	0.84	2.97	0.85	2.90
SWD/SUA	SWD & SUA codes	23,855	0.87	2.94	0.87	2.88
ELL/MLL/SUA	SUA & ELL/MLL codes	3,658	0.81	2.92	0.81	2.87

Table 7.12. ELA Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	176,156	0.89	2.81	0.90	2.72
Gender	Female	86,668	0.88	2.77	0.89	2.68
	Male	89,488	0.90	2.84	0.90	2.75
Ethnicity	Asian	18,877	0.88	2.56	0.89	2.47
	African American	31,891	0.88	2.89	0.89	2.81
	Hispanic	49,763	0.88	2.88	0.89	2.80
	American Indian	1,255	0.88	2.86	0.89	2.79
	Multiracial	4,398	0.90	2.77	0.91	2.66
	Pacific Islander	527	0.88	2.70	0.88	2.62
	White	69,292	0.89	2.77	0.90	2.67
NRC	New York	65,700	0.90	2.78	0.90	2.68
	Big 4 Cities	7,424	0.89	2.93	0.90	2.84
	Urban/Suburban	12,659	0.88	2.92	0.89	2.84
	Rural	9,280	0.88	2.89	0.89	2.81
	Average Needs	38,542	0.88	2.81	0.89	2.73
	Low Needs	18,066	0.86	2.62	0.86	2.55
	Charter School	12,351	0.86	2.79	0.86	2.73
	Religious and Independent	12,134	0.91	2.85	0.92	2.72
SWD	All Codes	27,795	0.86	2.97	0.87	2.90
SUA	All Codes	26,393	0.86	2.97	0.87	2.90
ELL/MLL	ELL/MLL=Y	14,722	0.83	3.01	0.84	2.92
SWD/SUA	SWD & SUA codes	22,793	0.85	2.98	0.86	2.91
ELL/MLL/SUA	SUA & ELL/MLL codes	3,483	0.77	2.97	0.78	2.91

Table 7.13. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	159,903	0.90	2.85	0.91	2.71
Gender	Female	77,866	0.89	2.76	0.90	2.65

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Gender	Male	82,037	0.91	2.91	0.92	2.75
Ethnicity	Asian	18,013	0.91	2.54	0.91	2.45
	African American	27,457	0.89	2.97	0.90	2.81
	Hispanic	44,084	0.89	2.92	0.90	2.78
	American Indian	1,113	0.90	2.90	0.91	2.75
	Multiracial	3,636	0.91	2.80	0.92	2.65
	Pacific Islander	579	0.90	2.70	0.91	2.58
	White	64,888	0.90	2.81	0.91	2.66
NRC	New York	64,463	0.91	2.80	0.91	2.66
	Big 4 Cities	6,718	0.90	3.04	0.91	2.87
	Urban/Suburban	11,899	0.89	2.99	0.90	2.85
	Rural	9,025	0.88	2.95	0.89	2.81
	Average Needs	35,311	0.89	2.86	0.90	2.73
	Low Needs	17,858	0.88	2.62	0.89	2.53
	Charter School	4,326	0.88	2.80	0.89	2.72
	Religious and Independent	10,303	0.92	2.89	0.93	2.68
SWD	All Codes	25,162	0.87	3.04	0.88	2.91
SUA	All Codes	24,359	0.87	3.04	0.88	2.91
ELL/MLL	ELL/MLL=Y	12,070	0.83	3.06	0.85	2.92
SWD/SUA	SWD & SUA codes	20,779	0.86	3.04	0.87	2.91
ELL/MLL/SUA	SUA & ELL/MLL codes	2,865	0.78	2.99	0.80	2.90

Table 7.14. ELA Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	151,954	0.90	2.78	0.91	2.67
Gender	Female	73,604	0.89	2.69	0.90	2.61
	Male	78,350	0.91	2.83	0.91	2.72
Ethnicity	Asian	17,154	0.91	2.48	0.91	2.41
	African American	29,483	0.90	2.87	0.90	2.77
	Hispanic	42,954	0.90	2.85	0.90	2.75
	American Indian	1,215	0.90	2.86	0.90	2.75
	Multiracial	2,884	0.91	2.74	0.92	2.62
	Pacific Islander	447	0.90	2.71	0.91	2.61
	White	57,722	0.90	2.73	0.91	2.62
NRC	New York	63,247	0.90	2.73	0.91	2.63
	Big 4 Cities	6,282	0.91	3.00	0.92	2.85

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	Urban/Suburban	10,458	0.90	2.93	0.91	2.81
	Rural	8,213	0.90	2.86	0.90	2.75
	Average Needs	31,285	0.90	2.80	0.91	2.69
	Low Needs	15,688	0.89	2.55	0.89	2.46
	Charter School	9,987	0.86	2.68	0.86	2.64
	Religious and Independent	6,794	0.91	2.75	0.91	2.64
SWD	All Codes	24,634	0.87	3.02	0.88	2.91
SUA	All Codes	23,559	0.88	3.00	0.89	2.90
ELL/MLL	ELL/MLL=Y	10,855	0.84	3.07	0.85	2.93
SWD/SUA	SWD & SUA codes	20,217	0.87	3.02	0.88	2.91
ELL/MLL/SUA	SUA & ELL/MLL codes	2,458	0.80	3.02	0.81	2.92

Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	181,387	0.92	2.90	0.92	2.74
Gender	Female	89,440	0.91	2.90	0.92	2.75
	Male	91,947	0.92	2.89	0.93	2.73
Ethnicity	Asian	17,968	0.91	2.61	0.92	2.48
	African American	30,032	0.92	2.95	0.92	2.79
	Hispanic	51,274	0.91	2.96	0.92	2.81
	American Indian	1,205	0.91	2.92	0.92	2.75
	Multiracial	5,477	0.92	2.89	0.93	2.72
	Pacific Islander	402	0.91	2.86	0.92	2.70
	White	74,779	0.91	2.88	0.92	2.73
NRC	New York	64,670	0.92	2.89	0.93	2.73
	Big 4 Cities	7,544	0.91	2.98	0.92	2.80
	Urban/Suburban	14,688	0.91	2.98	0.92	2.82
	Rural	9,846	0.91	2.96	0.92	2.81
	Average Needs	43,230	0.90	2.91	0.91	2.77
	Low Needs	19,302	0.89	2.71	0.90	2.60
	Charter School	12,036	0.91	2.68	0.92	2.55
	Religious and Independent	10,071	0.91	3.00	0.92	2.82
SWD	All Codes	25,902	0.91	2.98	0.92	2.81
SUA	All Codes	21,797	0.90	2.98	0.91	2.83
ELL/MLL	ELL/MLL=Y	21,843	0.90	2.99	0.91	2.84
SWD/SUA	SWD & SUA codes	19,303	0.89	2.98	0.90	2.82

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL/MLL/SUA	SUA & ELL/MLL codes	3,568	0.89	2.95	0.90	2.81

Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	182,321	0.93	2.97	0.93	2.84
Gender	Female	89,931	0.92	2.98	0.93	2.86
	Male	92,390	0.93	2.94	0.94	2.82
Ethnicity	Asian	18,720	0.92	2.61	0.93	2.49
	African American	31,426	0.92	3.05	0.93	2.92
	Hispanic	51,764	0.92	3.05	0.93	2.93
	American Indian	1,208	0.92	3.01	0.93	2.87
	Multiracial	5,275	0.93	2.93	0.94	2.80
	Pacific Islander	440	0.93	2.94	0.93	2.81
	White	73,277	0.92	2.92	0.93	2.81
NRC	New York	66,452	0.93	2.95	0.94	2.81
	Big 4 Cities	7,844	0.92	3.05	0.93	2.92
	Urban/Suburban	14,218	0.92	3.06	0.92	2.95
	Rural	9,892	0.91	3.03	0.92	2.94
	Average Needs	42,303	0.92	2.97	0.92	2.87
	Low Needs	19,085	0.90	2.75	0.91	2.66
	Charter School	11,210	0.92	2.79	0.93	2.67
	Religious and Independent	11,317	0.92	3.05	0.92	2.94
SWD	All Codes	27,118	0.91	3.07	0.92	2.95
SUA	All Codes	24,502	0.90	3.08	0.91	2.96
ELL/MLL	ELL/MLL=Y	19,764	0.90	3.08	0.91	2.98
SWD/SUA	SWD & SUA codes	21,631	0.90	3.08	0.91	2.96
ELL/MLL/SUA	SUA & ELL/MLL codes	3,848	0.88	3.05	0.89	2.95

Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	170,939	0.93	3.05	0.94	2.90
Gender	Female	83,816	0.93	3.05	0.93	2.91
	Male	87,123	0.93	3.04	0.94	2.89
Ethnicity	Asian	18,883	0.93	2.67	0.93	2.54
	African American	30,271	0.93	3.12	0.93	2.98

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Ethnicity	Hispanic	48,715	0.92	3.13	0.93	2.99
	American Indian	1,172	0.92	3.10	0.93	2.96
	Multiracial	4,500	0.93	3.01	0.94	2.87
	Pacific Islander	463	0.92	2.98	0.93	2.85
	White	66,698	0.92	3.00	0.93	2.88
NRC	New York	65,832	0.93	3.05	0.94	2.89
	Big 4 Cities	7,051	0.92	3.09	0.93	2.94
	Urban/Suburban	12,677	0.92	3.14	0.92	3.01
	Rural	8,917	0.91	3.12	0.92	3.01
	Average Needs	39,165	0.92	3.05	0.92	2.94
	Low Needs	18,298	0.91	2.83	0.91	2.73
	Charter School	11,630	0.93	2.90	0.94	2.75
	Religious and Independent	7,369	0.93	3.13	0.93	3.00
SWD	All Codes	22,673	0.91	3.13	0.92	3.00
SUA	All Codes	20,176	0.91	3.13	0.91	3.01
ELL/MLL	ELL/MLL=Y	16,491	0.90	3.15	0.90	3.03
SWD/SUA	SWD & SUA codes	17,468	0.90	3.13	0.91	3.00
ELL/MLL/SUA	SUA & ELL/MLL codes	3,018	0.88	3.10	0.89	2.99

Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	169,404	0.94	3.04	0.94	2.91
Gender	Female	83,250	0.94	3.06	0.94	2.92
	Male	86,154	0.94	3.03	0.95	2.89
Ethnicity	Asian	19,148	0.94	2.79	0.94	2.64
	African American	30,427	0.93	3.08	0.94	2.95
	Hispanic	48,283	0.93	3.09	0.93	2.97
	American Indian	1,195	0.93	3.09	0.94	2.97
	Multiracial	4,127	0.94	3.00	0.95	2.86
	Pacific Islander	522	0.94	3.01	0.94	2.87
	White	65,495	0.93	3.01	0.94	2.90
NRC	New York	65,101	0.94	3.04	0.95	2.88
	Big 4 Cities	6,834	0.93	3.01	0.93	2.89
	Urban/Suburban	11,813	0.93	3.07	0.93	2.95
	Rural	8,628	0.92	3.08	0.93	2.98
	Average Needs	36,605	0.93	3.05	0.93	2.94

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	Low Needs	17,828	0.92	2.89	0.93	2.79
	Charter School	11,821	0.94	2.99	0.94	2.86
	Religious and Independent	10,774	0.93	3.10	0.93	2.99
SWD	All Codes	22,245	0.91	2.99	0.91	2.89
SUA	All Codes	19,216	0.90	3.00	0.91	2.90
ELL/MLL	ELL/MLL=Y	15,130	0.90	2.98	0.90	2.90
SWD/SUA	SWD & SUA codes	16,570	0.89	2.97	0.90	2.89
ELL/MLL/SUA	SUA & ELL/MLL codes	2,817	0.85	2.87	0.86	2.82

Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	159,904	0.94	3.28	0.95	3.04
Gender	Female	78,600	0.94	3.28	0.95	3.04
	Male	81,304	0.94	3.28	0.95	3.04
Ethnicity	Asian	17,972	0.95	2.85	0.95	2.64
	African American	28,751	0.93	3.34	0.94	3.12
	Hispanic	45,017	0.93	3.36	0.94	3.14
	American Indian	1,093	0.93	3.34	0.94	3.10
	Multiracial	3,458	0.94	3.23	0.95	2.98
	Pacific Islander	589	0.94	3.18	0.95	2.94
	White	62,841	0.93	3.24	0.94	3.02
NRC	New York	63,196	0.95	3.25	0.95	2.98
	Big 4 Cities	6,112	0.92	3.26	0.93	3.06
	Urban/Suburban	10,942	0.92	3.36	0.93	3.16
	Rural	8,204	0.92	3.38	0.93	3.18
	Average Needs	33,002	0.93	3.31	0.94	3.10
	Low Needs	17,170	0.93	3.04	0.93	2.87
	Charter School	10,439	0.94	3.22	0.95	2.99
	Religious and Independent	10,839	0.93	3.33	0.94	3.13
SWD	All Codes	20,960	0.90	3.23	0.91	3.07
SUA	All Codes	18,620	0.91	3.24	0.91	3.08
ELL/MLL	ELL/MLL=Y	13,208	0.90	3.22	0.91	3.08
SWD/SUA	SWD & SUA codes	15,858	0.89	3.20	0.90	3.07
ELL/MLL/SUA	SUA & ELL/MLL codes	2,328	0.83	3.08	0.84	3.00

Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	107,081	0.93	3.33	0.94	3.18
Gender	Female	50,689	0.93	3.33	0.94	3.18
	Male	56,392	0.93	3.32	0.94	3.17
Ethnicity	Asian	10,401	0.94	3.05	0.95	2.89
	African American	22,166	0.92	3.31	0.93	3.17
	Hispanic	32,920	0.92	3.34	0.93	3.19
	American Indian	881	0.93	3.29	0.94	3.13
	Multiracial	1,918	0.93	3.33	0.93	3.19
	Pacific Islander	337	0.94	3.27	0.95	3.09
	White	38,275	0.92	3.35	0.92	3.23
NRC	New York	47,988	0.94	3.29	0.95	3.11
	Big 4 Cities	4,992	0.92	3.15	0.93	3.02
	Urban/Suburban	7,263	0.88	3.31	0.89	3.22
	Rural	6,221	0.90	3.36	0.91	3.26
	Average Needs	20,470	0.90	3.40	0.91	3.29
	Low Needs	8,292	0.91	3.30	0.91	3.20
	Charter School	6,605	0.93	3.22	0.94	3.07
	Religious and Independent	5,250	0.93	3.33	0.93	3.19
SWD	All Codes	17,505	0.89	3.19	0.90	3.10
SUA	All Codes	15,435	0.90	3.21	0.90	3.11
ELL/MLL	ELL/MLL=Y	10,079	0.90	3.19	0.91	3.08
SWD/SUA	SWD & SUA codes	13,165	0.89	3.17	0.89	3.09
ELL/MLL/SUA	SUA & ELL/MLL codes	1,688	0.83	3.04	0.84	2.99

7.2. Standard Error of Measurement (SEM)

Table 7.2 and Table 7.4 present the SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, for ELA and Mathematics, respectively. The SEMs ranged from 2.32 to 3.33 across subjects, grades, and the two methods of estimation, which is reasonable and small. The SEMs are directly related to reliability: the higher the reliability, the lower the standard error. As discussed, the reliability of these tests is relatively high, so it was expected that the SEMs would be very low.

The SEMs for the subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Tables 7.9–7.14 and Tables 7.15–7.20. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations ranged from 2.23 to 3.40, which is acceptably close to those for the entire population. This narrow range indicates that across the Grades 3–8 ELA and Mathematics Tests, all students' test scores are reasonably reliable with minimal error.

7.3. Performance Level Classification Consistency and Accuracy

This subsection describes the analyses conducted to estimate performance level classification consistency and accuracy for the Grades 3–8 ELA and Mathematics Tests. In other words, this provides statistical information on the classification of students into the four performance categories. Classification consistency refers to the estimated degree of agreement between examinees' performance classification from two independent administrations of the same test (or from two parallel forms of the test). Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each administration, a psychometric model was used to obtain the estimated classification consistency indices, using test scores from a single administration. Classification accuracy can be defined as the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston and Lewis, 1995).

In conjunction with measures of internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes tests. As a form of reliability, classification consistency represents how reliably students can be classified into performance categories.

Classification consistency is most relevant for students whose performance is near the proficiency cut score. For example, consider the cut score delineating Levels II and III or simply the "Level III Cut." Students whose proficiency is far above or far below that cut score are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Examinees whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency; these numbers show the number of students who are at risk of being misclassified. Scoring tables with SEMs are located in Section 6: IRT Calibration and Scaling, and student scale score frequency distributions are located in Appendix Q. Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000). Appendix P includes a description of the calculations and procedure based on the paper by Lee et al. (2002).

7.3.1. *Consistency*

The results for classifying students into four performance levels are separated from the results based solely on the Level III cut. Table 7.21 and Table 7.22 include case counts (N-Count), classification consistency (Agreement), classification inconsistency (Inconsistency), and Cohen's kappa (Kappa). Consistency indicates the rate at which a second administration would yield the same performance category designation (or a different designation for the inconsistency rate). The agreement index is a sum of the diagonal element in the contingency table. Kappa is similar but corrects for chance agreement. The inconsistency index is equal to the "1 - agreement index."

Table 7.21 depicts the ELA and Mathematics consistency study results, based on the range of performance levels for all grades. For ELA, 64–72% of students were estimated to be classified consistently to one of the four performance categories with a hypothetical second administration. Kappa—that corrects for chance agreement—ranged from 0.52 to 0.58. These are between

"moderate" and "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 72–77% of students were estimated to be classified consistently to one of the four performance categories, and kappa ranged from 0.63 to 0.69. These are all considered "substantial" agreement, by Landis and Koch's (1977) rules of thumb for the kappa statistic. As mentioned above and for all tests, there is an acceptable amount of measurement error that all scores contain. By random chance, students testing twice may be classified first, for example, as a Level III and second as a Level IV. This is expected to occur more often for students scoring around the selected cut score, and less often for students closer to the middle of the performance level (i.e., close to the mid-point of two adjacent cut scores).

Table 7.21. Decision Consistency (All Cuts)*

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	177,503	72%	28%	0.58
4	180,309	66%	34%	0.53
5	175,383	66%	34%	0.54
6	176,182	64%	36%	0.52
7	159,912	68%	32%	0.57
8	151,966	69%	31%	0.58
Mathematics				
3	181,402	72%	28%	0.63
4	182,333	73%	27%	0.64
5	170,958	73%	27%	0.64
6	169,427	76%	24%	0.68
7	159,915	77%	23%	0.69
8	107,089	76%	24%	0.66

**Note*. Decision consistency was calculated for PBT students only as item parameters were disproportionally based on PBT.

Table 7.22 depicts the ELA and Mathematics consistency study results based on two performance levels (NYS Level II and NYS Level III) as defined by the Level III cut. For ELA, 86–93% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.62 to 0.68. These are considered "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 91–94% of the classifications were estimated consistently, and kappa coefficients ranged from 0.75 to 0.81. As with ELA, these statistics indicate at least "substantial" agreement (where kappa > 0.60) and some indicate "almost perfect" agreement (where kappa > 0.80), as per Landis and Koch's (1977) rules of thumb for kappa.

Table 7.22. Decision Consistency (Level III Cut)*

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	177,503	93%	7%	0.62
4	180,309	89%	11%	0.66
5	175,383	90%	10%	0.63
6	176,182	86%	14%	0.67
7	159,912	91%	9%	0.65
8	151,966	89%	11%	0.68
Mathematics				
3	181,402	91%	9%	0.75
4	182,333	91%	9%	0.77
5	170,958	92%	8%	0.77
6	169,427	93%	7%	0.80
7	159,915	94%	6%	0.81
8	107,089	94%	6%	0.77

**Note.* Decision consistency was calculated for PBT students only as item parameters were disproportionally based on PBT.

7.3.2. *Accuracy*

Table 7.23 presents the results of classification accuracy for ELA and Mathematics across all grades. Included in the table are case counts (N-Count) and classification accuracy (Accuracy) for all performance levels (All Cuts) and for the Level III cut score. By definition, accuracy associated with the Level III cut is at least as great as that with the entire set of cut scores because there are only two categories for the former, as opposed to the latter, which has four.

For ELA, the estimated accuracy rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her underlying proficiency from 73% to 80% of the time across all performance levels and 90% to 95% of the time in regard to the Level III cut score. For mathematics, the estimated accuracy rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her true proficiency from 79% to 84% of the time across all performance levels and 93% to 96% of the time in regard to the Level III cut score.

Table 7.23. Decision Agreement (Accuracy) Estimates*

		Accuracy	
Grade	N-Count	All Cuts	Level III Cut
ELA			
3	177,503	80%	95%
4	180,309	73%	92%
5	175,383	75%	93%
6	176,182	73%	90%

Grade	N-Count	Accuracy	
		All Cuts	Level III Cut
ELA			
7	159,912	75%	93%
8	151,966	77%	93%
Mathematics			
3	181,402	79%	93%
4	182,333	80%	94%
5	170,958	81%	94%
6	169,427	83%	95%
7	159,915	84%	96%
8	107,089	82%	95%

**Note*. Decision agreement was calculated for PBT students only as item parameters were disproportionally based on PBT.

Section 8: Summary of Operational Test Results

This section summarizes the distribution of scale score results on the NYSTP 2019 Grades 3–8 ELA and Mathematics Tests. These include the scale score means, standard deviations, percentile ranks, and performance level distributions for each grade's population and specific subgroups. Gender, ethnic identification, NRC, ELL/MLL, SWD, and SUA variables were used to calculate the results of subgroups required for federal reporting and test equity purposes for both the ELA and Mathematics tests. Additionally, the ELL/MLL/SUA subgroup is defined as ELL/MLL who use one or more ELL/MLL-related accommodation(s). The SWD/SUA subgroup is defined as examinees with disabilities who use one or more disability-related accommodation(s). For the mathematics analyses, the test translation language is also indicated. (Recall that the ELA tests are not translated, as they are a measure of mastery of the English language.) ELA and Mathematics data include examinees with valid scores from all public, non-public, and charter schools. Complete scale score frequency distribution tables for ELA and Mathematics are located in Appendix Q.

8.1. Scale Score Distribution Summary

Scale score distribution summary tables for ELA and Mathematics are presented and discussed. ELA scale score distributions are described first, followed by Mathematics. In the following two subsections, ELA and Mathematics scale score and subscore statistics are presented for all grades, and across selected subgroups in each grade level. Use caution when interpreting the statistics for subgroups with small number counts that are included in the scale score summaries.

8.1.1. *ELA Scale Score and Subscore Distributions*

Table 8.1 shows some key statistics characterizing the distribution of ELA scale scores, while Table 8.2 summarizes the ELA subscores derived from the test in each grade. Tables 8.3–8.8 break down the scale scores by selected subgroups. Some general observations from these tables include: Females outperformed Males; Asian and White students outperformed their peers from other reported ethnic groups; students from Low Needs (as identified by NRC) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter); and ELL/MLL students, SWD, SUA, and SWD/SUA tended to under-perform the State population (All Students). This pattern of achievement was consistent across all grades.

Table 8.1. ELA Scale Score Distribution Summary

		Scale Score		Percentile Ranks				
Grade	N-Count	Mean	SD	10th	25th	50th	75th	90th
3	182,559	599.39	19.46	574	586	602	612	624
4	186,205	598.42	19.56	573	587	599	612	622
5	180,679	599.14	21.44	572	587	601	612	625
6	179,908	598.04	22.49	570	585	600	615	626
7	170,595	599.61	21.05	572	587	600	613	626
8	156,304	599.58	20.14	574	587	601	613	624

Table 8.2. ELA Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Reading	182,559	18	11.01	4.00
	Writing	182,559	16	8.40	3.85
4	Reading	186,205	18	11.24	3.80
	Writing	186,205	16	9.11	3.73
5	Reading	180,679	28	17.74	6.00
	Writing	180,679	16	9.21	3.79
6	Reading	179,908	28	17.97	5.81
	Writing	179,908	16	10.41	3.63
7	Reading	170,595	28	17.34	5.57
	Writing	170,595	18	12.33	4.44
8	Reading	156,304	28	18.29	5.81
	Writing	156,304	18	12.91	4.22

8.1.1.1. ELA Grade 3

Table 8.3 presents the scale score statistics and n-counts of demographic subgroups for Grade 3. The population scale score mean was 599.39 with a standard deviation of 19.46. Female students tended to outperform male students by around four scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (608.79). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about 60% of a standard deviation below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and ELL/MLL subgroups scored, on average, about two-thirds of a standard deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 20 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (602): Female (603), Asian (610), Multiracial (603), Pacific Islander (605), and White (603) students, those attending schools in Low Needs districts (607), and students attending Charter schools (607).

Table 8.3. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	182,559	599.39	19.46	574	586	602	612	624
Gender	Female	90,073	601.52	18.85	577	589	603	615	624
	Male	92,486	597.32	19.83	571	585	598	612	620
Ethnicity	Asian	17,552	608.79	18.82	585	598	610	620	632
	African American	30,433	596.42	19.91	569	584	598	610	620
	Hispanic	50,693	595.70	19.01	572	584	596	610	619

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
Ethnicity	American Indian	1,214	598.75	18.63	575	586	598	612	620
	Multiracial	5,449	600.92	19.65	575	589	603	615	624
	Pacific Islander	400	601.22	19.86	575	591	605	615	624
	White	73,307	601.59	18.45	578	591	603	615	624
NRC	New York	64,815	600.22	20.44	572	586	602	615	624
	Big 4 Cities	7,520	587.77	20.36	560	574	588	603	614
	Urban/Suburban	14,590	592.25	18.53	569	583	594	605	615
	Rural	9,928	593.75	17.70	571	584	594	605	615
	Average Needs	43,452	599.33	17.24	577	589	602	612	620
	Low Needs	19,226	607.47	15.94	586	598	607	617	629
	Charter	12,608	606.67	17.47	584	596	607	617	629
	Religious and Independent	10,220	594.58	21.73	565	583	596	610	620
SWD	All Codes	25,838	584.74	19.13	560	572	585	598	610
SUA	All Codes	13,960	583.22	17.98	560	572	584	596	605
ELL/MLL	ELL/MLL=Y	19,640	583.36	18.14	560	572	584	596	605
SWD/SUA	SWD & SUA codes	9,518	582.02	17.85	557	569	584	594	605
ELL/MLL/SUA	SUA & ELL/MLL codes	1,290	579.05	16.92	555	569	578	589	602

8.1.1.2. ELA Grade 4

Table 8.4 contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 598.42 with a standard deviation of 19.56. Female students tended to outperform male students by around five scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (608.18). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, about five-sixths deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 22 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (599): Female (603), Asian (609), Multiracial (603), Pacific Islander (603), and White (603) students, those from Low Needs districts (607), and those enrolled at Charter schools (607).

Table 8.4. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	186,205	598.42	19.56	573	587	599	612	622
Gender	Female	91,797	601.11	19.17	576	589	603	615	624
	Male	94,408	595.80	19.58	570	584	597	609	619
Ethnicity	Asian	18,277	608.18	18.75	584	597	609	621	632
	African American	31,947	594.63	19.21	570	581	594	607	619
	Hispanic	51,288	594.56	18.89	570	581	594	607	619
	American Indian	1,207	597.15	18.57	573	584	599	609	619
	Multiracial	5,198	601.07	19.53	576	589	603	615	626
	Pacific Islander	439	600.75	18.58	576	589	603	615	622
	White	73,716	601.08	18.87	576	589	603	612	622
NRC	New York	66,915	599.26	20.44	573	587	599	612	622
	Big 4 Cities	7,836	586.30	20.05	560	572	587	601	612
	Urban/Suburban	14,243	591.18	18.31	567	579	592	604	615
	Rural	10,045	593.56	17.71	570	581	594	606	615
	Average Needs	42,281	598.48	17.55	576	587	599	609	621
	Low Needs	18,852	606.93	16.55	587	597	607	619	626
	Charter	11,816	604.25	17.23	581	594	607	615	626
	Religious and Independent	13,893	595.49	21.31	567	583	597	609	621
SWD	All Codes	26,941	583.26	18.35	560	570	584	596	607
SUA	All Codes	15,273	582.34	17.38	560	570	581	594	604
ELL/MLL	ELL/MLL=Y	18,195	579.25	16.95	556	567	581	592	603
SWD/SUA	SWD & SUA codes	9,631	580.79	17.04	560	570	581	592	603
ELL/MLL/SUA	SUA & ELL/MLL codes	1,241	576.31	15.32	556	567	576	587	597

8.1.1.3. ELA Grade 5

Table 8.5 provides the scale score summary statistics by key demographic subgroups for Grade 5 students. The population scale score mean was 599.14 with a standard deviation of 21.44. Female students tended to outperform male students by around six scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across all ethnic groups, Asian students earned the highest mean score (609.77). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, one standard deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 25 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601):

Female (603), Asian (612), Pacific Islander (605), and White (603) students, as well as those from Low Needs districts (610), and Charter schools (603).

Table 8.5. ELA Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	180,679	599.14	21.44	572	587	601	612	625
Gender	Female	88,646	602.31	20.72	576	589	603	615	629
	Male	92,033	596.08	21.68	566	583	598	611	622
Ethnicity	Asian	18,480	609.77	20.36	583	598	612	622	634
	African American	31,535	594.52	20.75	566	581	596	610	620
	Hispanic	49,244	595.11	20.51	569	583	596	610	622
	American Indian	1,195	597.25	21.05	572	585	598	611	622
	Multiracial	4,650	601.47	21.84	574	589	603	615	629
	Pacific Islander	470	602.96	20.82	577	590	605	615	626
	White	70,111	602.26	20.86	576	590	603	615	627
NRC	New York	66,345	600.20	21.89	572	587	601	615	629
	Big 4 Cities	7,686	585.42	22.21	554	571	585	601	614
	Urban/Suburban	13,886	591.14	20.52	563	579	592	605	615
	Rural	9,915	593.41	19.56	569	581	594	605	617
	Average Needs	41,384	600.04	19.57	574	589	601	612	624
	Low Needs	18,607	608.88	17.98	587	598	610	622	629
	Charter	12,322	602.47	19.01	579	590	603	615	625
	Religious and Independent	10,088	593.78	25.07	554	579	598	611	622
SWD	All Codes	26,666	582.28	19.89	554	569	583	596	609
SUA	All Codes	16,183	581.43	19.64	554	569	583	594	605
ELL/MLL	ELL/MLL=Y	15,369	576.54	18.76	550	563	579	590	599
SWD/SUA	SWD & SUA codes	9,992	579.36	19.1	554	566	581	592	603
ELL/MLL/SUA	SUA & ELL/MLL codes	1,240	573.81	16.94	550	563	574	587	594

8.1.1.4. ELA Grade 6

Table 8.6 contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 598.04 with a standard deviation of 22.49. Female students tended to outperform male students by around seven scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (610.50). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, one standard deviation below the mean scale score for

the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 25 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (600): Female (602), Asian (614), Multiracial (603), Pacific Islander (604), and White (604) students, and those enrolled in Low Needs districts (609), and Charter schools (602).

Table 8.6. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	179,908	598.04	22.49	570	585	600	615	626
Gender	Female	88,388	601.21	21.30	575	588	602	615	626
	Male	91,520	594.97	23.17	564	582	597	614	622
Ethnicity	Asian	18,948	610.50	21.09	584	597	614	626	637
	African American	32,036	592.17	21.22	564	579	593	606	618
	Hispanic	49,263	593.27	20.83	567	580	595	606	618
	American Indian	1,222	595.32	20.89	568	583	595	609	622
	Multiracial	4,263	601.52	22.98	571	588	603	618	631
	Pacific Islander	520	604.53	20.50	577	591	604	618	631
	White	68,267	601.93	22.11	575	590	604	616	627
NRC	New York	66,341	599.03	22.87	570	584	600	615	626
	Big 4 Cities	7,536	584.30	22.77	553	570	586	600	615
	Urban/Suburban	12,805	588.82	21.54	561	575	590	604	615
	Rural	9,522	593.56	21.72	565	580	594	609	619
	Average Needs	39,000	598.60	21.18	571	586	600	615	623
	Low Needs	18,213	608.16	19.03	586	597	609	622	631
	Charter	12,567	599.66	18.61	575	588	602	614	622
	Religious and Independent	12,964	596.12	25.16	561	585	600	614	623
SWD	All Codes	26,267	579.76	20.79	552	567	580	593	604
SUA	All Codes	16,188	579.12	20.97	552	567	579	593	604
ELL/MLL	ELL/MLL=Y	14,471	574.24	19.88	547	562	577	588	597
SWD/SUA	SWD & SUA codes	10,021	576.84	20.37	547	564	577	591	602
ELL/MLL/SUA	SUA & ELL/MLL codes	1,260	572.55	17.13	552	562	573	584	593

8.1.1.5. ELA Grade 7

Table 8.7 presents the Grade 7 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 599.61 with a standard deviation of 21.05. Female students tended to outperform male students by around seven scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (611.41). Across NRC subgroups,

students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, about one standard deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 27 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (600): Female (604), Asian (613), Multiracial (603), Pacific Islander (607), and White (604) students as well as those enrolled in New York City (602), Low Needs districts (609), and Charter (602) and Religious and Independent (602) schools.

Table 8.7. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	170,595	599.61	21.05	572	587	600	613	626
Gender	Female	83,291	603.44	19.66	578	591	604	616	627
	Male	87,304	595.96	21.67	568	582	597	611	623
Ethnicity	Asian	18,157	611.41	20.68	584	598	613	626	635
	African American	31,111	594.63	19.52	570	582	596	609	619
	Hispanic	46,376	595.65	19.58	570	584	597	609	619
	American Indian	1,162	597.50	20.42	572	585	598	611	623
	Multiracial	3,632	602.60	21.87	574	589	603	616	630
	Pacific Islander	599	605.99	20.75	579	593	607	619	630
	White	65,141	602.49	20.81	576	591	604	616	626
NRC	New York	65,094	601.33	21.42	574	588	602	616	630
	Big 4 Cities	6,835	585.49	21.42	559	571	587	599	612
	Urban/Suburban	12,098	589.79	19.86	565	576	591	603	613
	Rural	9,177	594.05	18.97	570	582	595	607	617
	Average Needs	35,784	599.14	19.49	574	588	600	612	623
	Low Needs	18,028	608.53	18.34	586	598	609	620	630
	Charter	11,214	602.62	16.99	582	593	602	613	623
	Religious and Independent	11,669	598.02	23.97	562	588	602	613	624
SWD	All Codes	25,094	582.69	18.58	559	572	584	595	604
SUA	All Codes	14,914	581.78	19.01	559	570	582	595	605
ELL/MLL	ELL/MLL=Y	12,322	574.77	17.87	551	565	576	588	597
SWD/SUA	SWD & SUA codes	9,010	579.13	18.37	556	568	580	591	602
ELL/MLL/SUA	SUA & ELL/MLL codes	994	573.05	16.22	551	563	574	584	593

8.1.1.6. ELA Grade 8

Table 8.8 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.58 with a standard deviation of 20.14.

Female students tended to outperform male students by around seven scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (610.31). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, one standard deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 25 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Female (605), Asian (612), Multiracial (602), Pacific Islander (605), and White (604) students, as well as those enrolled in New York City (603) and Low Needs (609) districts, Charter (603), and Religious and Independent (603) schools.

Table 8.8. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	156,304	599.58	20.14	574	587	601	613	624
Gender	Female	75,519	603.44	18.81	580	592	605	615	625
	Male	80,785	595.98	20.66	570	584	597	609	620
Ethnicity	Asian	17,132	610.31	19.62	585	599	612	624	635
	African American	29,663	595.17	18.98	570	584	597	608	617
	Hispanic	42,702	596.11	18.71	572	585	597	609	617
	American Indian	1,212	596.81	19.40	572	585	597	609	620
	Multiracial	2,870	600.88	21.23	574	588	602	614	625
	Pacific Islander	456	602.91	19.56	576	590	605	614	624
	White	58,409	602.26	19.98	577	591	604	615	624
NRC	New York	63,774	601.33	19.75	576	589	603	614	624
	Big 4 Cities	6,425	586.07	21.67	560	571	587	601	613
	Urban/Suburban	10,698	590.18	19.99	566	578	592	604	614
	Rural	8,339	594.77	19.06	570	584	596	608	617
	Average Needs	31,684	598.58	19.54	573	587	600	612	621
	Low Needs	15,813	607.87	17.95	585	597	609	620	629
	Charter	10,127	602.88	15.51	584	594	603	614	620
	Religious and Independent	8,604	598.42	22.61	568	589	603	614	624
SWD	All Codes	23,802	584.00	17.92	563	572	585	595	605
SUA	All Codes	13,644	582.78	18.88	560	571	584	595	606
ELL/MLL	ELL/MLL=Y	10,905	576.25	17.42	553	566	578	589	597
SWD/SUA	SWD & SUA codes	8,542	580.36	18.21	557	570	582	592	603
ELL/MLL/SUA	SUA & ELL/MLL codes	868	574.95	15.27	554	566	576	585	594

8.1.2. *Mathematics Scale Score Distributions*

Table 8.9 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 8.10 summarizes the mathematics subscores derived from the test in each grade. Tables 8.11–8.16 break down the scale scores by selected subgroups. Some general observations from the mathematics data are as follows: Female and Male students performed fairly consistently; Asian students scored considerably higher than other reported ethnic groups; schools belonging to Low Needs districts (as identified by the NRC code) and Charter schools outperformed most other school types (New York City, Big 4 Cities, High Needs Urban/Suburban, and Rural and Average Needs districts). Students taking the Chinese and Korean translations tended to outperform the other translation subgroups (Haitian-Creole, Spanish, and Russian); and ELL/MLLs, SWDs, and/or SUAs achieved below the State mean in most percentile ranks. This pattern of achievement was fairly consistent across all grades.

Table 8.9. Mathematics Scale Score Distribution Summary

		Scale Score		Percentile Ranks				
Grade	N-Count	Mean	SD	10th	25th	50th	75th	90th
3	184,576	599.81	19.69	574	588	601	613	622
4	188,143	600.01	20.38	574	588	601	614	624
5	181,771	599.74	20.80	572	586	601	613	624
6	179,611	600.24	20.17	573	588	602	614	625
7	168,909	600.78	20.68	575	588	602	615	626
8	115,886	599.39	21.67	569	588	602	614	624

Table 8.10. Mathematics Subscore Summary

			Subscore		
Grade	Subscore	N-Count	Max	Mean	SD
3	Operations and Algebraic Thinking	184,576	19	12.66	4.63
	Number and Operations—Fractions	184,576	8	4.36	2.28
	Measurement and Data	184,576	11	6.41	2.90
4	Operations and Algebraic Thinking	188,143	9	6.31	2.27
	Number and Operations in Base 10	188,143	12	6.56	3.19
	Number and Operations—Fractions	188,143	12	8.03	3.36
5	Number and Operations in Base 10	181,771	13	8.52	3.55
	Number and Operations—Fractions	181,771	18	9.62	5.14
	Measurement and Data	181,771	12	6.60	3.26
6	Ratios and Proportional Relationships	179,611	12	6.48	3.21
	The Number System	179,611	8	4.72	2.41
	Expressions and Equations	179,611	22	11.25	5.97
7	Ratios and Proportional Relationships	168,909	14	7.18	3.98

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
7	The Number System	168,909	9	5.57	2.50
	Expressions and Equations	168,909	17	9.82	4.90
8	Expressions and Equations	115,886	18	8.38	4.85
	Functions	115,886	15	7.40	4.07
	Geometry	115,886	12	5.94	3.46

8.1.2.1. Mathematics Grade 3

Table 8.11 presents the Grade 3 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 599.81 with a standard deviation of 19.69. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (611.21). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored about three-quarters of a standard deviation below the mean scale score for the population. SUA students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 20 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Asian (613), Pacific Islander (603), and White (604) students, as well as those enrolled at Low Needs (609) districts and Charter schools (609). In terms of the 50th-percentile ranks for students using translated forms, they ranged from 581 (Haitian-Creole, n = 75) to 609 (Chinese, n = 464).

Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	184,576	599.81	19.69	574	588	601	613	622
Gender	Female	90,611	599.85	18.92	576	588	601	613	622
	Male	93,965	599.77	20.40	573	587	601	613	624
Ethnicity	Asian	18,150	611.21	17.96	588	601	613	622	631
	African American	30,809	594.94	20.16	568	583	596	609	619
	Hispanic	52,488	595.39	18.99	571	583	596	607	619
	American Indian	1,232	598.72	19.56	574	588	600	611	622
	Multiracial	5,534	600.48	20.29	573	589	601	615	626
	Pacific Islander	404	602.93	18.79	579	592	603	616	626
	White	75,689	602.12	18.78	577	591	604	615	624
NRC	New York	66,200	599.91	19.86	574	588	601	613	626
	Big 4 Cities	7,607	587.44	20.39	559	574	588	601	615
	Urban/Suburban	14,796	592.50	19.48	568	581	594	606	617
	Rural	9,926	594.85	18.92	570	583	596	607	617

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
NRC	Average Needs	43,469	600.13	18.19	577	590	601	613	622
	Low Needs	19,377	608.52	16.43	588	598	609	619	628
	Charter	12,618	608.72	18.14	585	597	609	619	631
	Religious and Independent	10,447	595.13	20.44	568	583	597	609	619
SWD	All Codes	28,400	585.57	20.39	557	571	587	600	611
SUA	All Codes	14,896	582.56	19.79	557	570	583	597	607
ELL/MLL	ELL/MLL=Y	22,341	587.13	18.86	563	574	588	600	611
SWD/SUA	SWD & SUA codes	12,667	580.95	19.67	554	568	581	594	606
ELL/MLL/SUA	SUA & ELL/MLL codes	1,797	580.19	18.46	554	568	581	593	603
ELL/MLL Test Language	Chinese	464	606.94	19.05	583	596	609	619	631
	English	180,270	600.09	19.59	574	588	601	613	622
	Haitian-Creole	75	581.60	16.17	563	571	581	591	603
	Korean	47	605.11	19.86	583	592	604	619	626
	Russian	195	600.55	18.34	574	591	601	611	622
	Spanish	3,525	584.51	18.38	557	574	585	597	607
	All Translations	4,306	587.83	19.98	563	574	588	601	613

8.1.2.2. Mathematics Grade 4

Table 8.12 presents the Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 600.01 with a standard deviation of 20.38. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (613.56). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored about three-quarters of a standard deviation below the mean scale score for the population. Students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 21 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Male (602), Asian (614), Multiracial (603), Pacific Islander (603), and White (604) students, and those enrolled in Average (602) and Low (610) Needs districts, and Charter schools (609). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 579 (Haitian-Creole, n = 79) to 610 (Chinese, n = 500 and Korean, n = 38).

Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	188,143	600.01	20.38	574	588	601	614	624
Gender	Female	92,231	600.01	19.61	574	589	600	612	624
Gender	Male	95,912	600.00	21.09	571	588	602	614	624
Ethnicity	Asian	18,930	613.56	18.98	590	602	614	628	643
	African American	32,333	593.51	20.33	568	581	594	607	618
	Hispanic	53,110	595.06	19.53	569	583	596	607	618
	American Indian	1,242	598.38	19.65	572	586	599	610	622
	Multiracial	5,356	601.35	20.68	574	589	603	616	628
	Pacific Islander	451	602.12	19.82	574	590	603	614	628
	White	76,490	602.80	19.07	579	592	604	615	624
NRC	New York	68,186	600.41	21.20	574	588	600	614	628
	Big 4 Cities	7,934	585.17	20.92	558	571	586	600	611
	Urban/Suburban	14,347	591.82	19.51	564	580	593	605	616
	Rural	9,972	595.70	18.38	571	585	598	608	617
	Average Needs	42,546	600.85	18.32	578	590	602	613	622
	Low Needs	19,148	609.74	16.74	590	600	610	621	628
	Charter	11,784	607.65	19.14	584	595	609	621	633
	Religious and Independent	13,988	595.79	19.62	568	584	598	609	618
SWD	All Codes	29,692	584.25	20.15	558	571	584	598	609
SUA	All Codes	17,975	582.97	19.72	554	569	584	597	607
ELL/MLL	ELL/MLL=Y	20,999	584.03	18.68	558	571	584	597	607
SWD/SUA	SWD & SUA codes	15,246	581.33	19.52	553	568	582	595	606
ELL/MLL/SUA	SUA & ELL/MLL codes	2,252	578.72	18.17	553	568	579	591	602
ELL/MLL Test Language	Chinese	500	607.44	18.80	583	597	610	618	628
	English	183,630	600.39	20.21	574	589	602	614	624
	Haitian-Creole	79	580.62	17.61	558	568	579	593	600
	Korean	38	610.63	17.75	590	599	610	624	633
	Russian	156	597.44	20.44	571	586	598	609	624
	Spanish	3,740	580.61	19.02	553	568	581	594	604
	All Translations	4,513	584.42	21.07	558	571	584	598	612

8.1.2.3. Mathematics Grade 5

Table 8.13 presents the Grade 5 demographic subgroup n-counts and scale score statistics. The population scale score mean was 599.74 with a standard deviation of 20.80. Female and male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students'

scale score means exceeded the State mean scale score, as did those of students from New York City, Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (613.84). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, about 0.85 standard deviation below the mean scale score for the population. ELL/MLL students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 24 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Asian (616), Pacific Islander (605), and White (604) students, as well as those enrolled at Average (602) and Low (610) Needs districts, and Charter schools (607). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 580 (Haitian-Creole, n = 105 and Spanish, n = 3,189) to 609 (Chinese, n = 540).

Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	181,771	599.74	20.80	572	586	601	613	624
Gender	Female	88,580	600.07	20.08	575	588	601	613	624
	Male	93,191	599.43	21.47	572	586	601	614	624
Ethnicity	Asian	19,137	613.84	19.48	589	602	616	628	644
	African American	31,995	593.34	20.84	565	580	593	607	619
	Hispanic	51,150	594.82	19.98	569	583	596	607	619
	American Indian	1,217	597.89	19.96	572	586	598	612	621
	Multiracial	4,763	600.18	21.23	572	586	601	616	624
	Pacific Islander	481	604.46	19.21	580	593	605	617	628
	White	72,749	602.37	19.32	577	591	604	616	624
NRC	New York	67,518	600.33	21.38	572	586	601	616	628
	Big 4 Cities	7,768	584.65	21.21	555	569	585	599	612
	Urban/Suburban	13,879	591.07	20.00	565	580	592	605	616
	Rural	9,835	594.80	18.62	571	583	596	607	617
	Average Needs	41,144	600.76	18.60	577	589	602	613	621
	Low Needs	18,661	609.61	17.33	588	600	610	621	628
	Charter	12,290	606.60	20.51	581	593	607	621	634
	Religious and Independent	10,415	594.04	20.29	569	581	595	609	619
SWD	All Codes	29,549	583.28	19.88	555	569	583	596	609
SUA	All Codes	15,344	581.36	19.41	555	569	581	595	606
ELL/MLL	ELL/MLL=Y	18,154	582.88	18.57	559	572	583	595	606
SWD/SUA	SWD & SUA codes	12,465	579.12	18.92	555	565	580	592	604
ELL/MLL/SUA	SUA & ELL/MLL codes	1,474	575.93	16.80	555	565	577	588	597

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
ELL/MLL Test Language	Chinese	540	608.54	16.45	588	598	609	619	628
	English	177,709	600.10	20.66	572	588	601	613	624
	Haitian-Creole	105	579.15	16.22	555	569	580	591	600
	Korean	45	605.89	19.77	583	589	605	619	634
	Russian	183	596.30	18.45	575	585	596	609	621
	Spanish	3,189	579.13	18.37	555	565	580	592	602
	All Translations	4,062	584.11	20.94	555	569	583	597	612

8.1.2.4. Mathematics Grade 6

Table 8.14 presents the Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 600.24 with a standard deviation of 20.17. Female and male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (613.94). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, 0.85 standard deviation below the mean scale score for the population. ELL/MLL students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 22 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (602): Asian (616), Multiracial (604), Pacific Islander (606), and White (606) students, as well as those enrolled in Average (604) and Low (612) Needs districts, and Charter schools (606). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 581 (Haitian-Creole, n = 118, Spanish, n = 4,171) to 612 (Korean, n = 29).

Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	179,611	600.24	20.17	573	588	602	614	625
Gender	Female	87,744	600.82	19.89	573	589	602	614	625
	Male	91,867	599.68	20.41	573	588	602	614	625
Ethnicity	Asian	19,458	613.94	18.64	591	603	616	625	636
	African American	32,230	592.80	20.00	565	581	594	606	617
	Hispanic	50,787	594.52	19.25	565	584	596	607	617
	American Indian	1,257	596.61	19.39	567	586	598	609	621
	Multiracial	4,336	602.55	20.67	575	590	604	617	628
	Pacific Islander	533	604.70	18.66	581	594	606	617	628
	White	70,771	603.91	18.34	581	594	606	616	625

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
NRC	New York	67,095	599.50	21.40	573	586	600	614	625
	Big 4 Cities	7,524	587.48	20.40	561	573	588	602	614
	Urban/Suburban	12,890	591.44	19.22	563	580	593	605	614
	Rural	9,409	596.70	18.26	573	588	599	609	618
	Average Needs	38,256	602.31	17.84	580	593	604	614	623
	Low Needs	18,179	610.82	16.24	591	602	612	621	631
	Charter	12,513	604.46	18.87	581	593	606	617	628
	Religious and Independent	13,557	598.08	19.01	573	588	600	610	621
SWD	All Codes	28,654	583.14	18.79	556	567	584	596	607
SUA	All Codes	12,872	583.11	18.69	556	567	584	596	607
ELL/MLL	ELL/MLL=Y	16,987	583.22	18.16	556	573	584	596	606
SWD/SUA	SWD & SUA codes	10,250	580.53	17.96	556	565	581	593	604
ELL/MLL/SUA	SUA & ELL/MLL codes	1,038	577.89	16.39	556	565	578	589	600
ELL/MLL Test Language	Chinese	738	607.36	16.43	588	598	607	619	625
	English	174,376	600.70	20.00	573	589	602	614	625
	Haitian-Creole	118	577.47	16.66	556	561	581	591	599
	Korean	29	607.55	17.23	573	603	612	619	628
	Russian	179	597.41	16.90	578	589	598	608	619
	Spanish	4,171	580.48	17.16	556	565	581	592	602
	All Translations	5,235	584.93	19.65	556	573	586	598	609

8.1.2.5. Mathematics Grade 7

Table 8.15 presents the Grade 7 n-counts and scale score statistics for key demographic subgroups. The population scale score mean was 600.78 with a standard deviation of 20.68. Female students tended to outperform male students by around two scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (614.67). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, 0.85 standard deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 24 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (602): Female (603), Asian (616), Multiracial (604), Pacific Islander (607), and White (606) students, those enrolled in Average (604) and Low Needs (613) districts, and Charter schools (606). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 582 (Spanish, n = 4,072 and Haitian-Creole, n = 95) to 612 (Chinese, n = 744).

Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	168,909	600.78	20.68	575	588	602	615	626
Gender	Female	82,198	601.88	19.84	579	590	603	615	627
	Male	86,711	599.74	21.40	575	588	602	614	624
Ethnicity	Asian	18,290	614.67	18.85	591	604	616	627	639
	African American	30,766	593.14	20.41	569	582	594	607	618
	Hispanic	47,488	595.26	19.74	569	584	596	608	618
	American Indian	1,171	597.65	19.87	575	586	599	612	621
	Multiracial	3,625	602.86	21.00	577	590	604	616	627
	Pacific Islander	606	605.62	20.03	579	594	607	618	631
	White	66,753	604.38	19.08	582	594	606	616	627
NRC	New York	65,590	600.90	21.84	575	588	602	615	627
	Big 4 Cities	6,728	585.55	21.34	552	575	586	600	612
	Urban/Suburban	11,923	590.53	19.56	557	581	593	604	613
	Rural	8,959	596.61	18.35	575	586	598	609	618
	Average Needs	34,741	602.25	18.31	581	593	604	614	623
	Low Needs	17,517	611.24	16.44	591	603	613	621	631
	Charter	11,147	604.96	18.41	582	594	606	616	627
	Religious and Independent	12,094	598.81	20.03	575	588	601	612	621
SWD	All Codes	27,019	583.05	19.84	552	571	584	596	607
SUA	All Codes	13,573	583.03	19.76	552	571	584	596	607
ELL/MLL	ELL/MLL=Y	14,640	583.33	19.40	552	575	584	595	606
SWD/SUA	SWD & SUA codes	10,882	580.36	19.07	552	569	582	593	603
ELL/MLL/SUA	SUA & ELL/MLL codes	1,147	577.02	18.02	552	569	579	590	598
ELL/MLL Test Language	Chinese	744	608.76	18.62	586	599	612	621	631
	English	163,834	601.28	20.48	575	590	603	615	627
	Haitian-Creole	95	580.55	16.63	557	575	582	590	599
	Korean	23	605.13	14.85	588	595	605	615	624
	Russian	141	592.75	19.06	575	584	593	604	616
	Spanish	4,072	580.15	18.31	552	569	582	593	601
	All Translations	5,075	584.81	21.01	552	575	586	597	612

8.1.2.6. Mathematics Grade 8

Table 8.16 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.39 with a standard deviation of 21.67. Female students tended to outperform male students by around four scale score points. Asian,

Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average Needs districts, Low Needs districts, Charter schools, and Religious and Independent schools. Across ethnic groups, Asian students earned the highest mean score (614.66). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, two-thirds of a standard deviation below the mean scale score for the population. English Language Learners/Multilingual Learners tested under accommodations were the lowest performing subgroup analyzed for English forms, scoring about 22 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (602): Female (604), Asian (616), Pacific Islander (608), and White (605) students, as well as those enrolled in Average Needs districts (603), Low Needs (610) districts, and Charter (609) and Religious and Independent (603) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 583 (Spanish, n = 3,310 and Haitian-Creole, n = 81) to 614 (Chinese, n = 662).

Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	115,886	599.39	21.67	569	588	602	614	624
Gender	Female	54,845	601.60	21.04	575	590	604	615	625
	Male	61,041	597.40	22.04	569	586	600	612	623
Ethnicity	Asian	10,599	614.66	21.04	588	603	616	627	639
	African American	23,625	593.11	21.92	557	580	595	608	618
	Hispanic	34,785	595.86	21.13	569	583	597	610	620
	American Indian	948	595.76	22.26	569	583	596	611	622
	Multiracial	2,072	598.98	21.10	571	588	601	613	623
	Pacific Islander	344	604.18	22.79	575	592	608	620	630
	White	43,293	602.07	19.79	577	592	605	614	623
NRC	New York	49,471	600.68	22.83	569	588	602	616	627
	Big 4 Cities	5,474	584.61	23.14	552	569	585	600	615
	Urban/Suburban	8,082	588.72	19.23	557	580	592	602	611
	Rural	6,892	594.91	19.13	569	585	598	608	616
	Average Needs	21,791	599.56	18.11	575	592	603	612	618
	Low Needs	8,545	608.24	17.12	588	600	610	618	627
	Charter	7,047	607.23	19.98	583	595	609	620	630
	Religious and Independent	8,399	599.74	22.29	569	588	603	614	625
SWD	All Codes	22,362	584.17	20.54	552	571	586	598	609
SUA	All Codes	10,941	584.47	20.66	552	571	586	598	610
ELL/MLL	ELL/MLL=Y	11,743	585.70	21.44	552	575	588	600	612
SWD/SUA	SWD & SUA codes	8,668	581.88	20.20	552	569	583	596	607

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
ELL/MLL/SUA	SUA & ELL/MLL codes	752	577.07	18.63	552	559	580	590	601
ELL/MLL Test Language	Chinese	662	613.88	19.66	590	603	614	625	639
	English	111,680	599.82	21.50	571	588	602	614	625
	Haitian-Creole	81	582.59	21.21	552	569	583	600	608
	Korean	15	600.47	29.63	552	580	604	625	634
	Russian	138	602.21	20.23	580	592	604	615	627
	Spanish	3,310	582.35	19.93	552	569	583	596	607
	All Translations	4,206	588.03	23.16	557	575	588	603	617

8.2. Performance Level Distribution Summary

Students are classified as NYS Level I, NYS Level II, NYS Level III, or NYS Level IV. The cut scores were established in 2018 during the standards review. It is inappropriate to compare scale scores across grades because they neither measure the same content, nor are they on the same scale. During the standards review process, the established cut scores were revisited and updated separately for different grades within a subject. Additional care was taken to vertically articulate performance levels; see *2018 Standards Review Report* in Appendix T for details. While vertical articulation helps to build consistent meaning to the performance levels, the very nature of grade-specific content, differing performance expectations, and panel-set cut scores result in cut score differences across grades.

8.2.1. *ELA Test Performance Level Distributions*

Table 8.17 shows the performance level distribution for all examinees from public, charter, and non-public schools with valid ELA scores. Performance level data for selected subgroups of students were also examined. In general, these distributions reflect the same achievement trends in the scale score summary discussion. Across Tables 8.18 through 8.23, more Female students were classified in Level III and above subgroups than were Male students. Similarly, more Asian and Pacific Islander students were classified in Level III and above subgroups than were their peers from other reported ethnic groups. Consistent with the pattern shown in scale score distribution across the subgroups, students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, Urban/Suburban, and Rural). The Level III and above rates for students in the ELL/MLL, SWD, and SUA subgroups were low, compared to the total population of examinees.

Table 8.17. ELA Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
3	182,559	15.81	32.36	44.86	6.96	51.82
4	186,205	21.09	31.46	31.44	16.01	47.45
5	180,679	36.31	26.53	21.76	15.40	37.16
6	179,908	31.16	21.64	18.24	28.96	47.20

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
7	170,595	30.59	29.54	25.40	14.48	39.87
8	156,304	18.65	33.62	27.64	20.10	47.74

8.2.1.1. ELA Grade 3

Table 8.18 presents the ELA Grade 3 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 52% of students achieved Level III and Level IV. About 56% of Female students were at Level III or above, as compared to 48% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (72%) students and students from Low Needs districts (71%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 28–46% of students in those same performance categories. Only about 21% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (52%): Female (56%), Asian (72%), Multiracial (55%), Pacific Islander (58%), White (57%) students, and those enrolled in New York City (53%), Low Needs districts (71%), and Charter schools (69%).

Table 8.18. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	182,559	15.81	32.36	44.86	6.96	51.82
Gender	Female	90,073	12.91	31.07	47.77	8.25	56.02
	Male	92,486	18.64	33.63	42.03	5.71	47.74
Ethnicity	Asian	17,552	6.81	21.60	54.82	16.77	71.59
	African American	30,433	20.03	34.43	39.82	5.73	45.55
	Hispanic	50,693	19.63	36.75	39.11	4.51	43.62
	American Indian	1,214	14.74	37.31	41.35	6.59	47.94
	Multiracial	5,449	14.46	30.89	46.38	8.28	54.65
	Pacific Islander	400	13.25	28.50	51.75	6.50	58.25
	White	73,307	12.26	30.81	49.88	7.05	56.92
NRC	New York	64,815	15.64	31.11	43.82	9.42	53.25
	Big 4 Cities	7,520	35.63	35.97	26.36	2.05	28.40
	Urban/Suburban	14,590	24.66	39.83	33.02	2.49	35.51
	Rural	9,928	21.72	40.70	35.52	2.06	37.58
	Average Needs	43,452	13.67	35.78	46.33	4.22	50.55
	Low Needs	19,226	5.00	23.73	60.72	10.55	71.27
	Charter	12,608	7.44	23.74	57.02	11.80	68.82
	Religious and Independent	10,220	23.52	31.28	40.08	5.12	45.20

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
SWD	All Codes	25,838	39.42	38.55	20.56	1.47	22.03
SUA	All Codes	13,960	42.47	39.60	17.31	0.62	17.93
ELL/MLL	ELL/MLL=Y	19,640	40.51	40.79	18.12	0.58	18.70
SWD/ SUA	SWD & SUA codes	9,518	44.18	39.51	15.83	0.47	16.31
ELL/MLL/SUA	SUA & ELL/MLL codes	1,290	51.40	37.75	10.62	0.23	10.85

8.2.1.2. ELA Grade 4

Table 8.19 presents the ELA Grade 4 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 47% of students achieved Level III and Level IV. About 53% of Female students were at Level III or above, as compared to 42% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (69%) students and students from Low Needs districts (67%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 24–39% of students in those same performance categories. Only about 12% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (47%): Female (53%), Asian (69%), Multiracial (51%), Pacific Islander (53%), and White (53%) students as well as those enrolled in New York City (50%), Low Needs districts (67%), and Charter schools (62%).

Table 8.19. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	186,205	21.09	31.46	31.44	16.01	47.45
Gender	Female	91,797	17.31	29.80	33.43	19.46	52.89
	Male	94,408	24.77	33.07	29.50	12.66	42.16
Ethnicity	Asian	18,277	9.24	21.26	37.24	32.26	69.50
	African American	31,947	26.80	33.72	27.76	11.72	39.47
	Hispanic	51,288	25.98	35.01	28.19	10.81	39.00
	American Indian	1,207	22.12	33.89	30.82	13.17	43.99
	Multiracial	5,198	18.16	30.92	30.86	20.07	50.92
	Pacific Islander	439	15.95	30.98	33.71	19.36	53.08
	White	73,716	16.33	30.63	35.02	18.01	53.04
NRC	New York	66,915	20.64	29.80	30.60	18.96	49.56
	Big 4 Cities	7,836	44.60	31.06	18.57	5.77	24.34
	Urban/Suburban	14,243	32.43	36.45	23.88	7.25	31.12
	Rural	10,045	28.45	37.13	26.57	7.84	34.42
	Average Needs	42,281	19.31	35.31	32.79	12.59	45.38
	Low Needs	18,852	7.47	25.53	41.11	25.89	67.00

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Charter	11,816	11.28	27.20	38.78	22.74	61.52
	Religious and Independent	13,893	25.05	30.56	30.66	13.73	44.39
SWD	All Codes	26,941	49.21	33.43	14.05	3.31	17.36
SUA	All Codes	15,273	52.27	33.14	12.59	2.00	14.59
ELL/MLL	ELL/MLL=Y	18,195	56.31	33.66	9.29	0.74	10.03
SWD/ SUA	SWD & SUA codes	9,631	54.81	32.72	11.02	1.45	12.47
ELL/MLL/SUA	SUA & ELL/MLL codes	1,241	65.03	29.49	5.08	0.40	5.48

8.2.1.3. ELA Grade 5

Table 8.20 presents the ELA Grade 5 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 37% of students achieved Level III and Level IV. About 42% of Female students were at Level III or above, as compared to 32% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (59%) students and students from Low Needs districts (57%). The Big 4 Cities, High Needs/Urban/ Suburban, African American, and Hispanic students had a range of 17–29% of students in those same performance categories. Only about 6% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (37%): Female (42%), Asian (59%), Multiracial (42%), Pacific Islander (44%), and White (43%) students, as well as those enrolled in New York City (40%), Low Needs districts (57%), and Charter schools (43%).

Table 8.20. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	180,679	36.31	26.53	21.76	15.40	37.16
Gender	Female	88,646	31.01	26.64	23.59	18.76	42.35
	Male	92,033	41.41	26.43	19.99	12.17	32.16
Ethnicity	Asian	18,480	18.03	22.53	28.60	30.84	59.44
	African American	31,535	45.32	25.89	18.85	9.94	28.79
	Hispanic	49,244	43.64	27.50	18.73	10.13	28.86
	American Indian	1,195	41.00	25.94	20.17	12.89	33.05
	Multiracial	4,650	33.85	24.60	22.69	18.86	41.55
	Pacific Islander	470	29.36	26.17	25.96	18.51	44.47
	White	70,111	29.65	27.80	24.36	18.19	42.55
NRC	New York	66,345	35.56	24.68	21.80	17.96	39.76
	Big 4 Cities	7,686	63.22	20.08	11.31	5.40	16.71
	Urban/Suburban	13,886	51.54	26.22	15.31	6.93	22.24

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Rural	9,915	47.70	29.14	16.18	6.99	23.17
	Average Needs	41,384	34.11	29.85	22.34	13.71	36.05
	Low Needs	18,607	16.95	26.37	30.62	26.07	56.68
	Charter	12,322	28.15	28.36	26.59	16.90	43.49
	Religious and Independent	10,088	42.62	26.01	19.47	11.90	31.36
SWD	All Codes	26,666	70.69	18.98	7.58	2.76	10.34
SUA	All Codes	16,183	72.16	19.20	6.62	2.01	8.64
ELL/MLL	ELL/MLL=Y	15,369	80.83	15.67	2.93	0.57	3.51
SWD/ SUA	SWD & SUA codes	9,992	76.18	17.28	5.27	1.26	6.54
ELL/MLL/SUA	SUA & ELL/MLL codes	1,240	88.23	10.65	0.89	0.24	1.13

8.2.1.4. ELA Grade 6

Table 8.21 presents the ELA Grade 6 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 47% of students achieved Level III and Level IV. About 52% of Female students were at Level III or above, as compared to 42% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 23–37% of students in those same performance categories. Only about 7% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (47%): Female (52%), Asian (71%), Multiracial (53%), Pacific Islander (60%), and White (55%) students, as well as those from New York City (48%), Average (47%) and Low (68%) Needs districts, and Charter (50%) and Religious and Independent (49%) schools.

Table 8.21. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	179,908	31.16	21.64	18.24	28.96	47.20
Gender	Female	88,388	26.01	21.84	19.19	32.95	52.15
	Male	91,520	36.13	21.44	17.32	25.11	42.43
Ethnicity	Asian	18,948	13.86	14.87	18.40	52.87	71.27
	African American	32,036	40.30	24.07	17.00	18.63	35.63
	Hispanic	49,263	38.27	24.59	17.83	19.31	37.14
	American Indian	1,222	35.35	25.37	17.43	21.85	39.28
	Multiracial	4,263	27.09	19.68	17.05	36.17	53.23
	Pacific Islander	520	18.85	20.96	21.35	38.85	60.19

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	White	68,267	24.02	20.60	19.89	35.49	55.38
NRC	New York	66,341	30.74	20.97	17.19	31.11	48.30
	Big 4 Cities	7,536	56.46	20.25	11.64	11.65	23.29
	Urban/Suburban	12,805	47.50	23.38	14.67	14.44	29.11
	Rural	9,522	39.72	22.64	16.82	20.81	37.64
	Average Needs	39,000	29.89	22.93	19.22	27.96	47.18
	Low Needs	18,213	13.55	18.62	21.76	46.07	67.83
	Charter	12,567	25.09	24.58	22.50	27.83	50.33
	Religious and Independent	12,964	29.96	21.11	20.19	28.73	48.93
SWD	All Codes	26,267	66.21	19.23	8.66	5.90	14.57
SUA	All Codes	16,188	67.22	18.64	8.53	5.60	14.13
ELL/MLL	ELL/MLL=Y	14,471	76.16	16.91	4.99	1.94	6.93
SWD/ SUA	SWD & SUA codes	10,021	71.30	17.91	7.11	3.68	10.79
ELL/MLL/SUA	SUA & ELL/MLL codes	1,260	83.02	13.33	3.17	0.48	3.65

8.2.1.5. ELA Grade 7

Table 8.22 presents the ELA Grade 7 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 40% of students achieved Level III and Level IV. About 47% of Female students were at Level III or above, as compared to 33% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (64%) students and students from Low Needs (59%) districts. The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 17–31% of students in those same performance categories. Only about 6% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (40%): Female (47%), Asian (64%), Multiracial (46%), Pacific Islander (54%), and White (46%) students, as well as those enrolled in New York City (43%), Low Needs (59%) districts, Charter (44%) and Religious and Independent schools (42%).

Table 8.22. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	170,595	30.59	29.54	25.40	14.48	39.87
Gender	Female	83,291	23.41	29.88	29.02	17.69	46.71
	Male	87,304	37.44	29.22	21.94	11.41	33.34
Ethnicity	Asian	18,157	14.33	21.25	31.50	32.91	64.42
	African American	31,111	38.43	32.17	21.69	7.71	29.39

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	Hispanic	46,376	36.69	32.33	22.15	8.83	30.98
	American Indian	1,162	34.94	30.81	23.06	11.19	34.25
	Multiracial	3,632	26.73	27.34	26.46	19.47	45.93
	Pacific Islander	599	20.87	25.38	33.22	20.53	53.76
	White	65,141	24.80	29.15	28.88	17.17	46.05
NRC	New York	65,094	28.88	28.42	24.92	17.79	42.71
	Big 4 Cities	6,835	57.85	25.28	12.82	4.05	16.87
	Urban/Suburban	12,098	49.38	29.40	16.30	4.92	21.22
	Rural	9,177	39.64	33.62	20.41	6.33	26.74
	Average Needs	35,784	30.90	31.62	25.82	11.65	37.48
	Low Needs	18,028	14.63	26.82	34.36	24.19	58.55
	Charter	11,214	21.37	34.32	31.03	13.28	44.31
	Religious and Independent	11,669	29.28	28.48	28.48	13.76	42.24
SWD	All Codes	25,094	65.63	24.68	7.93	1.75	9.68
SUA	All Codes	14,914	67.25	23.31	7.87	1.58	9.45
ELL/MLL	ELL/MLL=Y	12,322	81.28	16.11	2.30	0.31	2.61
SWD/ SUA	SWD & SUA codes	9,010	72.71	20.78	5.74	0.78	6.51
ELL/MLL/SUA	SUA & ELL/MLL codes	994	86.32	12.78	0.80	0.10	0.91

8.2.1.6. ELA Grade 8

Table 8.23 presents the ELA Grade 8 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 48% of students achieved Level III and Level IV. About 55% of Female students were at Level III or above, as compared to 41% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs (67%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 23–40% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (48%): Female (55%), Asian (71%), Multiracial (50%), Pacific Islander (56%), and White (54%) students, as well as those enrolled in New York City (51%) and Low Needs (67%) districts, Charter (54%) and Religious and Independent (50%) schools.

Table 8.23. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	156,304	18.65	33.62	27.64	20.10	47.74
Gender	Female	75,519	12.94	31.59	30.71	24.75	55.46
	Male	80,785	23.98	35.51	24.77	15.75	40.52
Ethnicity	Asian	17,132	8.50	20.92	29.16	41.42	70.58
	African American	29,663	23.37	38.74	25.54	12.35	37.89
	Hispanic	42,702	21.84	38.19	26.90	13.08	39.98
	American Indian	1,212	22.28	37.21	25.33	15.18	40.51
	Multiracial	2,870	18.22	32.13	25.44	24.22	49.65
	Pacific Islander	456	16.01	27.85	32.24	23.90	56.14
	White	58,409	14.79	31.32	30.02	23.87	53.89
NRC	New York	63,774	16.25	33.17	27.78	22.80	50.58
	Big 4 Cities	6,425	43.42	33.49	15.56	7.52	23.08
	Urban/Suburban	10,698	33.70	38.18	19.36	8.77	28.13
	Rural	8,339	24.68	39.32	24.42	11.58	36.00
	Average Needs	31,684	19.48	35.79	27.49	17.24	44.73
	Low Needs	15,813	8.34	24.55	33.93	33.18	67.11
	Charter	10,127	9.23	36.33	34.84	19.60	54.44
	Religious and Independent	8,604	18.99	31.00	30.00	20.01	50.01
SWD	All Codes	23,802	44.59	40.76	11.83	2.82	14.65
SUA	All Codes	13,644	48.51	37.44	10.79	3.26	14.05
ELL/MLL	ELL/MLL=Y	10,905	61.72	33.50	4.25	0.53	4.78
SWD/ SUA	SWD & SUA codes	8,542	52.81	36.99	8.44	1.76	10.20
ELL/MLL/SUA	SUA & ELL/MLL codes	868	68.89	28.57	2.53	0.00	2.53

8.2.2. *Mathematics Test Performance Level Distributions*

Table 8.24 shows the performance level distributions for all examinees from public, charter, and non-public schools with valid scores, and presents Mathematics performance level data for total populations of students in Grades 3–8. Performance level data for selected subgroups of students were also examined. In general, these summaries reflect the same achievement trends as in the scale score summary discussion. Across Table 8.25 through Table 8.30, Male and Female students performed similarly across grades. More White, Pacific Islander, and Asian students were classified in Level III and above, as compared to their peers from other ethnic subgroups. Students from Low and Average Needs districts and Charter schools outperformed students from High Needs districts (New York City, Big 4 Cities, High Needs Urban/Suburban, and High Needs Rural), and Religious and Independent schools. The subgroups that used the Korean or Chinese translations outperformed other test translation subgroups. The Level III and above rates for SWD and SUA subgroups were low, compared to the total population of examinees. The n-counts for

the Haitian-Creole, Korean, and Russian translation subgroups were very low, and the results might have been heavily influenced by very high and/or very low achieving individual students.

Table 8.24. Mathematics Test Performance Level Distributions

		Performance Levels				
Grade	N-Count	Level I	Level II	Level III	Level IV	Level III & IV
3	184,576	21.96	24.01	30.36	23.68	54.03
4	188,143	23.50	26.82	24.08	25.60	49.68
5	181,771	32.69	21.78	21.63	23.90	45.53
6	179,611	29.90	23.62	23.82	22.65	46.48
7	168,909	30.12	26.85	22.80	20.23	43.03
8	115,886	38.31	28.35	19.84	13.50	33.34

8.2.2.1. Mathematics Grade 3

Table 8.25 presents the Mathematics Grade 3 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 54% of students achieved Level III and Level IV. About 54% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (77%) students and students from Low Needs districts (74%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 29–43% of students in those same performance categories. Only about 17% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (54%): Asian (77%), Multiracial (56%), Pacific Islander (58%), and White (60%) students, as well as those enrolled at Average (55%) and Low (74%) Needs districts and Charter schools (72%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 15% (Haitian-Creole) to 39% (Russian).

Table 8.25. Mathematics Grade 3 Performance Level Distribution by Subgroup

			Performance Levels				
Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	184,576	21.96	24.01	30.36	23.68	54.03
Gender	Female	90,611	21.09	25.32	30.98	22.60	53.58
	Male	93,965	22.80	22.73	29.75	24.71	54.46
Ethnicity	Asian	18,150	7.88	14.77	31.26	46.09	77.34
	African American	30,809	30.44	26.54	26.09	16.93	43.02
	Hispanic	52,488	28.66	28.03	27.61	15.70	43.32
	American Indian	1,232	22.24	27.27	29.87	20.62	50.49
	Multiracial	5,534	21.36	22.97	29.56	26.11	55.67
	Pacific Islander	404	16.34	25.99	30.20	27.48	57.67
	White	75,689	17.23	22.40	33.88	26.49	60.37

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	66,200	22.41	24.44	29.00	24.16	53.15
	Big 4 Cities	7,607	46.23	24.85	18.71	10.21	28.92
	Urban/Suburban	14,796	33.89	28.04	25.49	12.58	38.07
	Rural	9,926	28.93	27.40	28.94	14.72	43.66
	Average Needs	43,469	19.40	25.25	33.47	21.89	55.35
	Low Needs	19,377	7.79	17.77	36.81	37.62	74.43
	Charter	12,618	10.22	18.06	31.75	39.97	71.72
	Religious and Independent	10,447	28.73	25.34	29.21	16.72	45.94
SWD	All Codes	28,400	49.39	24.99	17.53	8.08	25.62
SUA	All Codes	14,896	55.40	23.89	15.40	5.31	20.71
ELL/MLL	ELL/MLL=Y	22,341	45.41	28.31	19.33	6.95	26.28
SWD/ SUA	SWD & SUA codes	12,667	58.93	22.70	13.88	4.48	18.36
ELL/MLL/SUA	SUA & ELL/MLL codes	1,797	61.49	22.76	12.80	2.95	15.75
ELL/MLL Test Language	Chinese	464	12.50	18.10	32.97	36.42	69.40
	English	180,270	21.41	23.95	30.63	24.02	54.65
	Haitian-Creole	75	58.67	25.33	14.67	1.33	16.00
	Korean	47	10.64	23.40	23.40	42.55	65.96
	Russian	195	16.41	25.64	38.97	18.97	57.95
	Spanish	3,525	51.43	27.63	15.80	5.13	20.94
	All Translations	4,306	45.33	26.43	18.76	9.48	28.24

8.2.2.2. Mathematics Grade 4

Table 8.26 presents the Mathematics Grade 4 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 50% of students achieved Level III and Level IV. About 50% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (76%) students and students from Low Needs (73%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 23–38% of students in those same performance categories. Only about 12% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (50%): Asian (76%), Multiracial (52%), Pacific Islander (53%), and White (57%) students, as well as students enrolled in Average (52%) and Low (73%) Needs districts and Charter schools (65%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 5% (Haitian-Creole) to 26% (Korean).

Table 8.26. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	188,143	23.50	26.82	24.08	25.60	49.68
Gender	Female	92,231	22.78	28.30	24.56	24.36	48.92
	Male	95,912	24.19	25.40	23.62	26.79	50.41
Ethnicity	Asian	18,930	7.89	15.79	23.06	53.26	76.32
	African American	32,333	34.65	29.39	19.91	16.05	35.96
	Hispanic	53,110	30.81	30.69	21.89	16.61	38.50
	American Indian	1,242	26.41	28.18	24.32	21.10	45.41
	Multiracial	5,356	22.39	25.28	23.39	28.94	52.33
	Pacific Islander	451	20.18	27.05	25.28	27.49	52.77
	White	76,490	17.52	25.86	27.70	28.91	56.61
NRC	New York	68,186	24.35	26.35	22.00	27.30	49.31
	Big 4 Cities	7,934	52.26	25.23	14.09	8.42	22.51
	Urban/Suburban	14,347	37.05	30.09	20.30	12.56	32.86
	Rural	9,972	27.80	31.89	25.33	14.98	40.31
	Average Needs	42,546	19.62	28.67	27.36	24.35	51.71
	Low Needs	19,148	7.56	19.81	30.08	42.55	72.63
	Charter	11,784	13.11	22.37	25.30	39.22	64.52
	Religious and Independent	13,988	28.00	30.89	23.74	17.37	41.11
SWD	All Codes	29,692	54.01	26.20	12.60	7.19	19.79
SUA	All Codes	17,975	56.03	26.31	12.17	5.49	17.66
ELL/MLL	ELL/MLL=Y	20,999	53.48	29.54	11.79	5.20	16.98
SWD/ SUA	SWD & SUA codes	15,246	59.86	24.80	10.72	4.61	15.34
ELL/MLL/SUA	SUA & ELL/MLL codes	2,252	66.21	23.31	8.26	2.22	10.48
ELL/MLL Test Language	Chinese	500	11.60	22.60	25.60	40.20	65.80
	English	183,630	22.75	26.86	24.38	26.01	50.39
	Haitian-Creole	79	64.56	26.58	5.06	3.80	8.86
	Korean	38	5.26	23.68	26.32	44.74	71.05
	Russian	156	26.92	28.85	24.36	19.87	44.23
	Spanish	3,740	60.83	25.48	9.81	3.88	13.69
	All Translations	4,513	53.80	25.28	12.12	8.80	20.92

8.2.2.3. Mathematics Grade 5

Table 8.27 presents the Mathematics Grade 5 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 46% of students achieved Level III and Level IV. About 45% of both Female and Male students were at Level III or above. The percentage of

students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (74%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 20–34% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (46%): Asian (74%), Pacific Islander (55%), and White (52%) students, as well as those enrolled in Average (47%) and Low (68%) Needs districts and Charter schools (58%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 6% (Spanish) to 28% (Chinese).

Table 8.27. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	181,771	32.69	21.78	21.63	23.90	45.53
Gender	Female	88,580	31.83	22.87	21.90	23.41	45.31
	Male	93,191	33.51	20.74	21.38	24.37	45.75
Ethnicity	Asian	19,137	12.09	14.00	22.34	51.57	73.91
	African American	31,995	45.55	22.37	16.70	15.38	32.08
	Hispanic	51,150	41.60	24.00	18.79	15.62	34.40
	American Indian	1,217	35.83	23.25	20.54	20.38	40.92
	Multiracial	4,763	32.94	20.60	20.83	25.64	46.46
	Pacific Islander	481	21.62	23.28	23.91	31.19	55.09
	White	72,749	26.03	22.07	25.73	26.16	51.90
NRC	New York	67,518	32.81	21.14	20.00	26.05	46.05
	Big 4 Cities	7,768	62.71	17.70	11.86	7.74	19.59
	Urban/Suburban	13,879	49.48	22.83	16.15	11.54	27.69
	Rural	9,835	40.59	24.92	21.27	13.22	34.49
	Average Needs	41,144	28.36	24.30	25.46	21.89	47.35
	Low Needs	18,661	12.98	18.95	29.10	38.96	68.06
	Charter	12,290	22.45	19.60	22.01	35.94	57.95
	Religious and Independent	10,415	43.47	22.27	18.53	15.74	34.27
SWD	All Codes	29,549	66.09	18.04	9.68	6.19	15.87
SUA	All Codes	15,344	69.79	17.08	8.69	4.44	13.13
ELL/MLL	ELL/MLL=Y	18,154	68.17	18.71	8.59	4.53	13.12
SWD/ SUA	SWD & SUA codes	12,465	74.57	15.11	6.95	3.37	10.32
ELL//MLL/SUA	SUA & ELL/MLL codes	1,474	81.55	13.64	3.66	1.15	4.82
ELL/MLL Test Language	Chinese	540	14.81	20.74	28.15	36.30	64.44
	English	177,709	31.98	21.85	21.91	24.26	46.17

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
ELL/MLL Test Language	Haitian-Creole	105	77.14	14.29	7.62	0.95	8.57
	Korean	45	28.89	17.78	17.78	35.56	53.33
	Russian	183	44.81	21.31	17.49	16.39	33.88
	Spanish	3,189	73.38	17.97	6.02	2.63	8.65
	All Translations	4,062	63.91	18.39	9.65	8.05	17.70

8.2.2.4. Mathematics Grade 6

Table 8.28 presents the Mathematics Grade 6 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 46% of students achieved Level III and Level IV. About 46% of Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (74%) students and students from Low Needs districts (72%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 22–33% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (46%): Asian (74%), Multiracial (52%), Pacific Islander (54%), and White (55%) students, as well as those enrolled in Average (51%) and Low (72%) Needs districts and Charter schools (55%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 3% (Haitian-Creole) to 31% (Korean).

Table 8.28. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	179,611	29.90	23.62	23.82	22.65	46.48
Gender	Female	87,744	28.74	24.35	23.78	23.12	46.90
	Male	91,867	31.01	22.92	23.87	22.21	46.08
Ethnicity	Asian	19,458	10.37	15.56	23.97	50.10	74.07
	African American	32,230	44.58	24.72	18.32	12.38	30.70
	Hispanic	50,787	40.12	26.46	20.52	12.90	33.41
	American Indian	1,257	36.36	26.65	20.92	16.07	36.99
	Multiracial	4,336	26.98	21.26	23.89	27.86	51.75
	Pacific Islander	533	19.51	26.27	24.58	29.64	54.22
	White	70,771	21.26	23.38	28.76	26.60	55.36
NRC	New York	67,095	33.12	23.03	20.56	23.30	43.86
	Big 4 Cities	7,524	57.12	20.43	13.85	8.60	22.45
	Urban/Suburban	12,890	46.99	25.24	18.03	9.74	27.77
	Rural	9,409	33.88	27.77	24.82	13.53	38.35
	Average Needs	38,256	23.43	25.20	28.92	22.45	51.37

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Low Needs	18,179	10.07	18.23	31.38	40.33	71.70
	Charter	12,513	22.06	22.92	25.97	29.05	55.02
	Religious and Independent	13,557	31.62	27.30	23.96	17.12	41.08
SWD	All Codes	28,654	66.47	19.89	9.51	4.14	13.65
SUA	All Codes	12,872	66.18	20.05	10.01	3.75	13.77
ELL/MLL	ELL/MLL=Y	16,987	66.52	21.22	8.69	3.57	12.26
SWD/ SUA	SWD & SUA codes	10,250	72.02	17.90	7.79	2.29	10.08
ELL/MLL/SUA	SUA & ELL/MLL codes	1,038	79.58	14.16	5.20	1.06	6.26
ELL/MLL Test Language	Chinese	738	14.77	24.12	27.78	33.33	61.11
	English	174,376	28.90	23.73	24.22	23.14	47.36
	Haitian-Creole	118	77.97	18.64	3.39	0.00	3.39
	Korean	29	17.24	17.24	31.03	34.48	65.52
	Russian	179	32.96	31.84	21.23	13.97	35.20
	Spanish	4,171	73.00	18.58	7.07	1.34	8.42
	All Translations	5,235	63.23	19.81	10.53	6.44	16.96

8.2.2.5. Mathematics Grade 7

Table 8.29 presents the Mathematics Grade 7 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 43% of students achieved Level III and Level IV. About 44% of Female students were at Level III or above, as compared to 42% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 18–30% of students in those same performance categories. Only about 8% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (43%): Female (44%), Asian (71%), Multiracial (48%), Pacific Islander (52%), and White (52%) students, as well as those enrolled in Average (47%) and Low (68%) Needs districts and Charter schools (50%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 5% (Spanish) to 26% (Korean).

Table 8.29. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	168,909	30.12	26.85	22.80	20.23	43.03
Gender	Female	82,198	27.88	28.14	23.14	20.84	43.98
	Male	86,711	32.24	25.63	22.48	19.65	42.13

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	Asian	18,290	11.08	17.50	23.56	47.87	71.43
	African American	30,766	44.92	28.27	16.58	10.23	26.81
	Hispanic	47,488	39.64	30.19	18.96	11.21	30.17
	American Indian	1,171	35.95	27.75	20.92	15.37	36.29
	Multiracial	3,625	27.61	24.77	23.39	24.22	47.61
	Pacific Islander	606	22.77	25.08	24.75	27.39	52.15
	White	66,753	21.73	26.51	28.23	23.53	51.76
NRC	New York	65,590	32.18	25.76	19.31	22.76	42.07
	Big 4 Cities	6,728	61.28	21.15	12.02	5.54	17.57
	Urban/Suburban	11,923	49.37	29.41	14.89	6.34	21.23
	Rural	8,959	35.27	31.85	22.87	10.01	32.88
	Average Needs	34,741	24.38	28.63	28.78	18.21	46.99
	Low Needs	17,517	10.30	21.81	31.85	36.03	67.89
	Charter	11,147	22.06	27.86	25.87	24.21	50.09
	Religious and Independent	12,094	31.02	31.10	22.42	15.45	37.88
SWD	All Codes	27,019	67.52	21.27	7.96	3.25	11.21
SUA	All Codes	13,573	67.27	21.22	8.66	2.85	11.51
ELL/MLL	ELL/MLL=Y	14,640	67.25	22.17	7.41	3.16	10.57
SWD/ SUA	SWD & SUA codes	10,882	73.54	18.80	5.97	1.68	7.65
ELL/MLL/SUA	SUA & ELL/MLL codes	1,147	80.56	15.78	3.14	0.52	3.66
ELL/MLL Test Language	Chinese	744	14.65	23.79	27.69	33.87	61.56
	English	163,834	29.04	27.05	23.25	20.66	43.91
	Haitian-Creole	95	76.84	17.89	5.26	0.00	5.26
	Korean	23	21.74	34.78	26.09	17.39	43.48
	Russian	141	49.65	26.24	16.31	7.80	24.11
	Spanish	4,072	74.46	19.65	4.62	1.28	5.89
	All Translations	5,075	64.81	20.47	8.43	6.29	14.72

8.2.2.6. Mathematics Grade 8

Table 8.30 presents the Mathematics Grade 8 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 33% of students achieved Level III and Level IV. About 37% of Female students were at Level III or above, as compared to 30% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (64%) students and students from Low Needs districts (52%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 12–26% of students in those same performance categories. Only about 8% of the SWD, SUA,

and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (33%): Female (37%), Asian (64%), Pacific Islander (46%), and White (38%) students, as well as those enrolled in New York City (36%) and Low Needs districts (52%), and Charter (48%) and Religious and Independent (35%) schools. For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 4% (Haitian-Creole) to 26% (Chinese).

Table 8.30. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	115,886	38.31	28.35	19.84	13.50	33.34
Gender	Female	54,845	33.78	29.26	21.64	15.32	36.96
	Male	61,041	42.38	27.53	18.22	11.86	30.09
Ethnicity	Asian	10,599	15.86	19.90	24.93	39.32	64.24
	African American	23,625	51.49	25.82	14.33	8.36	22.69
	Hispanic	34,785	45.67	28.43	16.19	9.70	25.89
	American Indian	948	47.78	24.47	16.77	10.97	27.74
	Multiracial	2,072	38.42	29.05	20.46	12.07	32.53
	Pacific Islander	344	29.07	24.71	23.84	22.38	46.22
	White	43,293	30.36	31.87	24.61	13.16	37.77
NRC	New York	49,471	38.35	25.68	18.21	17.76	35.97
	Big 4 Cities	5,474	69.18	15.82	9.50	5.50	15.00
	Urban/Suburban	8,082	60.08	28.22	9.55	2.14	11.69
	Rural	6,892	45.02	33.36	16.50	5.12	21.62
	Average Needs	21,791	33.37	35.65	23.99	6.99	30.98
	Low Needs	8,545	17.69	30.31	32.53	19.46	52.00
	Charter	7,047	25.00	26.56	24.99	23.44	48.43
	Religious and Independent	8,399	35.71	28.97	21.12	14.20	35.33
SWD	All Codes	22,362	70.13	20.21	7.41	2.24	9.66
SUA	All Codes	10,941	68.92	21.03	7.64	2.41	10.05
ELL/MLL	ELL/MLL=Y	11,743	67.05	20.47	8.34	4.14	12.48
SWD/ SUA	SWD & SUA codes	8,668	74.42	18.38	5.66	1.53	7.20
ELL/MLL/SUA	SUA & ELL/MLL codes	752	84.71	12.63	2.39	0.27	2.66
ELL/MLL Test Language	Chinese	662	14.65	23.41	25.83	36.10	61.93
	English	111,680	37.37	28.68	20.23	13.72	33.95
	Haitian-Creole	81	67.90	24.69	3.70	3.70	7.41
	Korean	15	40.00	20.00	6.67	33.33	40.00
	Russian	138	34.78	28.99	19.57	16.67	36.23

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
ELL/MLL Test Language	Spanish	3,310	74.29	18.16	5.83	1.72	7.55
	All Translations	4,206	63.36	19.47	9.39	7.77	17.17

Section 9: References

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

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37: 29–51.

Cai, L., Thissen, D. J., & du Toit, S. (2011). IRTPRO (Version 2.1). Skokie, IL: Scientific Software International, Inc.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334.

Dorans, N.J., A.P. Schmitt & C.A. Bleistein (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement* 29: 309–319.

Dorans, N.J. & P. W. Holland (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.

Hambleton, R. K, & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer Academic Publishers.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kim, S. & M. J. Kolen (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models*. Iowa City, IA: Iowa Testing Programs, The University of Iowa.

Kolen, M.J. & Z. Cui (2004). *POLYEQUATE*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Kolen, M.J. & R.L. Brennan (1995). *Test Equating: Methods and Practices.* New York: Springer-Verlag.

Landis, J. R. & G. G. Koch. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174.

Lee, W. C., B.A. Hanson & R.L. Brennan (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement* 26:412–432.

Lee, W. C. (2008). *Classification consistency and accuracy for complex assessments using item response theory.* (CASMA Research Report No. 27). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Lee, W. C. & M. J. Kolen (2006, Revised 2008). IRT-CLASS (Version 2.0). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Livingston, S.A. & C. Lewis (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement* 32: 179–197.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum.

Lord, F.M. & M.R. Novick (1968). *Statistical Theories of Mental Test Scores.* Menlo Park, CA: Addison-Wesley.

Mehrens, W.A. & I.J. Lehmann (1991). *Measurement and Evaluation in Education and Psychology, 3rd ed*. New York: Holt, Rinehart, and Winston.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement* 16: 159–176.

Muraki, E. & R.D. Bock (1991). *PARSCALE*: *Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.

NYSED. (2013) New York State Testing Program 2013: English Language Arts and Mathematics Grades 3–8 Technical Report. Albany, NY: New York State Education Department (NYSED). Retrieved from: http://www.p12.nysed.gov/assessment/reports/2013/ela-math-tr13.pdf

Qualls, A.L. (1995). Estimating the reliability of a test containing multiple-item formats. *Applied Measurement in Education* 8: 111–120.

Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics* 4: 207–230.

Sandoval, J.H. & M.P. Mille (1979) *Accuracy of judgments of WISC-R item difficulty for minority groups*. Paper presented at the annual meeting of the American Psychological Association, New York. August.

Sinharay, S. (2006) Bayesian item fit analysis for unidimensional item response theory models. *British Journal of Mathematical and Statistical Psychology*, 59: 429–449.

Stocking, M.L. & F.M. Lord (1983). Developing a common metric in item response theory. *Applied Psychological Measurement* 7: 201–210.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* 47: 175–186.

Thompson, S.J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments (NCEO Synthesis Report 44)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from: http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis44.html.

Wang, T.M., J. Kolen, & D.J. Harris (2000). Psychometric properties of scale scores and performance levels for performance assessment using polytomous IRT. *Journal of Educational Measurement* 37: 141–162.

Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement* 30: 187–213.

Yen, W.M. (1984). Obtaining maximum likelihood trait estimates from number correct scores for the three-parameter logistic model. *Journal of Educational Measurement* 21: 93–111.

Yen, W.M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement* 5: 245–262.

Zwick, R., J.R. Donoghue & A. Grima, (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement* 36: 225–33.

Appendix A: ELA and Mathematics Test Configurations

Table A1. ELA Test Configuration

			Number of Items				
			Multiple-Choice		Constructed-Response		
Grade	Day	Session	Operational	Embedded	Operational	Embedded	Total
3	1	1	18	6	0	0	24
	2	2	0	0	7	0	7
		Total	18	6	7	0	31
4	1	1	18	6	0	0	24
	2	2	0	0	7	0	7
		Total	18	6	7	0	31
5	1	1	28	7	0	0	35
	2	2	0	0	7	0	7
		Total	28	7	7	0	42
6	1	1	28	7	0	0	35
	2	2	0	0	7	0	7
		Total	28	7	7	0	42
7	1	1	28	7	0	0	35
	2	2	0	0	8	0	8
		Total	28	7	8	0	43
8	1	1	28	7	0	0	35
	2	2	0	0	8	0	8
		Total	28	7	8	0	43

Table A2. Mathematics Test Configuration

			Number of Items				
			Multiple-Choice		Constructed-Response		
Grade	Day	Session	Operational	Embedded	Operational	Embedded	Total
3	1	1	19	6	0	0	25
	2	2	8	0	7	0	15
		Total	27	6	7	0	40
4	1	1	23	7	0	0	30
	2	2	8	0	7	0	15
		Total	31	7	7	0	45
5	1	1	23	7	0	0	30
	2	2	8	0	7	0	15
		Total	31	7	7	0	45
6	1	1	24	7	0	0	31
	2	2	7	0	8	0	15
		Total	31	7	8	0	46
7	1	1	26	7	0	0	33
	2	2	7	0	8	0	15
		Total	33	7	8	0	48
8	1	1	26	7	0	0	33
	2	2	7	0	8	0	15
		Total	33	7	8	0	48

Table A3. ELA Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)
3	1	1	42
	2	2	44
		Total	**86**
4	1	1	42
	2	2	44
		Total	**86**
5	1	1	57.5
	2	2	44
		Total	**101.5**
6	1	1	57.5
	2	2	44
		Total	**101.5**
7	1	1	57.5
	2	2	47
		Total	**104.5**
8	1	1	57.5
	2	2	47
		Total	**104.5**

Source: *2019 ELA and Mathematics Test Guides.*

The ELA estimated times on task were based on the following rules of thumb:

- Average time to read a passage—5 minutes
- Average time to respond to a multiple-choice question—1 minute
- Average time to respond to a two-point constructed-response question—3 minutes
- Average time to respond to a four-point constructed-response question—20 minutes

Table A4. Mathematics Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)
3	1	1	37.5
	2	2	51
		Total	**88.5**
4	1	1	45
	2	2	51
		Total	**96**
5	1	1	45
	2	2	51
		Total	**96**
6	1	1	46.5
	2	2	54.5
		Total	**101**
7	1	1	49.5
	2	2	54.5
		Total	**104**

Grade	Day	Session	Estimated Time on Task (min.)
8	1	1	49.5
	2	2	54.5
		Total	**104**

Source: *2019 ELA and Mathematics Test Guides.*

The Mathematics estimated times on task were based on the following rules of thumb:

- Average time to respond to a multiple-choice question—1.5 minutes
- Average time to respond to a two-point constructed-response question—5 minutes
- Average time to respond to a three-point constructed-response question—9 minutes

The testing times listed above do not include approximately 10 minutes reserved for preparation at the beginning of each session for handing out materials and reading directions. Additional details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the *2019 Teacher's Directions* and the *School Administrator's Manual*, which are accessible online:

- *2019 ELA Teacher's Directions*
  - Grades 3–5: http://www.p12.nysed.gov/assessment/ei/directions/2019/ela3-5-td-19.pdf
  - Grades 6–8: http://www.p12.nysed.gov/assessment/ei/directions/2019/ela6-8-td-19.pdf
- *2019 Mathematics Teacher's Directions*
  - Grades 3–5: http://www.p12.nysed.gov/assessment/ei/directions/2019/m3-5-td-19.pdf
  - Grades 6–8: http://www.p12.nysed.gov/assessment/ei/directions/2019/m6-8-td-19.pdf
- *2019 ELA and Mathematics Tests School Administrator's Manual*
  - http://www.p12.nysed.gov/assessment/sam/ei/eisam19.pdf
- *2019 ELA and Mathematics Test Guides*
  - https://www.engageny.org/resource/test-guides-english-language-arts-and-mathematics

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Strand	Point Range		% of Test	
			Target	Actual	Target	Actual
3	34	Literature	18	18	53%	53%
		Informational Text	16	16	47%	47%
4	34	Literature	16-18	18	47%-53%	53%
		Informational Text	16-18	16	47%-53%	47%
5	44	Literature	20-24	24	45%-55%	55%
		Informational Text	20-24	20	45%-55%	45%
6	44	Literature	20-24	24	45%-55%	55%
		Informational Text	20-24	20	45%-55%	45%
7	46	Literature	20-26	20	43%-57%	43%
		Informational Text	20-26	26	43%-57%	57%
8	46	Literature	20-26	20	43%-57%	43%
		Informational Text	20-26	26	43%-57%	57%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
3	42	Operations and Algebraic Thinking	17-21	19	40%-50%	45%
		Number and Operations in Base Ten	2-4	3	5%-10%	7%
		Number and Operations – Fractions	6-10	8	14%-24%	19%
		Measurement and Data	9-13	11	21%-31%	26%
		Geometry*	1-2	1	2%-5%	2%
4	46	Operations and Algebraic Thinking	7-11	9	15%-26%	20%
		Number and Operations in Base Ten	10-14	12	22%-30%	26%
		Number and Operations – Fractions	10-14	12	22%-30%	26%
		Measurement and Data	7-11	9	15%-26%	20%
		Geometry	4-6	4	9%-13%	9%
5	46	Operations and Algebraic Thinking	2-4	2	4%-9%	4%
		Number and Operations in Base Ten	10-14	13	22%-30%	28%
		Number and Operations – Fractions	16-20	18	35%-43%	39%
		Measurement and Data	10-14	12	22%-30%	26%
		Geometry*	1-2	1	2%-4%	2%
6	48	Ratios and Proportional Relationships	10-14	12	21%-29%	25%
		The Number System	9-13	8	19%-27%	17%

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
6	48	Expressions and Equations	16-22	22	33%-46%	46%
		Geometry	5-9	6	10%-19%	12%
7	50	Ratios and Proportional Relationships	12-16	14	24%-32%	28%
		The Number System	8-12	9	16%-24%	18%
		Expressions and Equations	13-19	17	26%-38%	34%
		Geometry	3-5	2	6%-10%	4%
		Statistics and Probability	6-10	8	12%-20%	16%
8	50	Expressions and Equations	18-24	18	36%-48%	36%
		Functions	11-15	15	22%-30%	30%
		Geometry	10-14	12	20%-28%	24%
		Statistics and Probability	3-5	5	6%-10%	10%

*There is a slight difference between the "Target % of Test" shown in these tables and the tables presented in the Guides to the 2019 Mathematics Tests. The guides were intended to provide general guidance regarding content coverage of mathematics domains so that classroom instruction would continue to cover the depth and breadth of the mathematics standards.

Appendix C: Passage Selection Guidelines for Assessing ELA

General Guidelines

The New York State Learning Standards for ELA devote considerable attention to the types and nature of texts used in instruction and assessment. The foundation for preparing students for the linguistic rigors of college and of the workplace lies in the texts with which they interact. By the time that they graduate, students should be prepared to successfully read and analyze the types of complex texts that they will encounter after high school. Selecting passages of appropriate type and complexity for use in assessment is integral to this preparation.

The New York State Learning Standards for ELA emphasize developing skills for comprehending and analyzing both literary and informational texts. Increased exposure to informational texts better prepares students for the various types of texts that they will encounter in college and in the workplace. The array of passages selected for assessment from K–12 should support the development of the necessary skills to handle a range of literary and informational texts.

In addition to the usual fairness and sensitivity guidelines when selecting passages for assessment, attention should be dedicated to three additional considerations:

- *Text Complexity*
- *Text Types*
- *Text Suitability for Specific Standards*

These guidelines should inform the training of passage finders, in order to ensure a pool of acceptable passages that can support assessment of all the Reading Informational Texts standards. They should also alert form assemblers as they construct forms that will assess the complete range of skills.

Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
A.	**Precisely Designed Constructs**
Definition	The item construct is clearly defined so that all irrelevant cognitive, sensory, emotional, and physical barriers are removed.
√	The item does not add skills to those being measured (no extraneous skills tested).
B.	**Language Appropriateness**
Definition	The item avoids words or phrases that are sexist, racist, or otherwise offensive, inappropriate, or negative to any subgroup. Language should be simple and clear.
√	The item uses commonly used words—simpler is better.
√	The item uses vocabulary appropriate for the grade level.
√	Idiomatic speech and figurative language are avoided unless being measured.
√	The item avoids technical terms unrelated to the content.
√	The item contains no unnecessary words.
√	The sentence complexity contained in the item is appropriate for the grade level.
√	The item avoids ambiguous or multiple-meaning words (e.g., crane—the bird—can easily be confused with crane—heavy machinery).
√	All pronouns have clear referents.
√	The item avoids the use of proper names. (Such names may be unfamiliar or difficult for cultural subgroups.)
√	The item avoids irregularly spelled words.
C.	**Gender Stereotypes**
Definition	The item avoids stereotyping as results of associating genders with certain professions or activities. All groups of society should be portrayed accurately and fairly regarding gender.
√	The item is free of content that might offend a gender subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a gender subgroup.
D.	**Ethnic Stereotypes**
Definition	The item avoids unnecessary references to and uses the proper reference for ethnic, racial, or cultural groups.
√	The item is free of content that might offend an ethnic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage an ethnic subgroup.
√	The artwork included in an item adequately reflects the diversity of the student population.
E.	**Cultural Familiarity**
Definition	Does not rely on an assumed shared experience that is class oriented or native English speaking oriented. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments.
√	The item does not rely on an assumed shared experience that is class oriented or native English speaking oriented.
√	The item is free from content that might offend a socioeconomic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a socioeconomic subgroup.

Universal Design Item Checklist	
√	The item is free from unnecessary cultural references.
√	The item is free from religious references.
F.	**Geographic Bias**
Definition	All groups of society should be portrayed accurately and fairly regarding geographic setting. A particular geographic setting shouldn't be used repeatedly, and urban, suburban, and rural settings should be represented across items.
√	The item is free of content that might offend a geographic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a geographic subgroup.
G.	**Disability Bias**
Definition	All groups of society should be portrayed accurately and fairly regarding disability. Stereotypes related to any particular disability should be avoided. No undue restrictions should exist in the item that would interfere with the ability of a student to comprehend or respond to the item.
√	The item is free of content that might offend a disability subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a disability subgroup.
√	A graphic representation is used in the items, as appropriate. The complexity of the graphic is appropriate to the purpose—simpler is better.
√	The item avoids content that depends on sensory knowledge (such as references to movement, sound, smell, etc.) unless this is crucial to the overall item.
√	The item could be put into Braille.
√	The item avoids using both O and Q.
√	Letter pairs can be easily distinguished when read. (S and T are okay; S and X are not).
H.	**Art Supports Text**
Definition	The art is related to the item and supports the reader when possible. The item text and art are legible and accessible, and the art is appropriately placed in the item to support the reader. The art does not distract the test taker, but instead provides a scaffold to overall comprehension.
√	All pictures relate to items.
√	The item is free from pictorial clutter: All pictures are needed to answer the item.
√	Graphics are clear and non-fuzzy.
√	Any symbols used are highly distinguishable.
√	Visual load requirements are reasonable for the grade level.
√	Multi-dimensional graphics and complex shading are avoided.
√	Tables have replaced any cluttered graphs.
√	Labels read clockwise (as is easier for Braille readers).
I.	**Special Populations Considerations**
Definition	Consideration must be given for maximum accessibility to all students including, but not limited to, English Language Learners/Multilingual Learners, limited sight, hearing impaired, cognitively challenged, etc. These considerations will assist all students.
√	The item contains scaffolding techniques to support student understanding of what is being asked in the item.
√	Text is replaced with graphic representations, when appropriate.
√	The item is written with simplified text load.

Universal Design Item Checklist	
√	The item is written with simplified sentences.
√	The item has as little extraneous information as possible.
√	The item provides context, but it is simplified.
√	The item uses smaller or less complicated numbers or expressions where not otherwise required.
√	The item avoids negative phrasing or questions; for example, questions are not asked in the negative.

Appendix E: Criteria for Item Acceptability

The following criteria represent best practices in item development, and were implemented during the creation and review of the New York State 3–8 test questions; however, these criteria are not a substitute for the full, detailed criteria documents, which are available online at the following links:

- http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests; and
- http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests.

For Multiple-Choice Items:

Check that the content of each item:

- is targeted to assess only one objective or skill (unless specifications indicate otherwise)
- deals with material that is important in testing the targeted performance indicator
- uses grade-appropriate content and thinking skills
- is presented at a reading level suitable for the grade level being tested
- has a stem that facilitates answering the question or completing the statement without looking at the answer choices
- has a stem that does **not** present clues to the correct answer choice
- has answer choices that are plausible and attractive to the student who has not mastered the objective or skill
- has mutually exclusive distractors
- has one and only one correct answer choice
- is free of cultural, racial, ethnic, age, gender, disability, regional, or other apparent bias

Check that the format of each item:

- is worded in the positive unless it is absolutely necessary to use the negative form
- is free of extraneous words or expressions in both the stem and the answer choices (e.g., the same word or phrase does not begin each answer choice)
- indicates emphasis on key words, such as best, first, least, not, and others that are important and might be overlooked
- places the interrogative word at the **beginning** of a stem in the form of a question, or places the omitted portion of an incomplete statement at the **end** of the statement
- indicates the correct answer choice
- provides the rationale for all distractors
- is conceptually, grammatically, and syntactically consistent–between the stem and answer choices, and among the answer choices
- has answer choices balanced in length, or contains two long and two short answer choices
- clearly identifies the passage or other stimulus material associated with the item
- clearly identifies a need for art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that:

- one item does not present clues to the correct answer choice for any other item
- any item based on a passage is answerable from the information given in the passage and is not dependent on skills related to other content areas
- any item based on a passage is truly passage-dependent; that is, **not** answerable without reference to the passage
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art

For Constructed-Response Items:

Check that the content of each item is:

- designed to assess the targeted performance indicator
- appropriate for the grade level being tested
- presented at a reading level suitable for the grade level being tested
- appropriate in context
- written so that a student possessing the knowledge or skill being tested can construct a response that can be scored with the specified rubric or scoring tool; that is, the range of possible correct responses must be wide enough to allow for a diversity of responses, but narrow enough so that students who do not clearly show their grasp of the objective or skill being assessed cannot obtain the maximum score
- presented without clues to the correct response
- checked for accuracy and documented against reliable, up-to-date sources (including rubrics)
- free of cultural, racial, ethnic, age, gender, disability, or other apparent bias

Check that the format of each item is:

- appropriate for the question being asked and the intended response
- worded clearly and concisely, using simple vocabulary and sentence structure
- precise and unambiguous in its directions for the desired response
- free of extraneous words or expressions
- worded in the positive form rather than in the negative form
- conceptually, grammatically, and syntactically consistent
- marked with emphasis on key words, such as best, first, least, and others that are important and might be overlooked
- clearly identified as needing art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that:

- one item does not present clues to the correct response to any other item
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art
- for each set of items related to a reading passage, each item is designed to elicit a unique and independent response
- items designed to assess reading do not depend on prior knowledge of the subject matter used in the prompt/question

Appendix F: Psychometric Guidelines for Operational Item Selection

It is primarily up to the content development department to select items for the 2019 Operational Test. The psychometrics department will provide support, as necessary, and will review the final item selection. The psychometrics department will provide data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2019 item selection included 2013–2018 embedded and stand-alone field-test items.

Here are the general guidelines for item selection:

- Satisfy the content specifications in terms of objective coverage and the number and percentage of MC and CR items on the test. An often-used criterion for objective coverage is within 5% of the percentages of score points and items per objective.
- To the extent possible, select both easy and difficult items to provide good measurement information at both ends of the performance scale.
- Avoid selecting items with too high/low *p*-values, items with flagged point-biserials, and poorly fitting items.
- Minimize the number of items flagged for DIF (gender, ethnic, and High/Low Needs schools). Flagged items should be reviewed for content again. It needs to be remembered that some items may be flagged for DIF by chance only, and that their content may not necessarily be biased against any of the analyzed subgroups. The psychometrics department will provide DIF information for each item. It is also possible to get "significant" DIF, but not bias, if the content is a necessary part of the construct that is measured. That is, there may be some non-false positive DIF flags on items that do not exhibit bias.
- Provide the NYSED with the following summary information:
    - Overview of the statistical properties of the tests
    - Blueprint comparison between the test build and the target. The focus is on the total number of points on the test
    - Raw score proportion-correct comparison between the test build and the reference (i.e., Spring 2018 test)
    - Vertical linked average difficulty parameter (MC items only) across all grades
    - Vertically linked TCC based on the constructed test
    - TCC, Test Information Curves, and Conditional SEM Curves for each subject and grade, again using the Spring 2018 operational test as a reference

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2019 NYSTP Grades 3–8 ELA and Mathematics Tests. Field test items that do not contribute to students' scores have been omitted. Additional detail on the standards to which these items align may be found at: http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards.

Table G1. ELA Grade 3 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
Session 1					
1	Multiple Choice	1	CCSS.ELA-Literacy.L.3.4	Language Standards	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.4	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.5	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.2	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.2	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.7	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.4	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.5	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.4	Reading Standards for Literature	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.6	Reading Standards for Literature	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.2	Reading Standards for Literature	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.5	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.2	Reading Standards for Literature	Reading
Session 2					
25	Constructed Response	2	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Writing to Sources
26	Constructed Response	2	CCSS.ELA-Literacy.RI.3.2	Reading Standards for Informational Text	Writing to Sources

Question	Type	Points	Standard	Strand	Subscore
27	Constructed Response	2	CCSS.ELA-Literacy.RL.3.5	Reading Standards for Literature	Writing to Sources
28	Constructed Response	2	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Writing to Sources
29	Constructed Response	2	CCSS.ELA-Literacy.RL.3.2	Reading Standards for Literature	Writing to Sources
30	Constructed Response	2	CCSS.ELA-Literacy.RI.3.4	Reading Standards for Informational Text	Writing to Sources
31	Constructed Response	4	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Writing to Sources

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2019 operational ELA test. However, each constructed-response question measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, two-point and four-point constructed-response questions require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response questions, please refer to the rubrics shown in the Educator Guides.

Table G2. ELA Grade 4 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
Session 1					
1	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.6	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.4	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.2	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.2	Reading Standards for Literature	Reading
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.4	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.3	Reading Standards for Informational Text	Reading
15	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.5	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.4	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.7	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.2	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.L.4.4	Language Standards	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.6	Reading Standards for Literature	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.2	Reading Standards for Literature	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.4	Reading Standards for Literature	Reading

Question	Type	Points	Standard	Strand	Subscore
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Reading
Session 2					
25	Constructed Response	2	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Writing to Sources
26	Constructed Response	2	CCSS.ELA-Literacy.RL.4.6	Reading Standards for Literature	Writing to Sources
27	Constructed Response	2	CCSS.ELA-Literacy.RL.4.2	Reading Standards for Literature	Writing to Sources
28	Constructed Response	2	CCSS.ELA-Literacy.RI.4.2	Reading Standards for Informational Text	Writing to Sources
29	Constructed Response	2	CCSS.ELA-Literacy.RI.4.3	Reading Standards for Informational Text	Writing to Sources
30	Constructed Response	2	CCSS.ELA-Literacy.RI.4.4	Reading Standards for Informational Text	Writing to Sources
31	Constructed Response	4	CCSS.ELA-Literacy.RI.4.9	Reading Standards for Informational Text	Writing to Sources

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2019 operational ELA test. However, each constructed-response question measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, two-point and four-point constructed-response questions require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response questions, please refer to the rubrics shown in the Educator Guides.

Table G3. ELA Grade 5 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
Session 1					
1	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.4	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.5	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.2	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.6	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.4	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.4	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.L.5.4	Language Standards	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.2	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.6	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
15	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.5	Reading Standards for Literature	Reading
16	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.4	Reading Standards for Literature	Reading
17	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
18	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.2	Reading Standards for Literature	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.2	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.L.5.5	Language Standards	Reading
31	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.6	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.2	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RI.5.2	Reading Standards for Informational Text	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RL.5.4	Reading Standards for Literature	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RL.5.5	Reading Standards for Literature	Writing to Sources
42	Constructed Response	4	CCSS.ELA-Literacy.RL.5.6	Reading Standards for Literature	Writing to Sources

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2019 operational ELA test. However, each constructed-response question measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, two-point and four-point constructed-response questions require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response questions, please refer to the rubrics shown in the Educator Guides.

Table G4. ELA Grade 6 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
Session 1					
1	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.2	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.5	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.4	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.2	Reading Standards for Literature	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.4	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.5	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.L.6.4	Language Standards	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.4	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.2	Reading Standards for Informational Text	Reading
15	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.2	Reading Standards for Literature	Reading
16	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
17	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.6	Reading Standards for Literature	Reading
18	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.4	Reading Standards for Literature	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.2	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.6	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.2	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.7	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.4	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.2	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RI.6.4	Reading Standards for Informational Text	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RI.6.2	Reading Standards for Informational Text	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RI.6.5	Reading Standards for Informational Text	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RL.6.4	Reading Standards for Literature	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RL.6.6	Reading Standards for Literature	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RL.6.5	Reading Standards for Literature	Writing to Sources
42	Constructed Response	4	CCSS.ELA-Literacy.RL.6.9	Reading Standards for Literature	Writing to Sources

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2019 operational ELA test. However, each constructed-response question measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, two-point and four-point constructed-response questions require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response questions, please refer to the rubrics shown in the Educator Guides.

Table G5. ELA Grade 7 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
Session 1					
1	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.2	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.6	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.2	Reading Standards for Literature	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.2	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.L.7.4	Language Standards	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.5	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.4	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.6	Reading Standards for Informational Text	Reading
15	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.5	Reading Standards for Literature	Reading
16	Multiple Choice	1	CCSS.ELA-Literacy.L.7.4	Language Standards	Reading
17	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Reading
18	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.6	Reading Standards for Literature	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.2	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.5	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.5	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.2	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.4	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.6	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RL.7.5	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RL.7.6	Reading Standards for Literature	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RI.7.5	Reading Standards for Informational Text	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RI.7.4	Reading Standards for Informational Text	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RI.7.6	Reading Standards for Informational Text	Writing to Sources
42	Constructed Response	2	CCSS.ELA-Literacy.RI.7.2	Reading Standards for Informational Text	Writing to Sources
43	Constructed Response	4	CCSS.ELA-Literacy.RI.7.8	Reading Standards for Informational Text	Writing to Sources

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2019 operational ELA test. However, each constructed-response question measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, two-point and four-point constructed-response questions require students to first conduct the analyses described in the mapped standard and then produce written responses

Question	Type	Points	Standard	Strand	Subscore
that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response questions, please refer to the rubrics shown in the Educator Guides.					

Table G6. ELA Grade 8 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
Session 1					
1	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.L.8.4	Language Standards	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.6	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.2	Reading Standards for Literature	Reading
15	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.8	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.5	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	CCSS.ELA-Literacy.L.8.4	Language Standards	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.3	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.2	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.6	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.4	Reading Standards for Literature	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.2	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Reading
25	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Reading
26	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.4	Reading Standards for Literature	Reading
27	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.2	Reading Standards for Literature	Reading
28	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.6	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
31	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.5	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.6	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.3	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.3	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.2	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RL.8.4	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RL.8.6	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RL.8.2	Reading Standards for Literature	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RI.8.5	Reading Standards for Informational Text	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RI.8.2	Reading Standards for Informational Text	Writing to Sources
42	Constructed Response	2	CCSS.ELA-Literacy.RI.8.3	Reading Standards for Informational Text	Writing to Sources
43	Constructed Response	4	CCSS.ELA-Literacy.RI.8.8	Reading Standards for Informational Text	Writing to Sources

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2019 operational ELA test. However, each constructed-response question measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, two-point and four-point constructed-response questions require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response questions, please refer to the rubrics shown in the Educator Guides.

Table G7. Mathematics Grade 3 Operational Item Map

Question	Type	Points	Standard	Cluster	Subscore
Session 1					
1	Multiple Choice	1	CCSS.Math.Content.3.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
2	Multiple Choice	1	CCSS.Math.Content.3.OA.D.9	Operations and Algebraic Thinking	Operations and Algebraic Thinking
3	Multiple Choice	1	CCSS.Math.Content.3.OA.A.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
5	Multiple Choice	1	CCSS.Math.Content.3.OA.D.8	Operations and Algebraic Thinking	Operations and Algebraic Thinking
6	Multiple Choice	1	CCSS.Math.Content.3.OA.A.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking

Question	Type	Points	Standard	Cluster	Subscore
7	Multiple Choice	1	CCSS.Math.Content.3.OA.D.8	Operations and Algebraic Thinking	Operations and Algebraic Thinking
8	Multiple Choice	1	CCSS.Math.Content.3.NF.A.3a	Number and Operations - Fractions	Number and Operations - Fractions
10	Multiple Choice	1	CCSS.Math.Content.3.NF.A.2a	Number and Operations - Fractions	Number and Operations - Fractions
12	Multiple Choice	1	CCSS.Math.Content.3.MD.A.2	Measurement and Data	Measurement and Data
13	Multiple Choice	1	CCSS.Math.Content.3.G.A.2	Geometry	
15	Multiple Choice	1	CCSS.Math.Content.3.MD.C.7c	Measurement and Data	Measurement and Data
16	Multiple Choice	1	CCSS.Math.Content.3.OA.B.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
18	Multiple Choice	1	CCSS.Math.Content.3.NF.A.3c	Number and Operations - Fractions	Number and Operations - Fractions
19	Multiple Choice	1	CCSS.Math.Content.3.MD.B.3	Measurement and Data	Measurement and Data
21	Multiple Choice	1	CCSS.Math.Content.3.OA.A.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
22	Multiple Choice	1	CCSS.Math.Content.3.NBT.A.1	Number and Operations in Base Ten	
23	Multiple Choice	1	CCSS.Math.Content.3.OA.D.9	Operations and Algebraic Thinking	Operations and Algebraic Thinking
24	Multiple Choice	1	CCSS.Math.Content.3.MD.C.7a	Measurement and Data	Measurement and Data
25	Multiple Choice	1	CCSS.Math.Content.3.OA.A.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
Session 2					
26	Multiple Choice	1	CCSS.Math.Content.3.NF.A.3b	Number and Operations - Fractions	Number and Operations - Fractions
27	Multiple Choice	1	CCSS.Math.Content.3.OA.A.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
28	Multiple Choice	1	CCSS.Math.Content.3.MD.C.5b	Measurement and Data	Measurement and Data
29	Multiple Choice	1	CCSS.Math.Content.3.MD.B.3	Measurement and Data	Measurement and Data
30	Multiple Choice	1	CCSS.Math.Content.3.OA.A.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
31	Multiple Choice	1	CCSS.Math.Content.3.MD.C.7d	Measurement and Data	Measurement and Data

Question	Type	Points	Standard	Cluster	Subscore
32	Multiple Choice	1	CCSS.Math.Content.3.OA.B.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
33	Multiple Choice	1	CCSS.Math.Content.3.OA.D.8	Operations and Algebraic Thinking	Operations and Algebraic Thinking
34	Constructed Response	2	CCSS.Math.Content.3.OA.B.6	Operations and Algebraic Thinking	Operations and Algebraic Thinking
35	Constructed Response	2	CCSS.Math.Content.3.NF.A.3d	Number and Operations - Fractions	Number and Operations - Fractions
36	Constructed Response	2	CCSS.Math.Content.3.MD.A.1	Measurement and Data	Measurement and Data
37	Constructed Response	2	CCSS.Math.Content.3.NBT.A.3	Number and Operations in Base Ten	
38	Constructed Response	2	CCSS.Math.Content.3.MD.C.6	Measurement and Data	Measurement and Data
39	Constructed Response	2	CCSS.Math.Content.3.NF.A.1	Number and Operations - Fractions	Number and Operations - Fractions
40	Constructed Response	3	CCSS.Math.Content.3.OA.A.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question. However, some questions measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G8. Mathematics Grade 4 Operational Item Map

Question	Type	Points	Standard	Cluster	Subscore
Session 1					
1	Multiple Choice	1	CCSS.Math.Content.4.NF.B.4c	Number and Operations - Fractions	Number and Operations - Fractions
2	Multiple Choice	1	CCSS.Math.Content.4.OA.B.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
3	Multiple Choice	1	CCSS.Math.Content.4.NF.A.2	Number and Operations - Fractions	Number and Operations - Fractions
4	Multiple Choice	1	CCSS.Math.Content.4.OA.A.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
5	Multiple Choice	1	CCSS.Math.Content.4.G.A.1	Geometry	
6	Multiple Choice	1	CCSS.Math.Content.4.OA.A.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
8	Multiple Choice	1	CCSS.Math.Content.4.NF.A.1	Number and Operations - Fractions	Number and Operations - Fractions
9	Multiple Choice	1	CCSS.Math.Content.4.MD.C.7	Measurement and Data	

Question	Type	Points	Standard	Cluster	Subscore
11	Multiple Choice	1	CCSS.Math.Content.4.NF.A.2	Number and Operations - Fractions	Number and Operations - Fractions
13	Multiple Choice	1	CCSS.Math.Content.4.NF.B.3c	Number and Operations - Fractions	Number and Operations - Fractions
15	Multiple Choice	1	CCSS.Math.Content.4.NBT.B.5	Number and Operations in Base Ten	Number and Operations in Base Ten
16	Multiple Choice	1	CCSS.Math.Content.4.NBT.B.6	Number and Operations in Base Ten	Number and Operations in Base Ten
17	Multiple Choice	1	CCSS.Math.Content.4.OA.A.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
19	Multiple Choice	1	CCSS.Math.Content.4.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
20	Multiple Choice	1	CCSS.Math.Content.4.NBT.A.3	Number and Operations in Base Ten	Number and Operations in Base Ten
21	Multiple Choice	1	CCSS.Math.Content.4.MD.A.3	Measurement and Data	
22	Multiple Choice	1	CCSS.Math.Content.4.MD.B.4	Measurement and Data	
24	Multiple Choice	1	CCSS.Math.Content.4.NBT.A.1	Number and Operations in Base Ten	Number and Operations in Base Ten
26	Multiple Choice	1	CCSS.Math.Content.4.OA.C.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
27	Multiple Choice	1	CCSS.Math.Content.4.NF.B.3a	Number and Operations - Fractions	Number and Operations - Fractions
28	Multiple Choice	1	CCSS.Math.Content.4.G.A.3	Geometry	
29	Multiple Choice	1	CCSS.Math.Content.4.MD.C.5a	Measurement and Data	
30	Multiple Choice	1	CCSS.Math.Content.4.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
Session 2					
31	Multiple Choice	1	CCSS.Math.Content.4.NF.B.4a	Number and Operations - Fractions	Number and Operations - Fractions
32	Multiple Choice	1	CCSS.Math.Content.4.OA.A.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
33	Multiple Choice	1	CCSS.Math.Content.4.NBT.B.6	Number and Operations in Base Ten	Number and Operations in Base Ten
34	Multiple Choice	1	CCSS.Math.Content.4.MD.C.5b	Measurement and Data	
35	Multiple Choice	1	CCSS.Math.Content.4.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking

Question	Type	Points	Standard	Cluster	Subscore
36	Multiple Choice	1	CCSS.Math.Content.4.MD.C.6	Measurement and Data	
37	Multiple Choice	1	CCSS.Math.Content.4.NF.A.1	Number and Operations - Fractions	Number and Operations - Fractions
38	Multiple Choice	1	CCSS.Math.Content.3.MD.B.4	Measurement and Data	
39	Constructed Response	2	CCSS.Math.Content.4.NF.B.4b	Number and Operations - Fractions	Number and Operations - Fractions
40	Constructed Response	2	CCSS.Math.Content.4.NBT.A.2	Number and Operations in Base Ten	Number and Operations in Base Ten
41	Constructed Response	2	CCSS.Math.Content.4.G.A.2	Geometry	
42	Constructed Response	2	CCSS.Math.Content.4.NF.B.3d	Number and Operations - Fractions	Number and Operations - Fractions
43	Constructed Response	2	CCSS.Math.Content.4.NBT.A.1	Number and Operations in Base Ten	Number and Operations in Base Ten
44	Constructed Response	2	CCSS.Math.Content.4.MD.A.3	Measurement and Data	
45	Constructed Response	3	CCSS.Math.Content.4.NBT.B.5	Number and Operations in Base Ten	Number and Operations in Base Ten

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question. However, some questions measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G9. Mathematics Grade 5 Operational Item Map

Question	Type	Points	Standard	Cluster	Subscore
Session 1					
1	Multiple Choice	1	CCSS.Math.Content.5.MD.C.5b	Measurement and Data	Measurement and Data
2	Multiple Choice	1	CCSS.Math.Content.4.NF.C.5	Number and Operations - Fractions	Number and Operations - Fractions
3	Multiple Choice	1	CCSS.Math.Content.5.NF.B.6	Number and Operations - Fractions	Number and Operations - Fractions
4	Multiple Choice	1	CCSS.Math.Content.4.NF.C.6	Number and Operations in Base Ten	Number and Operations in Base Ten

Question	Type	Points	Standard	Cluster	Subscore
6	Multiple Choice	1	CCSS.Math.Content.5.NBT.A.3b	Number and Operations in Base Ten	Number and Operations in Base Ten
8	Multiple Choice	1	CCSS.Math.Content.5.OA.A.2	Operations and Algebraic Thinking	
9	Multiple Choice	1	CCSS.Math.Content.5.MD.C.4	Measurement and Data	Measurement and Data
10	Multiple Choice	1	CCSS.Math.Content.5.OA.A.1	Operations and Algebraic Thinking	
12	Multiple Choice	1	CCSS.Math.Content.5.MD.A.1	Measurement and Data	Measurement and Data
13	Multiple Choice	1	CCSS.Math.Content.5.NF.B.4b	Number and Operations - Fractions	Number and Operations - Fractions
15	Multiple Choice	1	CCSS.Math.Content.4.MD.A.2	Measurement and Data	Measurement and Data
16	Multiple Choice	1	CCSS.Math.Content.5.NBT.B.6	Number and Operations in Base Ten	Number and Operations in Base Ten
18	Multiple Choice	1	CCSS.Math.Content.5.MD.C.5a	Measurement and Data	Measurement and Data
19	Multiple Choice	1	CCSS.Math.Content.5.NBT.A.4	Number and Operations in Base Ten	Number and Operations in Base Ten
20	Multiple Choice	1	CCSS.Math.Content.5.NF.B.7c	Number and Operations - Fractions	Number and Operations - Fractions
21	Multiple Choice	1	CCSS.Math.Content.5.NF.A.2	Number and Operations - Fractions	Number and Operations - Fractions
23	Multiple Choice	1	CCSS.Math.Content.5.NF.B.6	Number and Operations - Fractions	Number and Operations - Fractions
24	Multiple Choice	1	CCSS.Math.Content.5.NF.B.5a	Number and Operations - Fractions	Number and Operations - Fractions
26	Multiple Choice	1	CCSS.Math.Content.5.NF.B.7a	Number and Operations - Fractions	Number and Operations - Fractions
27	Multiple Choice	1	CCSS.Math.Content.5.NBT.B.7	Number and Operations in Base Ten	Number and Operations in Base Ten
28	Multiple Choice	1	CCSS.Math.Content.5.NBT.B.6	Number and Operations in Base Ten	Number and Operations in Base Ten
29	Multiple Choice	1	CCSS.Math.Content.5.MD.B.2	Measurement and Data	Measurement and Data
30	Multiple Choice	1	CCSS.Math.Content.5.MD.C.5c	Measurement and Data	Measurement and Data
Session 2					
31	Multiple Choice	1	CCSS.Math.Content.5.G.B.3	Geometry	

Question	Type	Points	Standard	Cluster	Subscore
32	Multiple Choice	1	CCSS.Math.Content.5.NF.A.1	Number and Operations - Fractions	Number and Operations - Fractions
33	Multiple Choice	1	CCSS.Math.Content.4.MD.A.1	Measurement and Data	Measurement and Data
34	Multiple Choice	1	CCSS.Math.Content.5.NF.A.1	Number and Operations - Fractions	Number and Operations - Fractions
35	Multiple Choice	1	CCSS.Math.Content.5.NBT.B.7	Number and Operations in Base Ten	Number and Operations in Base Ten
36	Multiple Choice	1	CCSS.Math.Content.5.NBT.A.3a	Number and Operations in Base Ten	Number and Operations in Base Ten
37	Multiple Choice	1	CCSS.Math.Content.5.NF.B.6	Number and Operations - Fractions	Number and Operations - Fractions
38	Multiple Choice	1	CCSS.Math.Content.5.NF.B.4a	Number and Operations - Fractions	Number and Operations - Fractions
39	Constructed Response	2	CCSS.Math.Content.5.MD.C.5c	Measurement and Data	Measurement and Data
40	Constructed Response	2	CCSS.Math.Content.5.NF.A.2	Number and Operations - Fractions	Number and Operations - Fractions
41	Constructed Response	2	CCSS.Math.Content.5.NBT.A.1	Number and Operations in Base Ten	Number and Operations in Base Ten
42	Constructed Response	2	CCSS.Math.Content.5.NF.B.7c	Number and Operations - Fractions	Number and Operations - Fractions
43	Constructed Response	2	CCSS.Math.Content.5.MD.A.1	Measurement and Data	Measurement and Data
44	Constructed Response	2	CCSS.Math.Content.5.NF.B.5b	Number and Operations - Fractions	Number and Operations - Fractions
45	Constructed Response	3	CCSS.Math.Content.5.NBT.B.7	Number and Operations in Base Ten	Number and Operations in Base Ten

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question. However, some questions measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G10. Mathematics Grade 6 Operational Item Map

Question	Type	Points	Standard	Cluster	Subscore
Session 1					
1	Multiple Choice	1	CCSS.Math.Content.6.G.A.3	Geometry	
2	Multiple Choice	1	CCSS.Math.Content.6.EE.A.1	Expressions and Equations	Expressions and Equations
3	Multiple Choice	1	CCSS.Math.Content.6.G.A.1	Geometry	
4	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
5	Multiple Choice	1	CCSS.Math.Content.6.EE.A.4	Expressions and Equations	Expressions and Equations
7	Multiple Choice	1	CCSS.Math.Content.6.EE.A.2a	Expressions and Equations	Expressions and Equations
8	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
10	Multiple Choice	1	CCSS.Math.Content.6.NS.A.1	The Number System	The Number System
11	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
13	Multiple Choice	1	CCSS.Math.Content.6.NS.A.1	The Number System	The Number System
14	Multiple Choice	1	CCSS.Math.Content.6.EE.A.4	Expressions and Equations	Expressions and Equations
16	Multiple Choice	1	CCSS.Math.Content.6.NS.B.4	The Number System	The Number System
17	Multiple Choice	1	CCSS.Math.Content.6.NS.C.7d	The Number System	The Number System
19	Multiple Choice	1	CCSS.Math.Content.6.RP.A.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
20	Multiple Choice	1	CCSS.Math.Content.6.EE.B.7	Expressions and Equations	Expressions and Equations
22	Multiple Choice	1	CCSS.Math.Content.6.EE.C.9	Expressions and Equations	Expressions and Equations
23	Multiple Choice	1	CCSS.Math.Content.6.NS.C.6b	The Number System	The Number System
25	Multiple Choice	1	CCSS.Math.Content.6.RP.A.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
26	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships

Question	Type	Points	Standard	Cluster	Subscore
27	Multiple Choice	1	CCSS.Math.Content.6.EE.A.2c	Expressions and Equations	Expressions and Equations
28	Multiple Choice	1	CCSS.Math.Content.6.NS.A.1	The Number System	The Number System
29	Multiple Choice	1	CCSS.Math.Content.6.RP.A.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
30	Multiple Choice	1	CCSS.Math.Content.6.EE.B.7	Expressions and Equations	Expressions and Equations
31	Multiple Choice	1	CCSS.Math.Content.6.EE.A.3	Expressions and Equations	Expressions and Equations
Session 2					
32	Multiple Choice	1	CCSS.Math.Content.6.EE.B.5	Expressions and Equations	Expressions and Equations
33	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
34	Multiple Choice	1	CCSS.Math.Content.6.EE.A.2a	Expressions and Equations	Expressions and Equations
35	Multiple Choice	1	CCSS.Math.Content.6.EE.B.5	Expressions and Equations	Expressions and Equations
36	Multiple Choice	1	CCSS.Math.Content.6.G.A.3	Geometry	
37	Multiple Choice	1	CCSS.Math.Content.6.EE.B.6	Expressions and Equations	Expressions and Equations
38	Multiple Choice	1	CCSS.Math.Content.6.G.A.4	Geometry	
39	Constructed Response	2	CCSS.Math.Content.6.G.A.1	Geometry	
40	Constructed Response	2	CCSS.Math.Content.6.NS.B.4	The Number System	The Number System
41	Constructed Response	2	CCSS.Math.Content.6.RP.A.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
42	Constructed Response	2	CCSS.Math.Content.6.EE.B.7	Expressions and Equations	Expressions and Equations
43	Constructed Response	2	CCSS.Math.Content.5.OA.B.3	Expressions and Equations	Expressions and Equations
44	Constructed Response	2	CCSS.Math.Content.6.EE.A.2a	Expressions and Equations	Expressions and Equations

Question	Type	Points	Standard	Cluster	Subscore
45	Constructed Response	2	CCSS.Math.Content.6.RP.A.3d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
46	Constructed Response	3	CCSS.Math.Content.6.EE.C.9	Expressions and Equations	Expressions and Equations

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question. However, some questions measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G11. Mathematics Grade 7 Operational Item Map

Question	Type	Points	Standard	Cluster	Subscore
Session 1					
1	Multiple Choice	1	CCSS.Math.Content.7.EE.B.4a	Expressions and Equations	Expressions and Equations
2	Multiple Choice	1	CCSS.Math.Content.7.NS.A.2d	The Number System	The Number System
3	Multiple Choice	1	CCSS.Math.Content.7.RP.A.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
4	Multiple Choice	1	CCSS.Math.Content.7.G.B.4	Geometry	
5	Multiple Choice	1	CCSS.Math.Content.7.RP.A.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
7	Multiple Choice	1	CCSS.Math.Content.7.NS.A.3	The Number System	The Number System
9	Multiple Choice	1	CCSS.Math.Content.7.RP.A.2b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
10	Multiple Choice	1	CCSS.Math.Content.7.NS.A.1a	The Number System	The Number System
12	Multiple Choice	1	CCSS.Math.Content.7.EE.B.3	Expressions and Equations	Expressions and Equations
13	Multiple Choice	1	CCSS.Math.Content.7.RP.A.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
14	Multiple Choice	1	CCSS.Math.Content.7.EE.A.2	Expressions and Equations	Expressions and Equations
15	Multiple Choice	1	CCSS.Math.Content.7.NS.A.3	The Number System	The Number System
16	Multiple Choice	1	CCSS.Math.Content.7.RP.A.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
18	Multiple Choice	1	CCSS.Math.Content.7.RP.A.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships

Question	Type	Points	Standard	Cluster	Subscore
19	Multiple Choice	1	CCSS.Math.Content.7.RP.A.2a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
20	Multiple Choice	1	CCSS.Math.Content.7.EE.A.1	Expressions and Equations	Expressions and Equations
21	Multiple Choice	1	CCSS.Math.Content.7.SP.C.7b	Statistics and Probability	
23	Multiple Choice	1	CCSS.Math.Content.7.EE.B.3	Expressions and Equations	Expressions and Equations
25	Multiple Choice	1	CCSS.Math.Content.7.SP.A.2	Statistics and Probability	
27	Multiple Choice	1	CCSS.Math.Content.7.G.A.1	Geometry	
28	Multiple Choice	1	CCSS.Math.Content.7.NS.A.1c	The Number System	The Number System
29	Multiple Choice	1	CCSS.Math.Content.7.EE.B.4b	Expressions and Equations	Expressions and Equations
30	Multiple Choice	1	CCSS.Math.Content.7.SP.A.1	Statistics and Probability	
31	Multiple Choice	1	CCSS.Math.Content.7.SP.C.6	Statistics and Probability	
32	Multiple Choice	1	CCSS.Math.Content.7.NS.A.1a	The Number System	The Number System
33	Multiple Choice	1	CCSS.Math.Content.7.EE.B.3	Expressions and Equations	Expressions and Equations
Session 2					
34	Multiple Choice	1	CCSS.Math.Content.7.RP.A.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
35	Multiple Choice	1	CCSS.Math.Content.7.EE.A.1	Expressions and Equations	Expressions and Equations
36	Multiple Choice	1	CCSS.Math.Content.6.SP.B.5c	Statistics and Probability	
37	Multiple Choice	1	CCSS.Math.Content.7.SP.C.6	Statistics and Probability	
38	Multiple Choice	1	CCSS.Math.Content.7.EE.B.4b	Expressions and Equations	Expressions and Equations
39	Multiple Choice	1	CCSS.Math.Content.7.NS.A.2c	The Number System	The Number System
40	Multiple Choice	1	CCSS.Math.Content.7.EE.B.4a	Expressions and Equations	Expressions and Equations
41	Constructed Response	2	CCSS.Math.Content.7.EE.B.3	Expressions and Equations	Expressions and Equations

Question	Type	Points	Standard	Cluster	Subscore
42	Constructed Response	2	CCSS.Math.Content.7.RP.A.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
43	Constructed Response	2	CCSS.Math.Content.7.RP.A.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
44	Constructed Response	2	CCSS.Math.Content.7.EE.B.4a	Expressions and Equations	Expressions and Equations
45	Constructed Response	2	CCSS.Math.Content.7.SP.A.2	Statistics and Probability	
46	Constructed Response	2	CCSS.Math.Content.7.NS.A.3	The Number System	The Number System
47	Constructed Response	2	CCSS.Math.Content.7.RP.A.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
48	Constructed Response	3	CCSS.Math.Content.7.EE.B.3	Expressions and Equations	Expressions and Equations

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question. However, some questions measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G12. Mathematics Grade 8 Operational Item Map

Question	Type	Points	Standard	Cluster	Subscore
Session 1					
1	Multiple Choice	1	CCSS.Math.Content.8.F.A.3	Functions	Functions
2	Multiple Choice	1	CCSS.Math.Content.8.EE.A.4	Expressions and Equations	Expressions and Equations
3	Multiple Choice	1	CCSS.Math.Content.8.F.B.4	Functions	Functions
4	Multiple Choice	1	CCSS.Math.Content.8.G.C.9	Geometry	Geometry
5	Multiple Choice	1	CCSS.Math.Content.8.EE.B.5	Expressions and Equations	Expressions and Equations
7	Multiple Choice	1	CCSS.Math.Content.8.F.A.3	Functions	Functions
8	Multiple Choice	1	CCSS.Math.Content.8.EE.B.6	Expressions and Equations	Expressions and Equations

Question	Type	Points	Standard	Cluster	Subscore
9	Multiple Choice	1	CCSS.Math.Content.8.EE.A.4	Expressions and Equations	Expressions and Equations
11	Multiple Choice	1	CCSS.Math.Content.8.G.A.2	Geometry	Geometry
12	Multiple Choice	1	CCSS.Math.Content.8.EE.C.8b	Expressions and Equations	Expressions and Equations
13	Multiple Choice	1	CCSS.Math.Content.8.EE.C.8b	Expressions and Equations	Expressions and Equations
14	Multiple Choice	1	CCSS.Math.Content.8.F.B.4	Functions	Functions
16	Multiple Choice	1	CCSS.Math.Content.8.G.A.4	Geometry	Geometry
17	Multiple Choice	1	CCSS.Math.Content.7.G.B.6	Geometry	Geometry
19	Multiple Choice	1	CCSS.Math.Content.8.EE.B.6	Expressions and Equations	Expressions and Equations
21	Multiple Choice	1	CCSS.Math.Content.8.EE.C.7a	Expressions and Equations	Expressions and Equations
22	Multiple Choice	1	CCSS.Math.Content.8.SP.A.3	Statistics and Probability	
24	Multiple Choice	1	CCSS.Math.Content.8.SP.A.2	Statistics and Probability	
25	Multiple Choice	1	CCSS.Math.Content.8.EE.B.5	Expressions and Equations	Expressions and Equations
26	Multiple Choice	1	CCSS.Math.Content.8.G.C.9	Geometry	Geometry
27	Multiple Choice	1	CCSS.Math.Content.8.EE.A.3	Expressions and Equations	Expressions and Equations
29	Multiple Choice	1	CCSS.Math.Content.8.EE.A.1	Expressions and Equations	Expressions and Equations
30	Multiple Choice	1	CCSS.Math.Content.8.F.A.2	Functions	Functions
31	Multiple Choice	1	CCSS.Math.Content.8.SP.A.1	Statistics and Probability	
32	Multiple Choice	1	CCSS.Math.Content.8.G.A.5	Geometry	Geometry
33	Multiple Choice	1	CCSS.Math.Content.8.F.B.5	Functions	Functions
Session 2					
34	Multiple Choice	1	CCSS.Math.Content.8.F.A.1	Functions	Functions

Question	Type	Points	Standard	Cluster	Subscore
35	Multiple Choice	1	CCSS.Math.Content.8.G.C.9	Geometry	Geometry
36	Multiple Choice	1	CCSS.Math.Content.8.SP.A.2	Statistics and Probability	
37	Multiple Choice	1	CCSS.Math.Content.8.EE.A.1	Expressions and Equations	Expressions and Equations
38	Multiple Choice	1	CCSS.Math.Content.8.SP.A.1	Statistics and Probability	
39	Multiple Choice	1	CCSS.Math.Content.8.G.A.3	Geometry	Geometry
40	Multiple Choice	1	CCSS.Math.Content.8.F.A.1	Functions	Functions
41	Constructed Response	2	CCSS.Math.Content.8.EE.A.4	Expressions and Equations	Expressions and Equations
42	Constructed Response	2	CCSS.Math.Content.8.F.A.3	Functions	Functions
43	Constructed Response	2	CCSS.Math.Content.8.EE.C.8b	Expressions and Equations	Expressions and Equations
44	Constructed Response	2	CCSS.Math.Content.8.F.B.4	Functions	Functions
45	Constructed Response	2	CCSS.Math.Content.8.EE.B.5	Expressions and Equations	Expressions and Equations
46	Constructed Response	2	CCSS.Math.Content.8.G.A.3	Geometry	Geometry
47	Constructed Response	2	CCSS.Math.Content.8.G.C.9	Geometry	Geometry
48	Constructed Response	3	CCSS.Math.Content.8.F.A.3	Functions	Functions

*This item map is intended to identify the primary analytic skills necessary to successfully answer each question. However, some questions measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Appendix H: ELA Short-Response Rubric

2-Point Rubric–Short Response

Score	Response Features
2 Point	The features of a 2-point response are: • Valid inferences and/or claims from the text where required by the prompt • Evidence of analysis of the text where required by the prompt • Relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Sufficient number of facts, definitions, concrete details, and/or other information from the text as required by the prompt • Complete sentences where errors do not affect readability
1 Point	The features of a 1-point response are: • A mostly literal recounting of events or details from the text as required by the prompt • Some relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Incomplete sentences or bullets
0 Point*	The features of a 0-point response are: • A response that does not address any of the requirements of the prompt or is totally inaccurate • A response that is not written in English • A response that is unintelligible or indecipherable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 1.

Appendix I: ELA Extended-Response Rubric

New York State Grade 3 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 **Essays at this level:**	**3** **Essays at this level:**	**2** **Essays at this level:**	**1** **Essays at this level:**	**0*** **Essays at this level:**
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support analysis of topics or text	W.2, R.1–9	–clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate comprehension and analysis of the text	–clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension of the text	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a confused comprehension of the text	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text	–demonstrate a lack of comprehension of the text or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided text to support analysis and reflection	W.2 R.1–8	–develop the topic with relevant, well-chosen facts, definitions, and details throughout the essay	–develop the topic with relevant facts, definitions, and details throughout the essay	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–clearly and consistently group related information together –skillfully connect ideas within categories of information using linking words and phrases – provide a concluding statement that follows clearly from the topic and information presented	–generally group related information together –connect ideas within categories of information using linking words and phrases –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt to group related information together –inconsistently connect ideas using some linking words and phrases –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization –lack the use of linking words and phrases –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 4-5 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 **Essays at this level:**	**3** **Essays at this level:**	**2** **Essays at this level:**	**1** **Essays at this level**	**0*** **Essays at this level:**
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support an analysis of topics or texts	W.2 R.1–9	– clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate insightful comprehension and analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension and analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.2 R.1–8	–develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) –sustain the use of varied, relevant evidence	–develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) –sustain the use of relevant evidence, with some lack of variety	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant –use relevant evidence with inconsistency	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–exhibit clear, purposeful organization –skillfully link ideas using grade-appropriate words and phrases –use grade-appropriate, stylistically sophisticated language and domain-specific vocabulary –provide a concluding statement that follows clearly from the topic and information presented	–exhibit clear organization –link ideas using grade-appropriate words and phrases –use grade-appropriate precise language and domain-specific vocabulary –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt at organization –inconsistently link ideas using words and phrases –inconsistently use appropriate language and domain-specific vocabulary –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack the use of linking words and phrases –use language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –exhibit no use of linking words and phrases –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 6-8 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys complex ideas and information clearly and accurately in order to support claims in an analysis of topics or texts	W.2, R.1–9	–clearly introduce a topic in a manner that is compelling and follows logically from the task and purpose –demonstrate insightful analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.2, R.1–8	–develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) –sustain the use of varied, relevant evidence	–develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) –sustain the use of relevant evidence, with some lack of variety	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant –use relevant evidence with inconsistency	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2, L.3, L.6	–exhibit clear organization, with the skillful use of appropriate and varied transitions to create a unified whole and enhance meaning –establish and maintain a formal style, using grade-appropriate, stylistically sophisticated language and domain-specific vocabulary with a notable sense of voice –provide a concluding statement or section that is compelling and follows clearly from the topic and information presented	–exhibit clear organization, with the use of appropriate transitions to create a unified whole –establish and maintain a formal style using precise language and domain-specific vocabulary –provide a concluding statement or section that follows from the topic and information presented	–exhibit some attempt at organization, with inconsistent use of transitions –establish but fail to maintain a formal style, with inconsistent use of language and domain-specific vocabulary –provide a concluding statement or section that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack a formal style, using language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement or section that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement or section
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2, L.1, L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

Appendix J: Mathematics Short-Response Rubric

2-Point Holistic Rubric

2 Points	A two-point response includes the correct solution to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task. This response: • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution and the demonstration of a thorough understanding
1 Point	A one-point response demonstrates only a partial understanding of the mathematical concepts and/or procedures in the task. This response: • correctly addresses only some elements of the task • may contain an incorrect solution but applies a mathematically appropriate process • may contain the correct solution but required work is incomplete
0 Points*	A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix K: Mathematics Extended-Response Rubric

3-Point Holistic Rubric

3 Points	A three-point response includes the correct solution(s) to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task. This response: • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution(s) and the demonstration of a thorough understanding
2 Points	A two-point response demonstrates a partial understanding of the mathematical concepts and/or procedures in the task. This response: • appropriately addresses most, but not all, aspects of the task using mathematically sound procedures • may contain an incorrect solution but provides sound procedures, reasoning, and/or explanations • may reflect some minor misunderstanding of the underlying mathematical concepts and/or procedures
1 Point	A one-point response demonstrates only a limited understanding of the mathematical concepts and/or procedures in the task. This response: • may address some elements of the task correctly but reaches an inadequate solution and/or provides reasoning that is faulty or incomplete • exhibits multiple flaws related to misunderstanding of important aspects of the task, misuse of mathematical procedures, or faulty mathematical reasoning • reflects a lack of essential understanding of the underlying mathematical concepts • may contain the correct solution(s) but required work is limited
0 Points*	A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix L: Factor Analysis Results for Select Subgroups

As described in Section 3: Validity, a principal components factor analysis was conducted on the Grades 3–8 ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: ELL/MLL, SWD, SUA, SWD/SUA students using disability accommodations, and ELL/MLL students using ELL/MLL-related accommodations (ELL/MLL & SUA). Tables L1 and L2 contain the results of factor analysis on the subpopulation data for the Grades 3–8 ELA and Mathematics Tests, respectively.

Table L.1. ELA Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL=Y	**5.39**	**21.58**	**21.58**
		1.43	5.70	27.28
		1.12	4.48	31.76
		1.03	4.11	35.87
		1.02	4.06	39.93
		1.00	4.00	43.94
SWD	All Codes	**5.98**	**23.93**	**23.93**
		1.51	6.02	29.95
		1.18	4.73	34.68
		1.00	4.01	38.69
SUA	All Codes	**5.66**	**22.63**	**22.63**
		1.55	6.18	28.81
		1.19	4.75	33.56
		1.02	4.08	37.64
SWD/SUA	SWD & SUA Codes	**5.52**	**22.09**	**22.09**
		1.55	6.20	28.30
		1.19	4.76	33.06
		1.03	4.12	37.18
ELL/MLL/ SUA	SUA & ELL/MLL Codes	**4.76**	**19.06**	**19.06**
		1.42	5.69	24.74
		1.19	4.75	29.49
		1.09	4.35	33.84
		1.08	4.30	38.14
		1.08	4.30	42.44

Table L.2. ELA Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL=Y	**4.79**	**19.17**	**19.17**
		1.47	5.88	25.05
		1.18	4.72	29.77
		1.04	4.15	33.92
		1.01	4.04	37.96
SWD	All Codes	**5.50**	**21.98**	**21.98**
		1.53	6.13	28.11
		1.18	4.71	32.82
SUA	All Codes	**5.34**	**21.37**	**21.37**
		1.55	6.21	27.57
		1.17	4.69	32.27
SWD/SUA	SWD & SUA Codes	**5.23**	**20.91**	**20.91**
		1.55	6.21	27.12
		1.17	4.68	31.79
ELL/MLL/ SUA	SUA & ELL/MLL Codes	**4.24**	**16.95**	**16.95**
		1.56	6.24	23.18
		1.17	4.68	27.86
		1.06	4.25	32.11
		1.05	4.19	36.30
		1.03	4.10	40.40
		1.01	4.04	44.44
		1.01	4.02	48.46

Table L.3. ELA Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL=Y	**5.67**	**16.19**	**16.19**
		1.58	4.53	20.72
		1.09	3.10	23.82
		1.06	3.02	26.84
		1.05	2.99	29.83
		1.04	2.98	32.81
		1.03	2.95	35.76
		1.01	2.89	38.65

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD	All Codes	**6.82**	**19.48**	**19.48**
		1.60	4.57	24.05
		1.09	3.12	27.16
		1.03	2.93	30.10
		1.02	2.91	33.01
SUA	All Codes	**6.79**	**19.39**	**19.39**
		1.61	4.61	24.00
		1.09	3.13	27.13
		1.03	2.93	30.06
		1.02	2.91	32.97
SWD/SUA	SWD & SUA Codes	**6.49**	**18.53**	**18.53**
		1.64	4.67	23.20
		1.10	3.14	26.34
		1.04	2.98	29.32
		1.03	2.94	32.26
ELL/MLL/ SUA	SUA & ELL/MLL Codes	**4.85**	**13.87**	**13.87**
		1.59	4.55	18.41
		1.19	3.39	21.80
		1.14	3.26	25.06
		1.09	3.13	28.19
		1.08	3.08	31.27
		1.07	3.05	34.32
		1.06	3.03	37.35
		1.01	2.89	40.24

Table L.4. ELA Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL=Y	**5.54**	**15.84**	**15.84**
		1.61	4.60	20.43
		1.15	3.27	23.71
		1.08	3.08	26.79
		1.07	3.05	29.83
		1.05	3.00	32.84
		1.02	2.91	35.75

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL=Y	1.01	2.89	38.64
		1.00	2.86	41.50
SWD	All Codes	**6.32**	**18.06**	**18.06**
		1.69	4.82	22.88
		1.12	3.20	26.08
		1.03	2.95	29.03
		1.02	2.91	31.94
		1.01	2.89	34.83
SUA	All Codes	**6.38**	**18.24**	**18.24**
		1.70	4.85	23.10
		1.12	3.21	26.31
		1.03	2.95	29.26
		1.02	2.90	32.16
		1.00	2.87	35.03
SWD/SUA	SWD & SUA Codes	**6.05**	**17.29**	**17.29**
		1.70	4.86	22.15
		1.13	3.23	25.38
		1.04	2.98	28.36
		1.03	2.94	31.30
		1.02	2.93	34.22
ELL/MLL/ SUA	SUA & ELL/MLL Codes	**4.45**	**12.70**	**12.70**
		1.54	4.40	17.10
		1.25	3.58	20.68
		1.15	3.29	23.97
		1.11	3.17	27.14
		1.09	3.10	30.24
		1.07	3.06	33.31
		1.05	3.00	36.30
		1.03	2.95	39.26
		1.03	2.93	42.19

Table L.5. ELA Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL/MLL	ELL/MLL=Y	**6.06**	**16.84**	**16.84**
		1.73	4.81	21.65
		1.16	3.22	24.88
		1.11	3.07	27.95
		1.07	2.96	30.90
		1.06	2.95	33.85
		1.05	2.91	36.76
		1.03	2.86	39.62
		1.01	2.79	42.41
SWD	All Codes	**6.89**	**19.15**	**19.15**
		1.83	5.07	24.22
		1.30	3.62	27.84
		1.05	2.92	30.76
		1.03	2.85	33.61
		1.01	2.81	36.42
SUA	All Codes	**7.07**	**19.63**	**19.63**
		1.80	5.01	24.64
		1.32	3.66	28.30
		1.05	2.91	31.21
		1.02	2.82	34.04
		1.00	2.78	36.82
SWD/SUA	SWD & SUA Codes	**6.76**	**18.78**	**18.78**
		1.83	5.07	23.85
		1.29	3.58	27.43
		1.05	2.93	30.36
		1.03	2.86	33.22
		1.02	2.83	36.05
ELL/MLL/ SUA	SUA & ELL/MLL Codes	**5.12**	**14.23**	**14.23**
		1.75	4.86	19.09
		1.20	3.33	22.42
		1.18	3.27	25.69
		1.11	3.08	28.76
		1.10	3.07	31.83
		1.08	3.01	34.84
		1.07	2.98	37.83

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL /SUA	SUA & ELL/MLL Codes	1.06	2.95	40.78
		1.05	2.92	43.70

Table L.6. ELA Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL =Y	**6.10**	**16.95**	**16.95**
		1.75	4.86	21.81
		1.15	3.19	25.00
		1.11	3.08	28.08
		1.06	2.94	31.02
		1.04	2.90	33.92
		1.03	2.87	36.79
		1.02	2.84	39.63
		1.00	2.79	42.42
SWD	All Codes	**6.97**	**19.35**	**19.35**
		2.04	5.66	25.01
		1.20	3.33	28.34
		1.10	3.04	31.38
		1.00	2.78	34.17
SUA	All Codes	**7.19**	**19.97**	**19.97**
		2.02	5.62	25.59
		1.21	3.37	28.96
		1.09	3.03	31.98
		1.00	2.79	34.77
SWD/SUA	SWD & SUA Codes	**6.81**	**18.92**	**18.92**
		2.02	5.62	24.54
		1.19	3.31	27.85
		1.10	3.06	30.91
		1.01	2.81	33.72
ELL/MLL /SUA	SUA & ELL/MLL Codes	**5.22**	**14.51**	**14.51**
		1.65	4.60	19.11
		1.21	3.36	22.47
		1.15	3.20	25.67

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
		1.14	3.17	28.84
		1.11	3.07	31.91
		1.09	3.02	34.93
ELL//MLL SUA	SUA & ELL/MLL Codes	1.06	2.95	37.87
		1.06	2.94	40.81
		1.04	2.88	43.70
		1.01	2.82	46.51
		1.00	2.78	49.30

Table L.7. Math Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL =Y	**8.07**	**23.75**	**23.75**
		1.62	4.76	28.51
		1.11	3.27	31.78
		1.02	3.00	34.78
SWD	All Codes	**8.69**	**25.55**	**25.55**
		1.60	4.71	30.26
		1.04	3.05	33.30
SUA	All Codes	**8.04**	**23.64**	**23.64**
		1.57	4.62	28.26
		1.05	3.09	31.34
SWD/SUA	SWD & SUA Codes	**7.86**	**23.12**	**23.12**
		1.56	4.59	27.71
		1.06	3.13	30.84
ELL//MLL SUA	SUA & ELL/MLL Codes	**7.56**	**22.23**	**22.23**
		1.60	4.71	26.94
		1.11	3.27	30.21
		1.03	3.03	33.24
		1.01	2.98	36.22

Table L.8. Math Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL/MLL	ELL/MLL =Y	**8.48**	**22.31**	**22.31**
		1.38	3.64	25.94
		1.05	2.77	28.71
		1.04	2.74	31.45
		1.02	2.69	34.14
SWD	All Codes	**9.33**	**24.54**	**24.54**
		1.38	3.63	28.17
		1.05	2.77	30.94
SUA	All Codes	**8.82**	**23.21**	**23.21**
		1.36	3.58	26.80
		1.07	2.81	29.60
		1.01	2.65	32.25
SWD/SUA	SWD & SUA Codes	**8.48**	**22.30**	**22.30**
		1.35	3.56	25.86
		1.08	2.85	28.71
		1.02	2.67	31.38
		1.00	2.63	34.01
ELL//MLL SUA	SUA & ELL/MLL Codes	**7.67**	**20.18**	**20.18**
		1.35	3.56	23.74
		1.10	2.89	26.63
		1.07	2.83	29.45
		1.04	2.73	32.19
		1.03	2.70	34.88

Table L.9. Math Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL/MLL	ELL/MLL =Y	**8.35**	**21.98**	**21.98**
		1.53	4.04	26.02
		1.18	3.11	29.13
		1.11	2.93	32.05
		1.01	2.66	34.71
SWD	All Codes	**9.41**	**24.76**	**24.76**
		1.53	4.04	28.80
		1.15	3.02	31.82

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD	All Codes	1.07	2.83	34.65
		1.00	2.64	37.30
SUA	All Codes	**9.03**	**23.75**	**23.75**
		1.50	3.96	27.71
		1.17	3.09	30.80
		1.08	2.84	33.64
		1.02	2.68	36.33
SWD/SUA	SWD & SUA Codes	**8.65**	**22.77**	**22.77**
		1.51	3.97	26.74
		1.18	3.12	29.86
		1.10	2.89	32.74
		1.03	2.71	35.45
ELL/MLL /SUA	SUA & ELL/MLL Codes	**7.63**	**20.08**	**20.08**
		1.55	4.09	24.16
		1.22	3.21	27.37
		1.16	3.06	30.43
		1.07	2.83	33.26
		1.03	2.71	35.97
		1.02	2.68	38.65

Table L.10. Math Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL =Y	**8.24**	**21.12**	**21.12**
		1.22	3.13	24.24
		1.16	2.99	27.23
		1.03	2.64	29.87
SWD	All Codes	**8.89**	**22.79**	**22.79**
		1.21	3.10	25.88
		1.15	2.94	28.82
		1.01	2.58	31.40
SUA	All Codes	**8.74**	**22.42**	**22.42**
		1.21	3.10	25.51
		1.16	2.98	28.49
		1.01	2.60	31.09

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD/SUA	SWD & SUA Codes	**8.05**	**20.64**	**20.64**
		1.22	3.12	23.76
		1.16	2.99	26.74
		1.03	2.64	29.38
ELL//MLL SUA	SUA & ELL/MLL Codes	**6.32**	**16.21**	**16.21**
		1.27	3.25	19.46
		1.20	3.07	22.53
		1.15	2.94	25.47
		1.11	2.85	28.32
		1.08	2.77	31.09
		1.04	2.67	33.76
		1.03	2.65	36.41
		1.01	2.59	39.00
		1.01	2.58	41.58

Table L.11. Math Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL/MLL =Y	**8.30**	**20.25**	**20.25**
		1.51	3.69	23.94
		1.24	3.03	26.97
		1.03	2.50	29.47
		1.02	2.49	31.96
SWD	All Codes	**8.83**	**21.54**	**21.54**
		1.48	3.61	25.15
		1.19	2.90	28.05
		1.02	2.48	30.54
SUA	All Codes	**8.87**	**21.64**	**21.64**
		1.49	3.62	25.26
		1.19	2.89	28.16
		1.01	2.47	30.63
SWD/SUA	SWD & SUA Codes	**8.06**	**19.65**	**19.65**
		1.50	3.65	23.30
		1.20	2.93	26.23

Demographic Category		Extracted Factor		
		Initial Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
		1.03	2.52	28.75
		1.02	2.48	31.23
ELL//MLL SUA	SUA & ELL/MLL Codes	**5.82**	**14.20**	**14.20**
		1.55	3.78	17.98
		1.29	3.15	21.13
		1.16	2.82	23.95
		1.14	2.79	26.74
		1.10	2.68	29.42
		1.09	2.66	32.08
		1.06	2.58	34.67
		1.03	2.52	37.18
		1.02	2.50	39.68
		1.02	2.48	42.16
		1.00	2.44	44.60

Table L.12. Math Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL/MLL	ELL/MLL =Y	**8.75**	**21.35**	**21.35**
		1.30	3.18	24.52
		1.11	2.70	27.23
		1.08	2.62	29.85
		1.04	2.54	32.40
		1.02	2.48	34.88
		1.01	2.45	37.33
SWD	All Codes	**8.14**	**19.84**	**19.84**
		1.29	3.14	22.98
		1.16	2.83	25.81
		1.07	2.62	28.43
		1.05	2.55	30.98
		1.01	2.45	33.43
SUA	All Codes	**8.45**	**20.62**	**20.62**
		1.29	3.15	23.77
		1.18	2.87	26.64
		1.05	2.57	29.21

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
		1.03	2.52	31.72
		1.02	2.49	34.21
SWD/SUA	SWD & SUA Codes	**7.90**	**19.28**	**19.28**
		1.30	3.16	22.44
		1.15	2.81	25.25
		1.07	2.61	27.87
		1.04	2.54	30.41
		1.03	2.52	32.92
		1.00	2.45	35.37
ELL//MLL SUA	SUA & ELL/MLL Codes	**5.94**	**14.48**	**14.48**
		1.33	3.24	17.72
		1.25	3.06	20.78
		1.21	2.95	23.72
		1.17	2.85	26.57
		1.15	2.81	29.38
		1.12	2.73	32.11
		1.11	2.70	34.81
		1.09	2.67	37.48
		1.06	2.59	40.07
		1.03	2.51	42.58
		1.02	2.49	45.07

Appendix M: Classical Test Theory Statistics

These tables support the classical test theory analyses described in Section 5: Operational Test Data Collection and Classical Analysis. They include item type, sample size, *p*-value, percent of omitted responses and the point-biserial of the key. Field test items that do not contribute to students' scores have been omitted.

Table M.1. ELA Grade 3 Classical Item Analysis

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	177,503	0.85	0.00	0.41
2	MC	177,503	0.83	0.00	0.36
3	MC	177,503	0.82	0.00	0.43
4	MC	177,503	0.87	0.00	0.37
5	MC	177,503	0.61	0.00	0.35
6	MC	177,503	0.61	0.00	0.41
7	MC	177,503	0.66	0.00	0.39
8	MC	177,503	0.44	0.00	0.30
9	MC	177,503	0.55	0.00	0.43
10	MC	177,503	0.56	0.00	0.47
11	MC	177,503	0.59	0.00	0.43
12	MC	177,503	0.36	0.00	0.28
19	MC	177,503	0.78	0.00	0.44
20	MC	177,503	0.48	0.00	0.29
21	MC	177,503	0.44	0.00	0.42
22	MC	177,503	0.63	0.00	0.43
23	MC	177,503	0.58	0.00	0.40
24	MC	177,503	0.43	0.00	0.37
25	CR	177,503	0.58	0.00	0.63
26	CR	177,503	0.51	1.00	0.60
27	CR	177,503	0.55	1.00	0.56
28	CR	177,503	0.64	1.00	0.64
29	CR	177,503	0.48	2.00	0.64
30	CR	177,503	0.63	1.00	0.58
31	CR	177,503	0.42	1.00	0.63

Table M.2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	180,309	0.80	0.00	0.37
2	MC	180,309	0.91	0.00	0.37
3	MC	180,309	0.79	0.00	0.38
4	MC	180,309	0.60	0.00	0.32

Item	Type	N-Count	P-Value	% Omit	PBis Key
5	MC	180,309	0.69	0.00	0.27
6	MC	180,309	0.69	0.00	0.40
13	MC	180,309	0.85	0.00	0.43
14	MC	180,309	0.51	0.00	0.25
15	MC	180,309	0.61	0.00	0.40
16	MC	180,309	0.62	0.00	0.32
17	MC	180,309	0.38	0.00	0.25
18	MC	180,309	0.54	0.00	0.36
19	MC	180,309	0.66	0.00	0.40
20	MC	180,309	0.63	0.00	0.45
21	MC	180,309	0.74	0.00	0.43
22	MC	180,309	0.36	0.00	0.31
23	MC	180,309	0.48	0.00	0.40
24	MC	180,309	0.43	0.00	0.38
25	CR	180,309	0.70	0.00	0.57
26	CR	180,309	0.65	0.00	0.56
27	CR	180,309	0.49	1.00	0.58
28	CR	180,309	0.64	1.00	0.54
29	CR	180,309	0.68	1.00	0.57
30	CR	180,309	0.65	1.00	0.59
31	CR	180,309	0.39	1.00	0.64

Table M.3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	175,383	0.77	0.00	0.46
2	MC	175,383	0.49	0.00	0.28
3	MC	175,383	0.85	0.00	0.41
4	MC	175,383	0.34	0.00	0.18
5	MC	175,383	0.87	0.00	0.40
6	MC	175,383	0.55	0.00	0.42
7	MC	175,383	0.57	0.00	0.35
15	MC	175,383	0.70	0.00	0.35
16	MC	175,383	0.82	0.00	0.47
17	MC	175,383	0.71	0.00	0.33
18	MC	175,383	0.53	0.00	0.44
19	MC	175,383	0.52	0.00	0.27
20	MC	175,383	0.65	0.00	0.46
21	MC	175,383	0.51	0.00	0.32

Item	Type	N-Count	*p*-value	% Omit	PBis Key
22	MC	175,383	0.75	0.00	0.40
23	MC	175,383	0.70	0.00	0.38
24	MC	175,383	0.61	0.00	0.40
25	MC	175,383	0.75	0.00	0.54
26	MC	175,383	0.76	0.00	0.55
27	MC	175,383	0.53	0.00	0.41
28	MC	175,383	0.59	0.00	0.43
29	MC	175,383	0.71	0.00	0.49
30	MC	175,383	0.67	0.00	0.48
31	MC	175,383	0.48	0.00	0.38
32	MC	175,383	0.53	0.00	0.39
33	MC	175,383	0.76	0.00	0.49
34	MC	175,383	0.51	0.00	0.33
35	MC	175,383	0.61	0.00	0.37
36	CR	175,383	0.68	0.00	0.55
37	CR	175,383	0.72	0.00	0.60
38	CR	175,383	0.42	1.00	0.50
39	CR	175,383	0.81	0.00	0.54
40	CR	175,383	0.75	1.00	0.60
41	CR	175,383	0.54	1.00	0.55
42	CR	175,383	0.36	1.00	0.61

Table M.4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	176,182	0.83	0.00	0.34
2	MC	176,182	0.62	0.00	0.35
3	MC	176,182	0.46	0.00	0.32
4	MC	176,182	0.74	0.00	0.34
5	MC	176,182	0.79	0.00	0.39
6	MC	176,182	0.65	0.00	0.40
7	MC	176,182	0.67	0.00	0.42
15	MC	176,182	0.61	0.00	0.31
16	MC	176,182	0.69	0.00	0.38
17	MC	176,182	0.72	0.00	0.50
18	MC	176,182	0.73	0.00	0.41
19	MC	176,182	0.48	0.00	0.32
20	MC	176,182	0.73	0.00	0.48
21	MC	176,182	0.48	0.00	0.36

Item	Type	N-Count	*p*-value	% Omit	PBis Key
22	MC	176,182	0.62	0.00	0.47
23	MC	176,182	0.78	0.00	0.45
24	MC	176,182	0.58	0.00	0.33
25	MC	176,182	0.68	0.00	0.47
26	MC	176,182	0.41	0.00	0.23
27	MC	176,182	0.67	0.00	0.42
28	MC	176,182	0.74	0.00	0.43
29	MC	176,182	0.69	0.00	0.42
30	MC	176,182	0.52	0.00	0.39
31	MC	176,182	0.81	0.00	0.51
32	MC	176,182	0.80	0.00	0.52
33	MC	176,182	0.48	0.00	0.10
34	MC	176,182	0.49	0.00	0.35
35	MC	176,182	0.57	0.00	0.30
36	CR	176,182	0.81	0.00	0.55
37	CR	176,182	0.83	0.00	0.51
38	CR	176,182	0.72	0.00	0.50
39	CR	176,182	0.78	0.00	0.60
40	CR	176,182	0.56	1.00	0.55
41	CR	176,182	0.69	1.00	0.55
42	CR	176,182	0.43	2.00	0.64

Table M.5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	159,912	0.82	0.00	0.47
2	MC	159,912	0.67	0.00	0.37
3	MC	159,912	0.73	0.00	0.44
4	MC	159,912	0.55	0.00	0.30
5	MC	159,912	0.70	0.00	0.44
6	MC	159,912	0.41	0.00	0.20
7	MC	159,912	0.87	0.00	0.39
15	MC	159,912	0.81	0.00	0.41
16	MC	159,912	0.67	0.00	0.42
17	MC	159,912	0.81	0.00	0.45
18	MC	159,912	0.83	0.00	0.47
19	MC	159,912	0.74	0.00	0.34
20	MC	159,912	0.77	0.00	0.45
21	MC	159,912	0.79	0.00	0.51

Item	Type	N-Count	*p*-value	% Omit	PBis Key
22	MC	159,912	0.59	0.00	0.32
23	MC	159,912	0.48	0.00	0.30
24	MC	159,912	0.38	0.00	0.30
25	MC	159,912	0.53	0.00	0.45
26	MC	159,912	0.57	0.00	0.45
27	MC	159,912	0.64	0.00	0.48
28	MC	159,912	0.79	0.00	0.49
29	MC	159,912	0.34	0.00	0.19
30	MC	159,912	0.48	0.00	0.35
31	MC	159,912	0.42	0.00	0.39
32	MC	159,912	0.23	0.00	0.12
33	MC	159,912	0.77	0.00	0.49
34	MC	159,912	0.48	0.00	0.25
35	MC	159,912	0.51	0.00	0.25
36	CR	159,912	0.76	0.00	0.53
37	CR	159,912	0.71	1.00	0.54
38	CR	159,912	0.61	1.00	0.57
39	CR	159,912	0.78	0.00	0.65
40	CR	159,912	0.75	0.00	0.67
41	CR	159,912	0.66	1.00	0.63
42	CR	159,912	0.71	1.00	0.66
43	CR	159,912	0.61	1.00	0.72

Table M.6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	151,966	0.70	0.00	0.34
2	MC	151,966	0.86	0.00	0.43
3	MC	151,966	0.85	0.00	0.39
4	MC	151,966	0.88	0.00	0.43
5	MC	151,966	0.53	0.00	0.32
6	MC	151,966	0.82	0.00	0.45
7	MC	151,966	0.62	0.00	0.22
8	MC	151,966	0.78	0.00	0.51
9	MC	151,966	0.44	0.00	0.34
10	MC	151,966	0.53	0.00	0.37
11	MC	151,966	0.80	0.00	0.47
12	MC	151,966	0.76	0.00	0.44
13	MC	151,966	0.77	0.00	0.46

Item	Type	N-Count	*p*-value	% Omit	PBis Key
14	MC	151,966	0.62	0.00	0.38
15	MC	151,966	0.38	0.00	0.26
16	MC	151,966	0.65	0.00	0.42
17	MC	151,966	0.82	0.00	0.44
18	MC	151,966	0.58	0.00	0.41
19	MC	151,966	0.65	0.00	0.35
20	MC	151,966	0.61	0.00	0.36
21	MC	151,966	0.61	0.00	0.47
29	MC	151,966	0.70	0.00	0.50
30	MC	151,966	0.57	0.00	0.33
31	MC	151,966	0.63	0.00	0.37
32	MC	151,966	0.59	0.00	0.32
33	MC	151,966	0.64	0.00	0.33
34	MC	151,966	0.38	0.00	0.32
35	MC	151,966	0.60	0.00	0.42
36	CR	151,966	0.80	0.00	0.53
37	CR	151,966	0.76	0.00	0.58
38	CR	151,966	0.78	1.00	0.58
39	CR	151,966	0.78	0.00	0.64
40	CR	151,966	0.70	1.00	0.60
41	CR	151,966	0.79	1.00	0.57
42	CR	151,966	0.70	1.00	0.61
43	CR	151,966	0.60	1.00	0.68

Table M.7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	181,402	0.90	0.00	0.40
2	MC	181,402	0.81	0.00	0.43
3	MC	181,402	0.81	0.00	0.50
5	MC	181,402	0.59	0.00	0.40
6	MC	181,402	0.70	0.00	0.55
7	MC	181,402	0.35	0.00	0.38
8	MC	181,402	0.59	1.00	0.46
10	MC	181,402	0.66	0.00	0.53
12	MC	181,402	0.79	0.00	0.53
13	MC	181,402	0.71	0.00	0.38
15	MC	181,402	0.56	0.00	0.48
16	MC	181,402	0.48	0.00	0.45

Item	Type	N-Count	*p*-value	% Omit	PBis Key
18	MC	181,402	0.32	0.00	0.44
19	MC	181,402	0.61	0.00	0.49
21	MC	181,402	0.70	0.00	0.45
22	MC	181,402	0.45	0.00	0.33
23	MC	181,402	0.59	0.00	0.45
24	MC	181,402	0.65	0.00	0.47
25	MC	181,402	0.90	1.00	0.41
26	MC	181,402	0.84	0.00	0.41
27	MC	181,402	0.77	0.00	0.54
28	MC	181,402	0.93	0.00	0.29
29	MC	181,402	0.67	0.00	0.58
30	MC	181,402	0.69	0.00	0.50
31	MC	181,402	0.34	0.00	0.45
32	MC	181,402	0.70	0.00	0.48
33	MC	181,402	0.48	0.00	0.48
34	CR	181,402	0.60	0.00	0.52
35	CR	181,402	0.39	1.00	0.55
36	CR	181,402	0.53	1.00	0.63
37	CR	181,402	0.49	1.00	0.61
38	CR	181,402	0.41	1.00	0.56
39	CR	181,402	0.59	1.00	0.58
40	CR	181,402	0.68	1.00	0.69

Table M.8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	182,333	0.79	0.00	0.44
2	MC	182,333	0.84	0.00	0.48
3	MC	182,333	0.68	0.00	0.58
4	MC	182,333	0.55	0.00	0.60
5	MC	182,333	0.67	0.00	0.42
6	MC	182,333	0.79	2.00	0.48
8	MC	182,333	0.60	0.00	0.57
9	MC	182,333	0.74	0.00	0.51
11	MC	182,333	0.56	0.00	0.60
13	MC	182,333	0.52	0.00	0.55
15	MC	182,333	0.71	0.00	0.48
16	MC	182,333	0.70	0.00	0.49
17	MC	182,333	0.68	0.00	0.47

Item	Type	N-Count	*p*-value	% Omit	PBis Key
19	MC	182,333	0.83	0.00	0.51
20	MC	182,333	0.64	0.00	0.50
21	MC	182,333	0.48	0.00	0.39
22	MC	182,333	0.49	0.00	0.22
24	MC	182,333	0.40	0.00	0.37
26	MC	182,333	0.50	0.00	0.50
27	MC	182,333	0.68	0.00	0.51
28	MC	182,333	0.59	0.00	0.25
29	MC	182,333	0.41	0.00	0.52
30	MC	182,333	0.63	0.00	0.47
31	MC	182,333	0.83	0.00	0.40
32	MC	182,333	0.92	0.00	0.38
33	MC	182,333	0.76	0.00	0.43
34	MC	182,333	0.63	0.00	0.53
35	MC	182,333	0.58	0.00	0.43
36	MC	182,333	0.59	0.00	0.44
37	MC	182,333	0.74	0.00	0.60
38	MC	182,333	0.58	0.00	0.40
39	CR	182,333	0.73	0.00	0.60
40	CR	182,333	0.52	1.00	0.59
41	CR	182,333	0.49	1.00	0.62
42	CR	182,333	0.61	0.00	0.56
43	CR	182,333	0.52	1.00	0.59
44	CR	182,333	0.57	0.00	0.67
45	CR	182,333	0.44	1.00	0.64

Table M.9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	170,958	0.86	0.00	0.48
2	MC	170,958	0.68	0.00	0.60
3	MC	170,958	0.39	0.00	0.41
4	MC	170,958	0.69	0.00	0.42
6	MC	170,958	0.65	0.00	0.55
8	MC	170,958	0.76	0.00	0.46
9	MC	170,958	0.78	0.00	0.53
10	MC	170,958	0.63	0.00	0.27
12	MC	170,958	0.49	0.00	0.54
13	MC	170,958	0.35	0.00	0.52

Item	Type	N-Count	*p*-value	% Omit	PBis Key
15	MC	170,958	0.62	0.00	0.42
16	MC	170,958	0.68	0.00	0.52
18	MC	170,958	0.61	0.00	0.48
19	MC	170,958	0.74	0.00	0.43
20	MC	170,958	0.53	0.00	0.37
21	MC	170,958	0.48	0.00	0.46
23	MC	170,958	0.48	0.00	0.50
24	MC	170,958	0.62	0.00	0.45
26	MC	170,958	0.70	0.00	0.44
27	MC	170,958	0.40	0.00	0.50
28	MC	170,958	0.77	0.00	0.49
29	MC	170,958	0.59	0.00	0.53
30	MC	170,958	0.71	1.00	0.53
31	MC	170,958	0.71	0.00	0.42
32	MC	170,958	0.74	0.00	0.60
33	MC	170,958	0.41	0.00	0.40
34	MC	170,958	0.74	0.00	0.57
35	MC	170,958	0.74	0.00	0.53
36	MC	170,958	0.67	0.00	0.45
37	MC	170,958	0.64	0.00	0.46
38	MC	170,958	0.49	0.00	0.39
39	CR	170,958	0.39	0.00	0.49
40	CR	170,958	0.52	0.00	0.74
41	CR	170,958	0.52	0.00	0.61
42	CR	170,958	0.51	0.00	0.61
43	CR	170,958	0.45	1.00	0.67
44	CR	170,958	0.47	1.00	0.64
45	CR	170,958	0.78	0.00	0.57

Table M.10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	169,427	0.76	0.00	0.44
2	MC	169,427	0.71	0.00	0.50
3	MC	169,427	0.62	0.00	0.52
4	MC	169,427	0.64	0.00	0.57
5	MC	169,427	0.65	1.00	0.50
7	MC	169,427	0.53	0.00	0.36
8	MC	169,427	0.69	0.00	0.56

Item	Type	N-Count	*p*-value	% Omit	PBis Key
10	MC	169,427	0.72	0.00	0.53
11	MC	169,427	0.49	0.00	0.40
13	MC	169,427	0.56	0.00	0.43
14	MC	169,427	0.43	0.00	0.54
16	MC	169,427	0.61	0.00	0.43
17	MC	169,427	0.60	0.00	0.52
19	MC	169,427	0.66	0.00	0.49
20	MC	169,427	0.62	0.00	0.60
22	MC	169,427	0.54	0.00	0.53
23	MC	169,427	0.41	0.00	0.39
25	MC	169,427	0.76	0.00	0.50
26	MC	169,427	0.70	0.00	0.44
27	MC	169,427	0.42	0.00	0.56
28	MC	169,427	0.65	0.00	0.54
29	MC	169,427	0.62	0.00	0.62
30	MC	169,427	0.70	0.00	0.47
31	MC	169,427	0.61	0.00	0.52
32	MC	169,427	0.57	0.00	0.41
33	MC	169,427	0.72	0.00	0.49
34	MC	169,427	0.63	0.00	0.52
35	MC	169,427	0.61	0.00	0.54
36	MC	169,427	0.60	0.00	0.49
37	MC	169,427	0.36	0.00	0.33
38	MC	169,427	0.51	0.00	0.52
39	CR	169,427	0.34	0.00	0.67
40	CR	169,427	0.63	1.00	0.64
41	CR	169,427	0.46	1.00	0.63
42	CR	169,427	0.41	1.00	0.60
43	CR	169,427	0.35	1.00	0.57
44	CR	169,427	0.49	1.00	0.68
45	CR	169,427	0.21	1.00	0.56
46	CR	169,427	0.54	1.00	0.72

Table M.11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	159,915	0.81	0.00	0.48
2	MC	159,915	0.84	0.00	0.41
3	MC	159,915	0.66	0.00	0.57

Item	Type	N-Count	*p*-value	% Omit	PBis Key
4	MC	159,915	0.47	0.00	0.50
5	MC	159,915	0.49	0.00	0.38
7	MC	159,915	0.54	0.00	0.30
9	MC	159,915	0.42	0.00	0.49
10	MC	159,915	0.72	0.00	0.47
12	MC	159,915	0.58	0.00	0.53
13	MC	159,915	0.56	0.00	0.58
14	MC	159,915	0.36	0.00	0.38
15	MC	159,915	0.64	0.00	0.58
16	MC	159,915	0.65	0.00	0.54
18	MC	159,915	0.46	0.00	0.52
19	MC	159,915	0.80	0.00	0.40
20	MC	159,915	0.68	0.00	0.48
21	MC	159,915	0.51	0.00	0.58
23	MC	159,915	0.54	0.00	0.37
25	MC	159,915	0.56	0.00	0.53
27	MC	159,915	0.53	0.00	0.62
28	MC	159,915	0.49	0.00	0.37
29	MC	159,915	0.58	0.00	0.51
30	MC	159,915	0.59	0.00	0.36
31	MC	159,915	0.51	0.00	0.51
32	MC	159,915	0.57	0.00	0.45
33	MC	159,915	0.56	0.00	0.43
34	MC	159,915	0.78	0.00	0.28
35	MC	159,915	0.68	0.00	0.45
36	MC	159,915	0.54	0.00	0.50
37	MC	159,915	0.63	0.00	0.54
38	MC	159,915	0.64	0.00	0.50
39	MC	159,915	0.66	0.00	0.51
40	MC	159,915	0.44	0.00	0.48
41	CR	159,915	0.62	1.00	0.66
42	CR	159,915	0.40	1.00	0.73
43	CR	159,915	0.52	1.00	0.59
44	CR	159,915	0.69	1.00	0.63
45	CR	159,915	0.46	1.00	0.76
46	CR	159,915	0.61	1.00	0.67
47	CR	159,915	0.36	2.00	0.70
48	CR	159,915	0.53	1.00	0.77

Table M.12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	107,089	0.61	0.00	0.33
2	MC	107,089	0.58	0.00	0.37
3	MC	107,089	0.65	0.00	0.48
4	MC	107,089	0.65	0.00	0.37
5	MC	107,089	0.66	0.00	0.51
7	MC	107,089	0.56	0.00	0.39
8	MC	107,089	0.56	0.00	0.46
9	MC	107,089	0.58	0.00	0.37
11	MC	107,089	0.54	0.00	0.43
12	MC	107,089	0.38	0.00	0.36
13	MC	107,089	0.44	0.00	0.50
14	MC	107,089	0.47	0.00	0.41
16	MC	107,089	0.56	0.00	0.31
17	MC	107,089	0.59	0.00	0.51
19	MC	107,089	0.40	0.00	0.43
21	MC	107,089	0.52	0.00	0.51
22	MC	107,089	0.60	0.00	0.56
24	MC	107,089	0.51	0.00	0.48
25	MC	107,089	0.47	0.00	0.61
26	MC	107,089	0.57	0.00	0.62
27	MC	107,089	0.42	0.00	0.34
29	MC	107,089	0.51	0.00	0.41
30	MC	107,089	0.43	0.00	0.48
31	MC	107,089	0.70	0.00	0.36
32	MC	107,089	0.53	0.00	0.43
33	MC	107,089	0.50	0.00	0.41
34	MC	107,089	0.60	0.00	0.41
35	MC	107,089	0.57	0.00	0.51
36	MC	107,089	0.69	0.00	0.39
37	MC	107,089	0.60	0.00	0.43
38	MC	107,089	0.69	0.00	0.43
39	MC	107,089	0.48	0.00	0.47
40	MC	107,089	0.59	0.00	0.45
41	CR	107,089	0.45	1.00	0.56
42	CR	107,089	0.53	2.00	0.59
43	CR	107,089	0.34	2.00	0.67

Item	Type	N-Count	*p*-value	% Omit	PBis Key
44	CR	107,089	0.41	2.00	0.72
45	CR	107,089	0.44	2.00	0.68
46	CR	107,089	0.26	3.00	0.60
47	CR	107,089	0.54	2.00	0.68
48	CR	107,089	0.44	2.00	0.57

Appendix N: Items Flagged for DIF

These tables support the DIF information in Section 5: Operational Test Data Collection and Classical Analysis. They include item numbers, focal group, and directions of DIF and DIF statistics. Tables N.1–N.4 show items flagged by the SMD, or Mantel-Haenszel methods. Positive values of SMD and Delta in Tables N.1–N.4 indicate DIF in favor of a focal group, and negative values of SMD and Delta indicate DIF against a focal group. Field test items that do not contribute to students' scores have been omitted.

Table N.1. ELA MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
03	1	African American	Against	1.79	751.3	-1.36
03	1	Hispanic	Against	2.17	1,816.3	-1.82
03	1	Asian	Against	2.48	981.6	-2.13
03	1	High Needs	Against	2.23	2,028.2	-1.88
03	1	ELL	Against	2.51	2,758.0	-2.16
03	1	CBT	In Favor	0.53	802.8	1.50
03	3	Hispanic	Against	1.60	785.4	-1.10
03	3	Asian	Against	1.73	397.4	-1.29
03	3	ELL	Against	1.65	757.5	-1.17
03	4	Hispanic	Against	1.71	828.3	-1.27
03	4	Asian	Against	1.79	358.1	-1.37
03	4	High Needs	Against	1.65	770.4	-1.18
03	4	ELL	Against	1.80	931.4	-1.38
03	19	ELL	Against	1.67	877.7	-1.21
03	22	CBT	In Favor	0.64	842.9	1.05
03	24	Asian	Against	1.68	738.2	-1.21
04	13	African American	Against	1.57	462.3	-1.05
04	13	Hispanic	Against	1.83	1,104.0	-1.42
04	13	High Needs	Against	1.81	1,122.5	-1.40
04	13	ELL	Against	2.04	1,450.6	-1.67
04	13	CBT	In Favor	0.65	418.9	1.02
04	16	ELL	Against	1.56	676.6	-1.05
04	19	African American	Against	1.60	889.2	-1.10
04	19	Hispanic	Against	2.09	3,033.1	-1.73
04	19	Asian	Against	2.41	1,944.1	-2.06
04	19	High Needs	Against	1.83	2,390.9	-1.42
04	19	ELL	Against	2.15	1,887.2	-1.80

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
04	22	Asian	Against	1.57	615.5	-1.06
05	6	African American	Against	1.55	845.9	-1.04
05	6	Hispanic	Against	1.59	1,258.3	-1.09
05	6	Asian	Against	1.72	779.3	-1.28
05	9	Hispanic	Against	1.73	976.2	-1.29
05	9	High Needs	Against	1.75	1,078.1	-1.32
05	9	ELL	Against	2.09	1,405.6	-1.73
05	9	CBT	In Favor	0.53	1,063.3	1.50
05	16	ELL	Against	1.59	623.5	-1.10
06	2	Hispanic	Against	1.80	2,050.9	-1.38
06	2	Asian	Against	2.08	1,493.8	-1.73
06	2	High Needs	Against	1.61	1,586.5	-1.13
06	2	ELL	Against	2.17	1,586.9	-1.82
06	2	CBT	In Favor	0.59	1,509.3	1.25
06	13	ELL	Against	1.68	721.8	-1.22
06	15	ELL	Against	1.61	555.3	-1.12
06	33	CBT	In Favor	0.58	1,981.1	1.28
06	34	Female	Against	1.53	1,666.5	-1.00
07	1	Hispanic	Against	1.78	1,024.4	-1.36
07	1	High Needs	Against	1.78	1,114.6	-1.35
07	1	ELL	Against	2.29	1,487.9	-1.95
07	3	ELL	Against	1.69	587.9	-1.23
07	5	ELL	Against	1.64	513.7	-1.16
07	9	African American	Against	1.58	766.2	-1.07
07	9	Hispanic	Against	1.59	1,050.7	-1.09
07	9	High Needs	Against	1.54	1,068.2	-1.01
07	11	ELL	Against	1.55	380.3	-1.02
07	12	ELL	Against	1.57	451.2	-1.06
07	16	African American	Against	1.67	1,117.6	-1.21
07	16	Hispanic	Against	1.57	1,182.3	-1.06
07	16	Asian	Against	1.70	840.7	-1.24
07	16	High Needs	Against	1.59	1,625.8	-1.10
07	33	Asian	Against	1.61	302.9	-1.11
07	33	High Needs	Against	1.66	1,044.7	-1.20
07	33	ELL	Against	1.55	409.3	-1.02

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
07	33	CBT	In Favor	0.65	592.9	1.02
08	2	African American	Against	1.58	390.4	-1.08
08	2	Hispanic	Against	1.95	1,041.4	-1.57
08	2	Asian	Against	2.10	537.6	-1.75
08	2	High Needs	Against	1.80	853.4	-1.38
08	2	ELL	Against	2.01	954.3	-1.64
08	3	ELL	Against	1.64	463.6	-1.17
08	4	ELL	Against	1.58	336.0	-1.07
08	5	Female	Against	1.56	1,554.8	-1.04
08	5	African American	Against	1.71	1,222.8	-1.25
08	5	Hispanic	Against	1.77	1,750.1	-1.34
08	5	Asian	Against	1.69	730.4	-1.23
08	5	High Needs	Against	1.74	2,046.1	-1.31
08	5	ELL	Against	1.79	595.2	-1.37
08	6	African American	Against	2.85	2,505.0	-2.46
08	6	Hispanic	Against	2.43	2,027.0	-2.09
08	6	Asian	Against	2.74	1,166.1	-2.37
08	6	High Needs	Against	3.03	3,234.6	-2.61
08	6	ELL	Against	2.21	1,322.6	-1.86
08	6	CBT	In Favor	0.47	1,129.1	1.76
08	18	ELL	Against	1.77	665.6	-1.35
08	20	ELL	Against	1.71	575.1	-1.26
08	29	African American	Against	1.88	1,278.8	-1.48
08	29	Hispanic	Against	1.69	1,045.7	-1.23
08	29	Asian	Against	1.73	443.4	-1.29
08	29	High Needs	Against	2.05	2,194.0	-1.69
08	29	ELL	Against	1.74	585.1	-1.30
08	29	CBT	In Favor	0.65	609.8	1.02

Table N.2. ELA CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect
03	26	Asian	In Favor	0.1226	0.173
03	27	Asian	In Favor	0.1285	0.204
03	31	Asian	In Favor	0.3682	0.357
03	31	High Needs	In Favor	0.2109	0.207
03	31	CBT	Against	-0.3984	0.389
04	28	Asian	In Favor	0.1227	0.195
04	31	Asian	In Favor	0.3477	0.294
04	31	CBT	Against	-0.3105	0.267
05	42	Female	In Favor	0.2331	0.210
05	42	Asian	In Favor	0.3509	0.310
05	42	CBT	Against	-0.3543	0.319
06	37	Hispanic	In Favor	0.0972	0.175
06	40	High Needs	In Favor	0.1408	0.183
06	42	Female	In Favor	0.2070	0.182
06	42	Asian	In Favor	0.3187	0.274
06	42	High Needs	In Favor	0.2267	0.199
06	42	CBT	Against	-0.3134	0.275
07	36	ELL	Against	-0.1205	0.197
08	39	High Needs	In Favor	0.1041	0.170
08	40	Asian	In Favor	0.1180	0.183
08	40	High Needs	In Favor	0.1530	0.231
08	41	Asian	In Favor	0.1002	0.172
08	41	High Needs	In Favor	0.1251	0.207
08	42	Asian	In Favor	0.1364	0.209
08	42	High Needs	In Favor	0.1540	0.229
08	43	Female	In Favor	0.2131	0.190
08	43	Asian	In Favor	0.3320	0.294
08	43	High Needs	In Favor	0.2633	0.231
08	43	CBT	Against	-0.3114	0.277

Table N.3. Mathematics MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
03	1	Asian	In Favor	0.64	94.7	1.05
03	2	African American	Against	1.54	547.1	-1.02
03	25	CBT	In Favor	0.62	281.4	1.14
03	29	African American	Against	1.62	755.4	-1.14
03	29	Hispanic	Against	1.55	857.0	-1.02
04	36	African American	Against	1.63	1,046.1	-1.16
04	36	CBT	In Favor	0.64	707.2	1.04
05	4	Female	Against	1.55	1,411.7	-1.03
05	19	African American	Against	1.75	1,054.7	-1.32
05	19	Hispanic	Against	1.78	1,433.6	-1.36
05	19	High Needs	Against	1.69	1,333.5	-1.24
05	19	ELL	Against	1.55	550.7	-1.03
06	19	Female	Against	1.57	1,436.6	-1.05
06	25	Asian	In Favor	0.65	197.7	1.02
06	34	ELL	Against	1.62	610.7	-1.14
07	10	Female	Against	1.72	1,780.2	-1.28
07	10	African American	Against	1.69	892.6	-1.23
07	10	Hispanic	Against	1.64	983.4	-1.16
07	10	Asian	Against	1.74	466.3	-1.30
07	10	High Needs	Against	1.59	935.2	-1.09
07	23	Asian	In Favor	0.65	462.6	1.02
07	32	African American	Against	1.67	1,029.3	-1.20
08	3	Female	Against	1.60	999.1	-1.11
08	31	African American	Against	1.56	505.9	-1.04
08	31	High Needs	Against	1.55	617.4	-1.02
08	31	CBT	In Favor	0.65	382.0	1.01
08	36	African American	Against	1.76	816.7	-1.33
08	36	Hispanic	Against	1.65	762.3	-1.17
08	36	Asian	Against	1.68	293.3	-1.22
08	36	High Needs	Against	1.72	937.7	-1.27
08	36	CBT	In Favor	0.63	434.2	1.07

Table N.4. Mathematics CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect
03	38	CBT	Against	-0.1546	0.191
03	39	CBT	Against	-0.3787	0.421
04	40	Hispanic	Against	-0.1626	0.194
04	40	ELL	Against	-0.1969	0.235
04	42	CBT	Against	-0.1369	0.179
04	43	Asian	In Favor	0.1333	0.173
04	43	CBT	Against	-0.1567	0.198
04	45	Asian	In Favor	0.1926	0.184
05	41	ELL	Against	-0.1513	0.176
05	45	ELL	Against	-0.2482	0.226
06	43	ELL	Against	-0.1443	0.186
06	46	ELL	Against	-0.2576	0.228
08	41	ELL	Against	-0.1629	0.195

Appendix O: IRT Statistics

Field test items that do not contribute to students' scores have been omitted.

Table O.1. ELA Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.999	0.053	0.222		
2	1	1.141	0.577	0.171		
3	1	0.884	-0.212	0.185		
4	1	1.089	0.706	0.200		
5	1	0.878	0.129	0.228		
6	1	0.506	0.499	0.129		
7	1	0.879	-1.010	0.139		
8	1	0.757	-0.345	0.199		
9	1	0.663	0.807	0.175		
10	1	1.186	0.265	0.244		
11	1	0.810	1.136	0.177		
12	1	0.886	-0.054	0.102		
13	1	0.875	-1.536	0.064		
14	1	0.697	-1.595	0.066		
15	1	0.969	-1.123	0.197		
16	1	0.793	-1.822	0.024		
17	1	0.625	-0.171	0.165		
18	1	0.752	-0.194	0.145		
19	2	1.084	-1.338	1.470		
20	2	0.947	-2.355	1.317		
21	2	1.169	-2.954	0.453		
22	2	0.997	-3.076	0.801		
23	2	1.055	-2.247	0.878		
24	2	0.924	-1.480	1.157		
25	4	0.799	-2.156	-0.265	1.333	2.864

Table O.2. ELA Grade 4 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	1.287	0.464	0.225		
2	1	0.670	0.061	0.137		
3	1	1.213	0.622	0.191		
4	1	0.930	0.974	0.169		
5	1	0.877	-0.809	0.184		
6	1	0.700	-1.367	0.090		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
7	1	0.926	-2.002	0.051		
8	1	0.470	1.094	0.112		
9	1	0.516	-0.473	0.123		
10	1	0.438	0.475	0.167		
11	1	0.767	-0.231	0.148		
12	1	0.399	-1.338	0.014		
13	1	1.049	-1.328	0.186		
14	1	0.647	-0.739	0.058		
15	1	0.682	-1.486	0.059		
16	1	0.484	-0.506	0.064		
17	1	0.943	-0.251	0.179		
18	1	0.678	-0.842	0.074		
19	2	0.842	-2.362	0.177		
20	2	0.956	-1.707	1.512		
21	2	0.946	-3.180	-0.101		
22	2	1.025	-3.203	0.402		
23	2	0.885	-2.877	0.505		
24	2	0.991	-3.354	0.142		
25	4	0.760	-1.169	-0.151	1.074	2.377

Table O.3. ELA Grade 5 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.649	0.845	0.249		
2	1	0.977	-0.823	0.233		
3	1	0.857	-0.013	0.188		
4	1	0.374	0.063	0.066		
5	1	0.835	0.013	0.232		
6	1	0.726	0.209	0.157		
7	1	0.693	0.498	0.210		
8	1	1.233	-0.394	0.264		
9	1	0.991	0.216	0.180		
10	1	0.839	-0.400	0.151		
11	1	0.789	0.465	0.165		
12	1	0.574	-0.310	0.121		
13	1	0.556	0.404	0.162		
14	1	1.620	-0.642	0.246		
15	1	1.451	-0.608	0.227		
16	1	0.590	-0.913	0.109		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
17	1	1.048	-0.295	0.224		
18	1	1.159	-0.651	0.264		
19	1	0.687	-0.980	0.163		
20	1	0.831	0.210	0.173		
21	1	0.545	-0.851	0.140		
22	1	1.017	-1.229	0.167		
23	1	0.877	-1.525	0.241		
24	1	0.461	-1.313	0.025		
25	1	0.344	1.990	0.120		
26	1	1.101	0.289	0.236		
27	1	0.971	-1.235	0.318		
28	1	0.679	0.227	0.227		
29	2	0.665	-2.153	-0.126		
30	2	0.507	0.107	0.496		
31	2	0.816	-2.605	-0.469		
32	2	0.807	-3.416	-1.163		
33	2	0.878	-3.003	-0.709		
34	2	0.645	-1.227	0.701		
35	4	0.599	-0.966	0.165	1.308	2.398

Table O.4. ELA Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.882	0.345	0.212		
2	1	0.671	-0.004	0.275		
3	1	0.959	-0.690	0.222		
4	1	0.784	-0.816	0.219		
5	1	0.649	-1.330	0.167		
6	1	0.521	-1.123	0.179		
7	1	0.938	-0.228	0.186		
8	1	0.805	-0.364	0.228		
9	1	0.899	-0.934	0.251		
10	1	0.536	0.581	0.138		
11	1	0.651	-1.356	0.330		
12	1	0.793	-0.230	0.249		
13	1	0.899	-0.510	0.196		
14	1	0.799	1.304	0.258		
15	1	0.590	0.095	0.217		
16	1	0.680	-0.654	0.137		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
17	1	0.843	-0.397	0.257		
18	1	0.449	-0.022	0.159		
19	1	1.326	-0.878	0.268		
20	1	0.113	0.776	0.046		
21	1	0.979	0.641	0.253		
22	1	1.363	-0.917	0.284		
23	1	0.769	0.522	0.189		
24	1	0.721	-0.363	0.280		
25	1	0.657	0.597	0.199		
26	1	0.547	0.056	0.259		
27	1	0.943	-0.418	0.335		
28	1	1.184	-0.520	0.258		
29	2	0.665	-2.271	-0.210		
30	2	0.591	-1.075	0.301		
31	2	0.878	-3.247	-1.113		
32	2	0.783	-3.794	-1.114		
33	2	0.721	-3.753	-1.258		
34	2	0.616	-2.799	-0.225		
35	4	0.622	-1.446	-0.379	0.978	2.182

Table O.5. ELA Grade 7 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.825	0.643	0.208		
2	1	1.045	-1.104	0.230		
3	1	1.627	0.385	0.245		
4	1	0.503	-0.119	0.156		
5	1	0.703	1.059	0.290		
6	1	1.321	-0.001	0.276		
7	1	1.253	1.087	0.224		
8	1	0.746	-1.961	0.038		
9	1	1.639	1.816	0.177		
10	1	1.082	-0.749	0.237		
11	1	0.685	-0.279	0.249		
12	1	0.977	-1.127	0.199		
13	1	1.496	0.291	0.281		
14	1	0.784	-1.207	0.201		
15	1	0.796	-1.388	0.076		
16	1	1.077	-0.967	0.169		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
17	1	0.799	-1.046	0.126		
18	1	0.612	0.913	0.282		
19	1	1.182	1.526	0.250		
20	1	0.497	1.595	0.229		
21	1	1.229	0.713	0.187		
22	1	1.126	-0.820	0.249		
23	1	0.779	-0.802	0.166		
24	1	0.573	0.410	0.230		
25	1	0.899	-0.218	0.265		
26	1	0.773	-0.665	0.155		
27	1	0.503	-1.261	0.095		
28	1	0.613	0.727	0.197		
29	1	0.825	0.643	0.208		
30	2	1.191	-3.373	-0.776		
31	2	0.954	-2.453	0.054		
32	2	1.168	-3.950	-0.974		
33	2	1.072	-2.725	-0.471		
34	2	0.740	-1.898	0.466		
35	2	0.717	-2.789	-0.151		
36	2	0.724	-3.018	-0.577		
37	4	0.860	-2.669	-1.636	-0.078	1.434

Table O.6. ELA Grade 8 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	1.510	0.922	0.210		
2	1	0.540	-0.424	0.149		
3	1	0.964	-1.287	0.252		
4	1	0.552	-0.741	0.167		
5	1	0.997	0.192	0.239		
6	1	0.587	-0.393	0.165		
7	1	0.891	-1.169	0.178		
8	1	0.536	-0.419	0.069		
9	1	0.890	1.133	0.200		
10	1	0.852	-0.209	0.217		
11	1	1.041	-0.076	0.184		
12	1	0.864	-1.214	0.327		
13	1	0.998	0.764	0.201		
14	1	0.961	-1.595	0.165		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
15	1	0.673	0.714	0.426		
16	1	0.550	-0.577	0.030		
17	1	0.616	0.094	0.093		
18	1	1.103	-0.769	0.280		
19	1	0.749	0.512	0.248		
20	1	0.985	-0.753	0.229		
21	1	1.198	-0.757	0.218		
22	1	1.014	-0.959	0.276		
23	1	0.875	-0.765	0.207		
24	1	0.885	-0.013	0.197		
25	1	1.317	-0.284	0.266		
26	1	0.691	-0.161	0.197		
27	1	0.673	0.234	0.263		
28	1	0.652	0.226	0.221		
29	2	0.953	-2.778	-0.150		
30	2	0.952	-2.855	-0.099		
31	2	1.193	-4.028	-0.875		
32	2	0.917	-3.344	-0.872		
33	2	0.933	-3.351	-0.816		
34	2	0.924	-3.436	-0.561		
35	2	0.840	-3.772	-0.821		
36	4	0.812	-2.485	-1.591	0.045	1.529

Table O.7. Mathematics Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	1.182	0.143	0.178	
2	1	1.361	-0.652	0.469	
3	1	1.458	0.882	0.151	
4	1	0.995	-0.246	0.179	
5	1	1.078	-1.784	0.023	
6	1	1.369	-0.394	0.096	
7	1	1.260	-0.329	0.222	
8	1	1.452	0.196	0.270	
9	1	1.257	0.385	0.150	
10	1	1.032	-0.367	0.234	
11	1	1.167	-0.611	0.053	
12	1	1.418	-0.733	0.163	
13	1	1.430	0.809	0.090	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
14	1	1.584	0.753	0.112	
15	1	1.427	-0.204	0.211	
16	1	1.297	-0.272	0.275	
17	1	1.318	-0.892	0.182	
18	1	0.963	-1.007	0.200	
19	1	0.888	0.112	0.232	
20	1	1.081	0.803	0.235	
21	1	0.930	-0.039	0.154	
22	1	1.134	-1.729	0.022	
23	1	1.228	-0.039	0.200	
24	1	1.442	0.474	0.198	
25	1	1.233	0.047	0.434	
26	1	0.813	-2.323	0.076	
27	1	1.450	-0.784	0.190	
28	2	0.705	0.633	-0.500	
29	2	0.814	-0.618	0.345	
30	2	0.594	0.311	-0.988	
31	2	0.643	0.455	0.521	
32	2	0.564	-0.742	-0.113	
33	2	0.669	0.064	0.742	
34	3	0.738	-1.199	0.370	-1.525

Table O.8. Mathematics Grade 4 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	1.110	0.345	0.292	
2	1	1.398	-0.932	0.294	
3	1	1.820	0.076	0.147	
4	1	1.405	0.369	0.172	
5	1	1.040	0.629	0.216	
6	1	2.077	0.992	0.374	
7	1	1.611	-0.608	0.126	
8	1	1.265	-1.662	0.247	
9	1	0.968	-0.361	0.199	
10	1	1.464	-0.046	0.220	
11	1	1.228	-0.683	0.273	
12	1	1.279	-1.080	0.135	
13	1	1.429	-0.392	0.129	
14	1	1.170	-0.337	0.241	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
15	1	1.100	-0.354	0.174	
16	1	0.925	-0.962	0.185	
17	1	1.481	0.028	0.094	
18	1	1.196	0.028	0.266	
19	1	1.189	-0.196	0.067	
20	1	0.898	-0.155	0.260	
21	1	1.256	0.322	0.290	
22	1	0.820	-1.295	0.160	
23	1	0.365	-0.265	0.099	
24	1	0.942	-0.600	0.284	
25	1	1.177	-0.556	0.202	
26	1	0.893	-0.650	0.115	
27	1	0.975	0.859	0.164	
28	1	1.198	-0.072	0.219	
29	1	1.351	0.509	0.094	
30	1	1.673	0.235	0.156	
31	1	0.870	0.029	0.174	
32	2	0.703	-0.717	0.554	
33	2	0.814	0.173	-0.677	
34	2	0.698	-0.201	-1.749	
35	2	0.640	-1.137	0.083	
36	2	0.655	-0.284	0.173	
37	2	0.733	-0.154	0.356	
38	3	0.666	-0.777	0.688	0.916

Table O.9. Mathematics Grade 5 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.757	-0.935	0.086	
2	1	1.173	0.474	0.175	
3	1	1.478	-0.722	0.028	
4	1	1.329	-1.347	0.039	
5	1	1.520	0.826	0.182	
6	1	1.221	-0.882	0.087	
7	1	0.932	-0.146	0.200	
8	1	1.521	0.651	0.133	
9	1	1.520	0.704	0.082	
10	1	1.023	-0.861	0.117	
11	1	1.271	-0.236	0.220	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
12	1	0.998	-0.057	0.183	
13	1	1.448	0.218	0.321	
14	1	1.842	0.521	0.205	
15	1	1.133	0.620	0.300	
16	1	0.882	-0.450	0.209	
17	1	0.903	-0.236	0.265	
18	1	0.833	-0.480	0.219	
19	1	0.378	-0.821	0.029	
20	1	0.788	-0.501	0.113	
21	1	1.204	-0.492	0.151	
22	1	1.469	-0.633	0.125	
23	1	1.138	0.805	0.174	
24	1	1.260	0.685	0.266	
25	1	1.309	-0.181	0.175	
26	1	1.402	0.338	0.138	
27	1	1.263	0.015	0.171	
28	1	0.908	-0.851	0.134	
29	1	0.790	-0.091	0.186	
30	1	1.482	-0.403	0.270	
31	1	1.400	-0.469	0.067	
32	2	0.599	1.085	-1.079	
33	2	0.679	-0.097	-0.019	
34	2	0.714	0.714	-0.311	
35	2	0.802	1.527	-0.943	
36	2	1.229	-0.546	0.497	
37	2	0.618	-0.566	1.979	
38	3	0.506	-0.129	-0.797	-1.807

Table O.10. Mathematics Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	1.289	0.440	0.290	
2	1	1.282	0.173	0.241	
3	1	1.443	1.111	0.207	
4	1	1.496	-0.314	0.258	
5	1	1.375	-0.169	0.152	
6	1	1.065	-0.174	0.296	
7	1	1.469	0.502	0.099	
8	1	1.773	0.427	0.215	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
9	1	1.315	0.058	0.213	
10	1	1.433	0.058	0.239	
11	1	0.936	0.838	0.156	
12	1	1.317	-0.125	0.327	
13	1	1.600	-0.236	0.220	
14	1	1.184	-0.362	0.357	
15	1	1.409	0.522	0.110	
16	1	1.739	0.011	0.261	
17	1	1.562	0.177	0.246	
18	1	1.023	0.625	0.226	
19	1	1.059	0.169	0.265	
20	1	1.861	0.003	0.192	
21	1	1.202	-0.101	0.175	
22	1	1.074	0.650	0.302	
23	1	1.603	0.304	0.200	
24	1	1.393	-0.194	0.320	
25	1	1.145	0.428	0.291	
26	1	1.631	-0.153	0.111	
27	1	1.310	-0.041	0.279	
28	1	1.310	-0.037	0.258	
29	1	1.355	0.004	0.184	
30	1	1.281	-0.280	0.252	
31	1	1.313	-0.594	0.226	
32	2	1.033	0.884	1.181	
33	2	0.917	-0.161	0.417	
34	2	0.908	2.012	1.405	
35	2	0.774	-0.147	1.263	
36	2	0.916	-0.900	1.543	
37	2	0.747	0.891	-1.846	
38	2	0.693	0.376	1.199	
39	3	0.868	-0.539	-0.741	1.062

Table O.11. Mathematics Grade 7 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.984	0.739	0.244	
2	1	1.325	-0.049	0.237	
3	1	2.308	0.389	0.190	
4	1	2.582	0.120	0.288	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
5	1	1.488	-0.174	0.157	
6	1	1.951	0.592	0.192	
7	1	1.947	0.315	0.261	
8	1	1.105	-0.113	0.174	
9	1	1.953	0.275	0.204	
10	1	0.844	0.637	0.195	
11	1	0.901	-0.930	0.200	
12	1	1.655	0.050	0.263	
13	1	0.921	0.569	0.265	
14	1	1.401	-0.764	0.207	
15	1	2.144	1.009	0.206	
16	1	0.494	-1.538	0.055	
17	1	1.227	-0.114	0.272	
18	1	1.356	0.391	0.280	
20	1	1.779	0.684	0.195	
21	1	1.032	0.369	0.238	
22	1	1.912	0.484	0.233	
23	1	1.559	0.299	0.214	
24	1	1.267	0.057	0.332	
25	1	0.721	0.187	0.215	
26	1	0.682	0.674	0.268	
27	1	1.645	0.581	0.192	
28	1	1.174	-0.307	0.255	
29	1	1.125	0.097	0.155	
30	1	1.989	0.155	0.289	
31	1	1.753	0.705	0.174	
32	1	1.077	-0.987	0.240	
33	1	1.726	0.198	0.114	
34	1	1.279	0.326	0.191	
35	2	0.914	-0.650	-0.290	
36	2	1.331	0.026	1.919	
37	2	1.067	1.175	0.764	
38	2	0.608	0.956	-0.867	
39	2	0.914	0.090	-0.894	
40	2	1.203	1.436	-0.493	
41	2	0.839	0.455	-1.878	
42	3	0.942	0.343	0.753	-0.984

Table O.12. Mathematics Grade 8 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.779	-0.113	0.133	
2	1	0.666	0.497	0.268	
3	1	1.195	-0.111	0.245	
4	1	1.471	-0.011	0.174	
5	1	1.249	0.360	0.409	
6	1	2.013	0.700	0.214	
7	1	0.913	0.402	0.262	
8	1	0.692	0.112	0.202	
9	1	1.000	0.172	0.191	
10	1	1.436	0.740	0.262	
11	1	1.561	-0.066	0.274	
12	1	1.382	0.348	0.201	
13	1	1.492	0.230	0.242	
14	1	0.958	0.365	0.121	
15	1	1.650	0.603	0.186	
16	1	1.910	0.021	0.136	
17	1	1.263	0.562	0.367	
18	1	1.136	0.377	0.178	
19	1	0.813	0.051	0.209	
20	1	1.789	0.333	0.112	
21	1	1.305	0.499	0.271	
22	1	0.875	0.255	0.180	
23	1	1.850	0.851	0.217	
24	1	0.593	-0.883	0.068	
25	1	0.733	-0.525	0.165	
26	1	0.994	0.123	0.214	
27	1	1.335	0.667	0.278	
28	1	0.741	-0.769	0.016	
29	1	1.154	1.022	0.246	
30	1	1.267	1.060	0.212	
31	1	1.321	0.119	0.220	
32	1	0.921	0.508	0.383	
33	1	1.105	0.575	0.246	
34	2	0.611	0.051	0.395	
35	2	0.801	1.461	0.970	
36	2	1.078	0.601	0.607	
37	2	0.879	0.636	0.059	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
38	2	0.898	1.261	0.411	
39	2	0.597	1.104	-1.246	
40	2	0.843	1.438	-1.658	
41	3	0.474	-0.144	0.209	0.736

Appendix P: Derivation and Estimation of Classification Consistency and Accuracy

Classification Consistency

Assume that θ is a single latent trait measured by a test and denote Φ as a latent random variable. When a test X consists of *K* items and its maximum number correct score is *N*, the marginal probability of the number correct (NC) score *x* is

$$P(X = x) = \int P(X = x \mid \Phi = \theta) g(\theta) d\theta, \quad x = 0,1,...,N$$

where

$g(\theta)$ is the density of θ.

In this report, the marginal distribution $P(X = x)$ is denoted as $f(x)$, and the conditional error distribution $P(X = x \mid \Phi = \theta)$ is denoted as $f(x \mid \theta)$. It is assumed that examinees are classified into one of H mutually exclusive categories on the basis of predetermined *H* - 1 observed score cutoffs, C_1, C_2, …, C_{H-1}. Let L_h represent the hth category into which examinees with $C_{h-1} \le X < C_h$ are classified. $C_0 = 0$ and $C_H =$ the maximum number-correct score plus one. Then, the conditional and marginal probabilities of each category classification are as follows:

$$P(X \in L_h \mid \theta) = \sum_{x=C_{h-1}}^{C_h - 1} f(x \mid \theta), \; h = 1, 2, \ldots, H$$

$$P(X \in L_h) = \int \sum_{x=C_{h-1}}^{C_h - 1} f(x \mid \theta) g(\theta) d\theta, \; h = 1, 2, \ldots, H$$

Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each OP administration, a psychometric model was used to obtain the estimated classification consistency indices using test scores from a single administration. Based on the psychometric model, a symmetric *H*-by-*H* contingency table can be constructed. The elements of the *H*-by-*H* contingency table consist of the joint probabilities of the row and column observed category classifications.

That two administrations are independent implies that if X_1 and X_2 represent the raw score random variables on the two administrations, then, conditioned on θ, X_1 and X_2 are independent and identically distributed. Consequently, the conditional bivariate distribution of X_1 and X_2 is

$$f(x_1, \; x_2 \mid \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta)$$

The marginal bivariate distribution of X_1 and X_2 can be expressed as follows:

$$f(x_1,\ x_2) = \int f(x_1, x_2 \mid \theta) f(\theta) d\theta$$

Consistent classification means that both X_1 and X_2 fall in the same category. The conditional probability of falling in the same category on the two administrations is

$$P(X_1 \in L_h,\ \ X_2 \in L_h \mid \theta) = \left[\sum_{x_1 = C_{h-1}}^{C_{h-1}} f(x_1 \mid \theta) \right]^2, \ h = 1, 2, \ldots, \mathrm{H}$$

The agreement index P, conditional on theta, is obtained by

$$P(\theta) = \sum_{h=1}^{H} P(X_1 \in L_h,\ \ X_2 \in L_h \mid \theta)$$

The agreement index (classification consistency) can be computed as

$$P = \int P(\theta) g(\theta) d(\theta)$$

The probability of consistent classification by chance, P_C, is the sum of squared marginal probabilities of each category classification.

$$P_C = \sum_{h=1}^{H} P(X_1 \in L_h) P(X_2 \in L_h) = \sum_{h=1}^{H} \left[P(X_1 \in L_h) \right]^2$$

Then, Kappa (Cohen, 1960) is

$$k = \frac{P - P_C}{1 - P_C}$$

Classification Accuracy

Let Γ_w denote true category. When an examinee has an observed score, $x \in L_h$ (h =1, 2,…, H), and a latent score, $\theta \in \Gamma_w$ (w =1, 2,…, H), an accurate classification is made when $h = w$. The conditional probability of accurate classification is

$$\gamma(\theta) = P(X \in L_w \mid \theta),$$

where

w is the category such that $\theta \in \Gamma_w$

Lee (2008) thoroughly discusses this IRT method for estimating decision indices, including the computational method used to estimate the results when integrating across the latent variable, θ.

Estimating Classification Indices

The classification consistency and accuracy estimates were obtained using an open-source software program, IRT-CLASS v2.0 (Lee & Kolen, 2006). Below is a brief description of the files that are used and their purpose. (See the IRT-CLASS v2.0 manual for complete instructions.)

Files needed:

- Raw-to-Scale score conversion file
    a. Contains the raw-to-scale score conversions
    b. This is used to provide both raw and scale score classification estimates, which is useful when the raw-to-scale score transformation is not one-to-one.

- Cut score file
    a. Contains the cut scores to be used
    b. Results are provided for all cut scores simultaneously (all performance levels), as well as the estimates based on each of the cut scores separately (Level III only).

- Item parameter file
    a. This contains the IRT model used and item parameter estimates.
    b. This information is used when calculating the classification indices.

- Theta file
    a. Contains the theta distribution in terms of quadrature points
    b. The theta and the item parameter files are used to solve the integrals mentioned above.

- Control card
    a. This is used to run the program.
    b. Identifies the names of the four files above and gives a name to the output file

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Tables Q.1–Q.12 show the PBT raw-to-scale score conversion tables, while Tables Q.13-Q.24 show the CBT raw-to-scale score conversion tables. Tables Q.25–Q.36 show the scale score distributions that include all students with valid scores, by frequency (n-count), percent, cumulative frequency, and cumulative percent.

Table Q.1. PBT ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	532	25	18	596	6
1	536	20	19	598	6
2	541	17	20	602	6
3	545	14	21	603	5
4	550	12	22	605	5
5	555	10	23	607	5
6	560	9	24	610	5
7	565	8	25	612	6
8	569	7	26	615	6
9	572	7	27	617	6
10	575	7	28	620	6
11	578	6	29	624	6
12	583	6	30	629	7
13	584	6	31	632	8
14	586	6	32	639	10
15	589	6	33	650	13
16	591	6	34	654	16
17	594	6			

Table Q.2. PBT ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	528	22	8	564	7
1	533	18	9	567	7
2	537	15	10	570	7
3	542	13	11	573	7
4	546	11	12	576	7
5	551	10	13	579	6
6	556	9	14	581	6
7	560	8	15	584	6

Raw Score	Scale Score	Standard Error
16	587	6
17	589	6
18	592	6
19	594	6
20	597	6
21	599	6
22	603	6
23	604	6
24	607	6
25	609	6
26	612	6
27	615	6
28	619	7
29	622	7
30	626	8
31	632	9
32	639	11
33	651	15
34	656	18

Table Q.3. PBT ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	513	40
1	518	34
2	522	29
3	527	25
4	532	21
5	536	18
6	541	16
7	545	14
8	550	12
9	554	10
10	559	9
11	563	8
12	566	8
13	569	7
14	572	7
15	574	7
16	577	6
17	579	6
18	581	6
19	583	5
20	585	5
21	587	5
22	589	5
23	590	5
24	592	5
25	594	5
26	596	5
27	598	5
28	599	5
29	601	5
30	603	5
31	605	5
32	609	5
33	610	5
34	612	6
35	615	6
36	618	6
37	622	7
38	625	7
39	629	8
40	634	9
41	640	11
42	649	14
43	654	16
44	658	18

Table Q.4. PBT ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	502	47	23	586	5
1	506	40	24	588	5
2	511	34	25	590	5
3	515	29	26	591	5
4	520	25	27	593	5
5	525	22	28	595	5
6	529	19	29	597	5
7	534	16	30	600	5
8	538	14	31	602	5
9	543	13	32	604	5
10	547	11	33	606	5
11	552	10	34	609	6
12	557	9	35	614	6
13	561	9	36	615	6
14	564	8	37	618	7
15	567	8	38	622	7
16	570	7	39	626	8
17	573	7	40	631	9
18	575	7	41	637	10
19	577	6	42	646	13
20	579	6	43	651	15
21	582	6	44	656	18
22	584	5			

Table Q.5. PBT ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	510	46	9	551	9
1	515	38	10	556	8
2	519	32	11	559	7
3	524	27	12	562	7
4	528	22	13	565	6
5	533	19	14	568	6
6	537	16	15	570	6
7	542	13	16	572	5
8	547	11	17	574	5

Raw Score	Scale Score	Standard Error
18	576	5
19	578	5
20	580	5
21	582	5
22	584	5
23	586	5
24	588	5
25	589	5
26	591	5
27	593	5
28	595	5
29	597	5
30	598	5
31	600	5
32	602	5
33	604	5
34	607	5
35	609	5
36	611	5
37	613	5
38	616	6
39	619	6
40	623	6
41	626	7
42	630	7
43	635	8
44	641	9
45	652	15
46	657	18

Table Q.6. PBT ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	507	43
1	511	36
2	516	30
3	520	25
4	525	21
5	530	17
6	534	15
7	539	12
8	543	11
9	548	9
10	553	8
11	557	7
12	560	7
13	563	6
14	566	6
15	568	6
16	570	6
17	572	6
18	574	5
19	576	5
20	578	5
21	580	5
22	582	5
23	584	5
24	585	5
25	587	5
26	589	5
27	590	5
28	592	5
29	594	5
30	595	5
31	597	5
32	599	5
33	601	5
34	603	5
35	605	5

Raw Score	Scale Score	Standard Error
36	607	5
37	609	6
38	612	6
39	614	6
40	617	6
41	620	7
42	624	7
43	629	8
44	635	10
45	646	14
46	651	17

Table Q.7. PBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	524	38
1	529	31
2	533	26
3	538	21
4	543	18
5	547	15
6	552	13
7	557	11
8	563	9
9	568	8
10	571	7
11	574	7
12	577	6
13	579	6
14	581	5
15	583	5
16	585	5
17	587	5
18	588	5
19	590	5
20	591	4
21	593	4
22	594	4
23	596	4
24	597	4
25	598	4
26	600	4
27	601	4
28	603	4
29	604	4
30	606	4
31	607	4
32	609	5
33	611	5
34	613	5
35	615	5
36	617	5
37	619	5
38	622	6
39	626	7
40	631	8
41	640	12
42	645	15

Table Q.8. PBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	521	53	24	594	4
1	526	46	25	595	4
2	530	39	26	597	4
3	535	34	27	598	4
4	539	29	28	599	4
5	544	24	29	600	4
6	549	20	30	602	4
7	553	16	31	603	4
8	558	13	32	604	4
9	564	10	33	606	4
10	568	8	34	607	4
11	571	7	35	609	4
12	574	7	36	610	4
13	577	6	37	612	4
14	579	6	38	614	5
15	581	5	39	616	5
16	583	5	40	618	5
17	584	5	41	621	5
18	586	5	42	624	6
19	588	4	43	628	7
20	589	4	44	633	9
21	590	4	45	643	14
22	591	4	46	648	17
23	593	4			

Table Q.9. PBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	527	49	8	565	9
1	532	42	9	569	8
2	536	36	10	572	7
3	541	30	11	575	6
4	546	25	12	577	6
5	550	20	13	580	5
6	555	16	14	581	5
7	559	12	15	583	5

Raw Score	Scale Score	Standard Error
16	585	5
17	586	4
18	588	4
19	589	4
20	591	4
21	592	4
22	593	4
23	595	4
24	596	4
25	597	4
26	598	4
27	600	4
28	601	4
29	602	4
30	604	4
31	605	4
32	606	4
33	607	4
34	609	4
35	610	4
36	612	4
37	613	4
38	616	4
39	617	5
40	619	5
41	621	5
42	624	6
43	628	7
44	634	9
45	644	15
46	649	18

Table Q.10. PBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	529	74
1	533	62
2	538	52
3	542	43
4	547	36
5	552	30
6	556	24
7	561	20
8	565	16
9	573	10
10	578	8
11	581	7
12	584	6
13	586	5
14	588	5
15	589	4
16	591	4
17	592	4
18	593	4
19	594	4
20	596	4
21	597	3
22	598	3
23	599	3
24	600	3
25	601	3
26	602	3
27	603	3
28	604	3
29	605	3
30	606	3
31	607	3
32	608	3
33	609	3

Raw Score	Scale Score	Standard Error
34	610	3
35	612	3
36	613	4
37	614	4
38	616	4
39	617	4
40	619	4
41	621	4
42	623	5
43	625	5
44	628	6
45	631	7
46	636	8
47	644	12
48	649	15

Table Q.11. PBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	520	74
1	525	64
2	529	56
3	534	48
4	538	41
5	543	35
6	548	30
7	552	24
8	557	20
9	569	11
10	575	8
11	579	7
12	582	6
13	584	5
14	586	5
15	588	5
16	590	4
17	591	4
18	593	4
19	594	4
20	595	4
21	596	3
22	597	3
23	598	3
24	599	3
25	600	3
26	601	3
27	602	3
28	603	3
29	604	3
30	604	3
31	605	3
32	606	3
33	607	3
34	608	3
35	609	3
36	610	3
37	611	3
38	612	3
39	613	3
40	614	3
41	615	3
42	616	4
43	618	4
44	620	4
45	621	4
46	624	5
47	627	6
48	631	7
49	639	11

Raw Score	Scale Score	Standard Error
50	643	14

Table Q.12. PBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	520	80	26	604	3
1	525	68	27	605	3
2	530	58	28	606	3
3	534	49	29	607	3
4	539	42	30	608	3
5	543	36	31	609	3
6	548	30	32	610	3
7	552	26	33	611	3
8	557	22	34	612	3
9	569	14	35	613	3
10	575	11	36	614	3
11	580	9	37	615	3
12	583	8	38	616	3
13	586	7	39	617	4
14	588	6	40	618	4
15	590	5	41	620	4
16	592	5	42	622	4
17	593	5	43	623	4
18	595	4	44	625	5
19	596	4	45	627	5
20	597	4	46	630	6
21	598	4	47	634	7
22	600	4	48	639	9
23	601	4	49	648	14
24	602	3	50	653	17
25	603	3			

Table Q.13. CBT ELA Grade 3 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	534	22	18	598	6
1	538	19	19	600	6
2	543	16	20	604	5
3	547	13	21	605	5
4	552	11	22	607	5
5	557	10	23	609	5
6	562	8	24	612	6
7	567	7	25	614	6
8	571	7	26	617	6
9	574	7	27	619	6
10	577	6	28	622	6
11	580	6	29	626	7
12	585	6	30	631	8
13	586	6	31	634	8
14	588	6	32	641	10
15	591	6	33	652	14
16	593	6	34	654	16
17	596	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.14. CBT ELA Grade 4 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	530	20	13	581	6
1	535	17	14	583	6
2	539	15	15	586	6
3	544	12	16	589	6
4	548	11	17	591	6
5	553	9	18	594	6
6	558	8	19	596	6
7	562	8	20	599	6
8	566	7	21	601	6
9	569	7	22	605	6
10	572	7	23	606	6
11	575	7	24	609	6
12	578	7	25	611	6

Raw Score	Scale Score*	Standard Error
26	614	6
27	617	7
28	621	7
29	624	7
30	628	8
31	634	9
32	641	11
33	653	16
34	656	18

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.15. CBT ELA Grade 5 RSSS Table

Raw Score	Scale Score*	Standard Error
0	515	38
1	520	32
2	524	28
3	529	23
4	534	20
5	538	17
6	543	15
7	547	13
8	552	11
9	556	10
10	561	9
11	565	8
12	568	8
13	571	7
14	574	7
15	576	6
16	579	6
17	581	6
18	583	5
19	585	5
20	587	5
21	589	5
22	591	5
23	592	5
24	594	5
25	596	5
26	598	5
27	600	5
28	601	5
29	603	5
30	605	5
31	607	5
32	611	6
33	612	6
34	614	6
35	617	6
36	620	7
37	624	7
38	627	8
39	631	9
40	636	10
41	642	11
42	651	14
43	656	17
44	658	18

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.16. CBT ELA Grade 6 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	503	45	23	587	5
1	507	39	24	589	5
2	512	33	25	591	5
3	516	29	26	592	5
4	521	24	27	594	5
5	526	21	28	596	5
6	530	18	29	598	5
7	535	16	30	601	5
8	539	14	31	603	5
9	544	12	32	605	5
10	548	11	33	607	5
11	553	10	34	610	6
12	558	9	35	615	6
13	562	8	36	616	6
14	565	8	37	619	7
15	568	8	38	623	7
16	571	7	39	627	8
17	574	7	40	632	9
18	576	6	41	638	11
19	578	6	42	647	14
20	580	6	43	652	16
21	583	6	44	656	18
22	585	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.17. CBT ELA Grade 7 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	511	44	8	548	11
1	516	36	9	552	9
2	520	31	10	557	8
3	525	25	11	560	7
4	529	22	12	563	7
5	534	18	13	566	6
6	538	15	14	569	6
7	543	13	15	571	5

Raw Score	Scale Score*	Standard Error
16	573	5
17	575	5
18	577	5
19	579	5
20	581	5
21	583	5
22	585	5
23	587	5
24	589	5
25	590	5
26	592	5
27	594	5
28	596	5
29	598	5
30	599	5
31	601	5
32	603	5
33	605	5
34	608	5
35	610	5
36	612	5
37	614	5
38	617	6
39	620	6
40	624	6
41	627	7
42	631	7
43	636	8
44	642	10
45	653	16
46	657	18

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.18. CBT ELA Grade 8 RSSS Table

Raw Score	Scale Score*	Standard Error
0	508	41
1	512	35
2	517	28
3	521	24
4	526	20
5	531	16
6	535	14
7	540	12
8	544	10
9	549	9
10	554	8
11	558	7
12	561	7
13	564	6
14	567	6
15	569	6
16	571	6
17	573	5
18	575	5
19	577	5
20	579	5
21	581	5
22	583	5
23	585	5
24	586	5
25	588	5
26	590	5
27	591	5
28	593	5
29	595	5
30	596	5
31	598	5

Raw Score	Scale Score*	Standard Error
32	600	5
33	602	5
34	604	5
35	606	5
36	608	5
37	610	6
38	613	6
39	615	6
40	618	6
41	621	7
42	625	7
43	630	8
44	636	10
45	647	15
46	651	17

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.19. CBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score*	Standard Error
0	526	35
1	531	28
2	535	24
3	540	20
4	545	16
5	549	14
6	554	12
7	559	11
8	565	9
9	570	8
10	573	7
11	576	6
12	579	6
13	581	6
14	583	5
15	585	5
16	587	5
17	589	5
18	590	5
19	592	4
20	593	4
21	595	4
22	596	4
23	598	4
24	599	4
25	600	4
26	602	4
27	603	4
28	605	4
29	606	4
30	608	4
31	609	5
32	611	5
33	613	5
34	615	5
35	617	5
36	619	5
37	621	6
38	624	6
39	628	7
40	633	9
41	642	13
42	645	15

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.20. CBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	522	52	24	595	4
1	527	44	25	596	4
2	531	38	26	598	4
3	536	32	27	599	4
4	540	28	28	600	4
5	545	23	29	601	4
6	550	19	30	603	4
7	554	16	31	604	4
8	559	12	32	605	4
9	565	10	33	607	4
10	569	8	34	608	4
11	572	7	35	610	4
12	575	6	36	611	4
13	578	6	37	613	4
14	580	5	38	615	5
15	582	5	39	617	5
16	584	5	40	619	5
17	585	5	41	622	6
18	587	4	42	625	6
19	589	4	43	629	8
20	590	4	44	634	10
21	591	4	45	644	15
22	592	4	46	648	17
23	594	4			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.21. CBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	529	46	7	561	12
1	534	39	8	567	9
2	538	34	9	571	7
3	543	28	10	574	6
4	548	22	11	577	6
5	552	18	12	579	5
6	557	14	13	582	5

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
14	583	5	31	607	4
15	585	5	32	608	4
16	587	4	33	609	4
17	588	4	34	611	4
18	590	4	35	612	4
19	591	4	36	614	4
20	593	4	37	615	4
21	594	4	38	618	5
22	595	4	39	619	5
23	597	4	40	621	5
24	598	4	41	623	6
25	599	4	42	626	6
26	600	4	43	630	8
27	602	4	44	636	10
28	603	4	45	646	16
29	604	4	46	649	18
30	606	4			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.22. CBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	531	68	15	591	4
1	535	58	16	593	4
2	540	48	17	594	4
3	544	40	18	595	4
4	549	33	19	596	3
5	554	27	20	598	3
6	558	22	21	599	3
7	563	18	22	600	3
8	567	15	23	601	3
9	575	10	24	602	3
10	580	7	25	603	3
11	583	6	26	604	3
12	586	5	27	605	3
13	588	5	28	606	3
14	590	4	29	607	3

Raw Score	Scale Score*	Standard Error
30	608	3
31	609	3
32	610	3
33	611	3
34	612	4
35	614	4
36	615	4
37	616	4
38	618	4
39	619	4
40	621	4
41	623	5
42	625	5
43	627	6
44	630	6
45	633	7
46	638	9
47	646	13
48	649	15

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.23. CBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score*	Standard Error
0	522	70
1	527	60
2	531	53
3	536	45
4	540	39
5	545	33
6	550	27
7	554	23
8	559	18
9	571	10
10	577	8
11	581	6
12	584	6
13	586	5
14	588	5
15	590	4
16	592	4
17	593	4
18	595	4
19	596	3
20	597	3
21	598	3
22	599	3
23	600	3
24	601	3
25	602	3
26	603	3
27	604	3
28	605	3
29	606	3
30	606	3
31	607	3
32	608	3
33	609	3
34	610	3
35	611	3
36	612	3
37	613	3
38	614	3
39	615	3
40	616	4
41	617	4
42	618	4
43	620	4

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
44	622	5	48	633	8
45	623	5	49	641	13
46	626	6	50	643	14
47	629	7			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q.24. CBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	522	76	26	606	3
1	527	63	27	607	3
2	532	53	28	608	3
3	536	46	29	609	3
4	541	38	30	610	3
5	545	33	31	611	3
6	550	28	32	612	3
7	554	24	33	613	3
8	559	20	34	614	3
9	571	13	35	615	3
10	577	10	36	616	3
11	582	8	37	617	4
12	585	7	38	618	4
13	588	6	39	619	4
14	590	5	40	620	4
15	592	5	41	622	4
16	594	4	42	624	5
17	595	4	43	625	5
18	597	4	44	627	5
19	598	4	45	629	6
20	599	4	46	632	6
21	600	4	47	636	8
22	602	3	48	641	10
23	603	3	49	650	15
24	604	3	50	653	17
25	605	3			

*A CBT mode adjustment has been taken into account for these scale scores

Table Q.25. ELA Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
532	50	0.03%	50	0.03%
534	4	0.00%	54	0.03%
536	135	0.07%	189	0.10%
538	20	0.01%	209	0.11%
541	406	0.22%	615	0.34%
543	68	0.04%	683	0.37%
545	790	0.43%	1,473	0.81%
547	161	0.09%	1,634	0.90%
550	1,310	0.72%	2,944	1.61%
552	239	0.13%	3,183	1.74%
555	1,751	0.96%	4,934	2.70%
557	351	0.19%	5,285	2.89%
560	2,233	1.22%	7,518	4.12%
562	463	0.25%	7,981	4.37%
565	2,590	1.42%	10,571	5.79%
567	567	0.31%	11,138	6.10%
569	2,993	1.64%	14,131	7.74%
571	635	0.35%	14,766	8.09%
572	3,397	1.86%	18,163	9.95%
574	745	0.41%	18,908	10.4%
575	3,822	2.09%	22,730	12.5%
577	843	0.46%	23,573	12.9%
578	4,326	2.37%	27,899	15.3%
580	967	0.53%	28,866	15.8%
583	4,820	2.64%	33,686	18.5%
584	5,168	2.83%	38,854	21.3%
585	1,041	0.57%	39,895	21.9%
586	6,839	3.75%	46,734	25.6%
588	1,213	0.66%	47,947	26.3%
589	6,000	3.29%	53,947	29.6%
591	7,824	4.29%	61,771	33.8%
593	1,412	0.77%	63,183	34.6%
594	6,492	3.56%	69,675	38.2%
596	8,265	4.53%	77,940	42.7%
598	8,541	4.68%	86,481	47.4%
600	1,469	0.80%	87,950	48.2%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
602	7,222	3.96%	95,172	52.1%
603	7,345	4.02%	102,517	56.2%
604	1,452	0.80%	103,969	57.0%
605	8,834	4.84%	112,803	61.8%
607	8,890	4.87%	121,693	66.7%
609	1,418	0.78%	123,111	67.4%
610	7,480	4.10%	130,591	71.5%
612	8,560	4.69%	139,151	76.2%
614	1,133	0.62%	140,284	76.8%
615	7,153	3.92%	147,437	80.8%
617	7,994	4.38%	155,431	85.1%
619	894	0.49%	156,325	85.6%
620	6,462	3.54%	162,787	89.2%
622	763	0.42%	163,550	89.6%
624	5,711	3.13%	169,261	92.7%
626	585	0.32%	169,846	93.0%
629	4,733	2.59%	174,579	95.6%
631	362	0.20%	174,941	95.8%
632	3,568	1.95%	178,509	97.8%
634	189	0.10%	178,698	97.9%
639	2,316	1.27%	181,014	99.2%
641	76	0.04%	181,090	99.2%
650	1,115	0.61%	182,205	99.8%
652	26	0.01%	182,231	99.8%
654	328	0.18%	182,559	100%

Table Q.26. ELA Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
528	20	0.01%	20	0.01%
530	8	0.00%	28	0.02%
533	73	0.04%	101	0.05%
535	12	0.01%	113	0.06%
537	229	0.12%	342	0.18%
539	48	0.03%	390	0.21%
542	467	0.25%	857	0.46%
544	78	0.04%	935	0.50%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
546	872	0.47%	1,807	0.97%
548	126	0.07%	1,933	1.04%
551	1,290	0.69%	3,223	1.73%
553	230	0.12%	3,453	1.85%
556	1,497	0.80%	4,950	2.66%
558	296	0.16%	5,246	2.82%
560	1,879	1.01%	7,125	3.83%
562	394	0.21%	7,519	4.04%
564	2,324	1.25%	9,843	5.29%
566	525	0.28%	10,368	5.57%
567	2,804	1.51%	13,172	7.07%
569	608	0.33%	13,780	7.40%
570	3,199	1.72%	16,979	9.12%
572	672	0.36%	17,651	9.48%
573	3,769	2.02%	21,420	11.5%
575	909	0.49%	22,329	12.0%
576	4,169	2.24%	26,498	14.2%
578	951	0.51%	27,449	14.7%
579	4,480	2.41%	31,929	17.1%
581	6,149	3.30%	38,078	20.4%
583	1,191	0.64%	39,269	21.1%
584	5,530	2.97%	44,799	24.1%
586	1,295	0.70%	46,094	24.8%
587	6,115	3.28%	52,209	28.0%
589	7,862	4.22%	60,071	32.3%
591	1,464	0.79%	61,535	33.0%
592	6,935	3.72%	68,470	36.8%
594	8,925	4.79%	77,395	41.6%
596	1,575	0.85%	78,970	42.4%
597	7,793	4.19%	86,763	46.6%
599	9,583	5.15%	96,346	51.7%
601	1,505	0.81%	97,851	52.6%
603	8,277	4.45%	106,128	57.0%
604	8,493	4.56%	114,621	61.6%
605	1,604	0.86%	116,225	62.4%
606	1,519	0.82%	117,744	63.2%
607	8,495	4.56%	126,239	67.8%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
609	10,005	5.37%	136,244	73.2%
611	1,416	0.76%	137,660	73.9%
612	8,213	4.41%	145,873	78.3%
614	1,278	0.69%	147,151	79.0%
615	8,040	4.32%	155,191	83.3%
617	1,198	0.64%	156,389	84.0%
619	7,292	3.92%	163,681	87.9%
621	1,026	0.55%	164,707	88.5%
622	6,461	3.47%	171,168	91.9%
624	787	0.42%	171,955	92.3%
626	5,207	2.80%	177,162	95.1%
628	552	0.30%	177,714	95.4%
632	3,758	2.02%	181,472	97.5%
634	375	0.20%	181,847	97.7%
639	2,551	1.37%	184,398	99.0%
641	190	0.10%	184,588	99.1%
651	1,196	0.64%	185,784	99.8%
653	82	0.04%	185,866	99.8%
656	339	0.18%	186,205	100%

Table Q.27. ELA Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
513	23	0.01%	23	0.01%
515	3	0.00%	26	0.01%
518	31	0.02%	57	0.03%
520	2	0.00%	59	0.03%
522	53	0.03%	112	0.06%
524	5	0.00%	117	0.06%
527	105	0.06%	222	0.12%
529	22	0.01%	244	0.14%
532	233	0.13%	477	0.26%
534	20	0.01%	497	0.28%
536	438	0.24%	935	0.52%
538	89	0.05%	1,024	0.57%
541	746	0.41%	1,770	0.98%
543	125	0.07%	1,895	1.05%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
545	957	0.53%	2,852	1.58%
547	202	0.11%	3,054	1.69%
550	1,363	0.75%	4,417	2.44%
552	308	0.17%	4,725	2.62%
554	1,607	0.89%	6,332	3.50%
556	362	0.20%	6,694	3.70%
559	1,877	1.04%	8,571	4.74%
561	438	0.24%	9,009	4.99%
563	2,156	1.19%	11,165	6.18%
565	517	0.29%	11,682	6.47%
566	2,240	1.24%	13,922	7.71%
568	615	0.34%	14,537	8.05%
569	2,509	1.39%	17,046	9.43%
571	655	0.36%	17,701	9.80%
572	2,699	1.49%	20,400	11.3%
574	3,635	2.01%	24,035	13.3%
576	795	0.44%	24,830	13.7%
577	3,017	1.67%	27,847	15.4%
579	4,019	2.22%	31,866	17.6%
581	4,122	2.28%	35,988	19.9%
583	4,267	2.36%	40,255	22.3%
585	4,540	2.51%	44,795	24.8%
587	4,892	2.71%	49,687	27.5%
589	5,104	2.82%	54,791	30.3%
590	4,090	2.26%	58,881	32.6%
591	1,120	0.62%	60,001	33.2%
592	5,595	3.10%	65,596	36.3%
594	5,785	3.20%	71,381	39.5%
596	5,953	3.29%	77,334	42.8%
598	6,277	3.47%	83,611	46.3%
599	5,307	2.94%	88,918	49.2%
600	1,363	0.75%	90,281	50.0%
601	6,899	3.82%	97,180	53.8%
603	7,344	4.06%	104,524	57.9%
605	7,552	4.18%	112,076	62.0%
607	1,458	0.81%	113,534	62.8%
609	6,357	3.52%	119,891	66.4%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
610	6,538	3.62%	126,429	70.0%
611	1,506	0.83%	127,935	70.8%
612	8,082	4.47%	136,017	75.3%
614	1,408	0.78%	137,425	76.1%
615	6,496	3.60%	143,921	79.7%
617	1,350	0.75%	145,271	80.4%
618	6,287	3.48%	151,558	83.9%
620	1,290	0.71%	152,848	84.6%
622	5,912	3.27%	158,760	87.9%
624	1,190	0.66%	159,950	88.5%
625	5,288	2.93%	165,238	91.5%
627	983	0.54%	166,221	92.0%
629	4,344	2.40%	170,565	94.4%
631	791	0.44%	171,356	94.8%
634	3,475	1.92%	174,831	96.8%
636	565	0.31%	175,396	97.1%
640	2,404	1.33%	177,800	98.4%
642	348	0.19%	178,148	98.6%
649	1,424	0.79%	179,572	99.4%
651	183	0.10%	179,755	99.5%
654	655	0.36%	180,410	99.9%
656	76	0.04%	180,486	99.9%
658	193	0.11%	180,679	100%

Table Q.28. ELA Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
502	13	0.01%	13	0.01%
503	3	0.00%	16	0.01%
506	15	0.01%	31	0.02%
511	32	0.02%	63	0.04%
512	1	0.00%	64	0.04%
515	69	0.04%	133	0.07%
516	11	0.01%	144	0.08%
520	127	0.07%	271	0.15%
521	29	0.02%	300	0.17%
525	210	0.12%	510	0.28%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
526	29	0.02%	539	0.30%
529	358	0.20%	897	0.50%
530	56	0.03%	953	0.53%
534	551	0.31%	1,504	0.84%
535	115	0.06%	1,619	0.90%
538	790	0.44%	2,409	1.34%
539	132	0.07%	2,541	1.41%
543	1,019	0.57%	3,560	1.98%
544	224	0.12%	3,784	2.10%
547	1,241	0.69%	5,025	2.79%
548	308	0.17%	5,333	2.96%
552	1,474	0.82%	6,807	3.78%
553	377	0.21%	7,184	3.99%
557	1,761	0.98%	8,945	4.97%
558	452	0.25%	9,397	5.22%
561	2,019	1.12%	11,416	6.35%
562	507	0.28%	11,923	6.63%
564	2,239	1.24%	14,162	7.87%
565	591	0.33%	14,753	8.20%
567	2,423	1.35%	17,176	9.55%
568	658	0.37%	17,834	9.91%
570	2,629	1.46%	20,463	11.4%
571	690	0.38%	21,153	11.8%
573	2,800	1.56%	23,953	13.3%
574	801	0.45%	24,754	13.8%
575	2,941	1.63%	27,695	15.4%
576	809	0.45%	28,504	15.8%
577	3,252	1.81%	31,756	17.7%
578	828	0.46%	32,584	18.1%
579	3,184	1.77%	35,768	19.9%
580	887	0.49%	36,655	20.4%
582	3,463	1.92%	40,118	22.3%
583	978	0.54%	41,096	22.8%
584	3,712	2.06%	44,808	24.9%
585	1,007	0.56%	45,815	25.5%
586	3,924	2.18%	49,739	27.6%
587	1,068	0.59%	50,807	28.2%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
588	4,094	2.28%	54,901	30.5%
589	1,151	0.64%	56,052	31.2%
590	4,399	2.45%	60,451	33.6%
591	5,860	3.26%	66,311	36.9%
592	1,233	0.69%	67,544	37.5%
593	5,074	2.82%	72,618	40.4%
594	1,341	0.75%	73,959	41.1%
595	5,275	2.93%	79,234	44.0%
596	1,380	0.77%	80,614	44.8%
597	5,504	3.06%	86,118	47.9%
598	1,515	0.84%	87,633	48.7%
600	5,896	3.28%	93,529	52.0%
601	1,454	0.81%	94,983	52.8%
602	6,323	3.51%	101,306	56.3%
603	1,529	0.85%	102,835	57.2%
604	6,436	3.58%	109,271	60.7%
605	1,653	0.92%	110,924	61.7%
606	6,702	3.73%	117,626	65.4%
607	1,719	0.96%	119,345	66.3%
609	6,795	3.78%	126,140	70.1%
610	1,658	0.92%	127,798	71.0%
614	6,815	3.79%	134,613	74.8%
615	8,430	4.69%	143,043	79.5%
616	1,632	0.91%	144,675	80.4%
618	6,585	3.66%	151,260	84.1%
619	1,529	0.85%	152,789	84.9%
622	6,053	3.36%	158,842	88.3%
623	1,394	0.77%	160,236	89.1%
626	5,532	3.07%	165,768	92.1%
627	1,196	0.66%	166,964	92.8%
631	4,492	2.50%	171,456	95.3%
632	919	0.51%	172,375	95.8%
637	3,333	1.85%	175,708	97.7%
638	636	0.35%	176,344	98.0%
646	2,002	1.11%	178,346	99.1%
647	290	0.16%	178,636	99.3%
651	902	0.50%	179,538	99.8%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
652	135	0.08%	179,673	99.9%
656	235	0.13%	179,908	100%

Table Q.29. ELA Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
510	14	0.01%	14	0.01%
511	3	0.00%	17	0.01%
515	15	0.01%	32	0.02%
516	2	0.00%	34	0.02%
519	45	0.03%	79	0.05%
520	4	0.00%	83	0.05%
524	81	0.05%	164	0.10%
525	16	0.01%	180	0.11%
528	174	0.10%	354	0.21%
529	39	0.02%	393	0.23%
533	298	0.17%	691	0.41%
534	38	0.02%	729	0.43%
537	444	0.26%	1,173	0.69%
538	87	0.05%	1,260	0.74%
542	596	0.35%	1,856	1.09%
543	122	0.07%	1,978	1.16%
547	767	0.45%	2,745	1.61%
548	169	0.10%	2,914	1.71%
551	891	0.52%	3,805	2.23%
552	230	0.13%	4,035	2.37%
556	1,118	0.66%	5,153	3.02%
557	247	0.14%	5,400	3.17%
559	1,212	0.71%	6,612	3.88%
560	327	0.19%	6,939	4.07%
562	1,409	0.83%	8,348	4.89%
563	382	0.22%	8,730	5.12%
565	1,582	0.93%	10,312	6.04%
566	426	0.25%	10,738	6.29%
568	1,709	1.00%	12,447	7.30%
569	459	0.27%	12,906	7.57%
570	1,950	1.14%	14,856	8.71%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
571	509	0.30%	15,365	9.01%
572	2,040	1.20%	17,405	10.2%
573	543	0.32%	17,948	10.5%
574	2,256	1.32%	20,204	11.8%
575	603	0.35%	20,807	12.2%
576	2,395	1.40%	23,202	13.6%
577	652	0.38%	23,854	14.0%
578	2,584	1.51%	26,438	15.5%
579	678	0.40%	27,116	15.9%
580	2,760	1.62%	29,876	17.5%
581	727	0.43%	30,603	17.9%
582	3,030	1.78%	33,633	19.7%
583	790	0.46%	34,423	20.2%
584	3,195	1.87%	37,618	22.1%
585	837	0.49%	38,455	22.5%
586	3,342	1.96%	41,797	24.5%
587	895	0.52%	42,692	25.0%
588	3,590	2.10%	46,282	27.1%
589	4,871	2.86%	51,153	30.0%
590	1,026	0.60%	52,179	30.6%
591	4,395	2.58%	56,574	33.2%
592	1,119	0.66%	57,693	33.8%
593	4,437	2.60%	62,130	36.4%
594	1,223	0.72%	63,353	37.1%
595	4,668	2.74%	68,021	39.9%
596	1,212	0.71%	69,233	40.6%
597	4,838	2.84%	74,071	43.4%
598	6,387	3.74%	80,458	47.2%
599	1,307	0.77%	81,765	47.9%
600	5,343	3.13%	87,108	51.1%
601	1,310	0.77%	88,418	51.8%
602	5,608	3.29%	94,026	55.1%
603	1,437	0.84%	95,463	56.0%
604	5,711	3.35%	101,174	59.3%
605	1,402	0.82%	102,576	60.1%
607	5,846	3.43%	108,422	63.6%
608	1,464	0.86%	109,886	64.4%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
609	5,940	3.48%	115,826	67.9%
610	1,450	0.85%	117,276	68.7%
611	6,023	3.53%	123,299	72.3%
612	1,360	0.80%	124,659	73.1%
613	5,925	3.47%	130,584	76.5%
614	1,403	0.82%	131,987	77.4%
616	5,946	3.49%	137,933	80.9%
617	1,251	0.73%	139,184	81.6%
619	5,592	3.28%	144,776	84.9%
620	1,123	0.66%	145,899	85.5%
623	5,288	3.10%	151,187	88.6%
624	984	0.58%	152,171	89.2%
626	4,720	2.77%	156,891	92.0%
627	801	0.47%	157,692	92.4%
630	3,980	2.33%	161,672	94.8%
631	607	0.36%	162,279	95.1%
635	3,219	1.89%	165,498	97.0%
636	399	0.23%	165,897	97.2%
641	2,318	1.36%	168,215	98.6%
642	276	0.16%	168,491	98.8%
652	1,418	0.83%	169,909	99.6%
653	144	0.08%	170,053	99.7%
657	542	0.32%	170,595	100%

Table Q.30. ELA Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
507	24	0.02%	24	0.02%
508	6	0.00%	30	0.02%
511	11	0.01%	41	0.03%
512	3	0.00%	44	0.03%
516	22	0.01%	66	0.04%
517	6	0.00%	72	0.05%
520	54	0.03%	126	0.08%
521	7	0.00%	133	0.09%
525	114	0.07%	247	0.16%
526	21	0.01%	268	0.17%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
530	163	0.10%	431	0.28%
531	41	0.03%	472	0.30%
534	278	0.18%	750	0.48%
535	53	0.03%	803	0.51%
539	407	0.26%	1,210	0.77%
540	84	0.05%	1,294	0.83%
543	517	0.33%	1,811	1.16%
544	101	0.06%	1,912	1.22%
548	664	0.42%	2,576	1.65%
549	148	0.09%	2,724	1.74%
553	750	0.48%	3,474	2.22%
554	192	0.12%	3,666	2.35%
557	937	0.60%	4,603	2.94%
558	222	0.14%	4,825	3.09%
560	1,081	0.69%	5,906	3.78%
561	260	0.17%	6,166	3.94%
563	1,137	0.73%	7,303	4.67%
564	293	0.19%	7,596	4.86%
566	1,402	0.90%	8,998	5.76%
567	346	0.22%	9,344	5.98%
568	1,483	0.95%	10,827	6.93%
569	381	0.24%	11,208	7.17%
570	1,699	1.09%	12,907	8.26%
571	400	0.26%	13,307	8.51%
572	1,739	1.11%	15,046	9.63%
573	448	0.29%	15,494	9.91%
574	1,932	1.24%	17,426	11.1%
575	467	0.30%	17,893	11.4%
576	2,009	1.29%	19,902	12.7%
577	516	0.33%	20,418	13.1%
578	2,111	1.35%	22,529	14.4%
579	541	0.35%	23,070	14.8%
580	2,401	1.54%	25,471	16.3%
581	612	0.39%	26,083	16.7%
582	2,437	1.56%	28,520	18.2%
583	625	0.40%	29,145	18.6%
584	2,793	1.79%	31,938	20.4%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
585	3,613	2.31%	35,551	22.7%
586	715	0.46%	36,266	23.2%
587	3,092	1.98%	39,358	25.2%
588	698	0.45%	40,056	25.6%
589	3,392	2.17%	43,448	27.8%
590	4,385	2.81%	47,833	30.6%
591	857	0.55%	48,690	31.2%
592	3,816	2.44%	52,506	33.6%
593	882	0.56%	53,388	34.2%
594	4,112	2.63%	57,500	36.8%
595	5,265	3.37%	62,765	40.2%
596	1,012	0.65%	63,777	40.8%
597	4,607	2.95%	68,384	43.8%
598	1,061	0.68%	69,445	44.4%
599	4,800	3.07%	74,245	47.5%
600	1,057	0.68%	75,302	48.2%
601	5,236	3.35%	80,538	51.5%
602	1,150	0.74%	81,688	52.3%
603	5,388	3.45%	87,076	55.7%
604	1,203	0.77%	88,279	56.5%
605	5,794	3.71%	94,073	60.2%
606	1,211	0.77%	95,284	61.0%
607	5,965	3.82%	101,249	64.8%
608	1,192	0.76%	102,441	65.5%
609	6,150	3.93%	108,591	69.5%
610	1,228	0.79%	109,819	70.3%
612	6,375	4.08%	116,194	74.3%
613	1,227	0.79%	117,421	75.1%
614	6,309	4.04%	123,730	79.2%
615	1,160	0.74%	124,890	79.9%
617	6,109	3.91%	130,999	83.8%
618	1,121	0.72%	132,120	84.5%
620	5,767	3.69%	137,887	88.2%
621	948	0.61%	138,835	88.8%
624	5,271	3.37%	144,106	92.2%
625	824	0.53%	144,930	92.7%
629	4,306	2.75%	149,236	95.5%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
630	561	0.36%	149,797	95.8%
635	3,281	2.10%	153,078	97.9%
636	378	0.24%	153,456	98.2%
646	1,955	1.25%	155,411	99.4%
647	167	0.11%	155,578	99.5%
651	726	0.46%	156,304	100%

Table Q.31. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
524	15	0.01%	15	0.01%
529	33	0.02%	48	0.03%
531	2	0.00%	50	0.03%
533	98	0.05%	148	0.08%
535	12	0.01%	160	0.09%
538	248	0.13%	408	0.22%
540	32	0.02%	440	0.24%
543	580	0.31%	1,020	0.55%
545	101	0.05%	1,121	0.61%
547	1,126	0.61%	2,247	1.22%
549	191	0.10%	2,438	1.32%
552	1,647	0.89%	4,085	2.21%
554	264	0.14%	4,349	2.36%
557	2,156	1.17%	6,505	3.52%
559	396	0.21%	6,901	3.74%
563	2,578	1.40%	9,479	5.14%
565	482	0.26%	9,961	5.40%
568	2,909	1.58%	12,870	6.97%
570	515	0.28%	13,385	7.25%
571	3,215	1.74%	16,600	8.99%
573	538	0.29%	17,138	9.29%
574	3,216	1.74%	20,354	11.0%
576	545	0.30%	20,899	11.3%
577	3,242	1.76%	24,141	13.1%
579	3,958	2.14%	28,099	15.2%
581	4,084	2.21%	32,183	17.4%
583	4,105	2.22%	36,288	19.7%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
585	4,251	2.30%	40,539	22.0%
587	4,334	2.35%	44,873	24.3%
588	3,798	2.06%	48,671	26.4%
589	632	0.34%	49,303	26.7%
590	4,599	2.49%	53,902	29.2%
591	4,108	2.23%	58,010	31.4%
592	664	0.36%	58,674	31.8%
593	4,960	2.69%	63,634	34.5%
594	4,433	2.40%	68,067	36.9%
595	668	0.36%	68,735	37.2%
596	5,275	2.86%	74,010	40.1%
597	4,713	2.55%	78,723	42.7%
598	5,419	2.94%	84,142	45.6%
599	705	0.38%	84,847	46.0%
600	5,668	3.07%	90,515	49.0%
601	5,131	2.78%	95,646	51.8%
602	727	0.39%	96,373	52.2%
603	6,005	3.25%	102,378	55.5%
604	5,478	2.97%	107,856	58.4%
605	733	0.40%	108,589	58.8%
606	6,264	3.39%	114,853	62.2%
607	5,640	3.06%	120,493	65.3%
608	755	0.41%	121,248	65.7%
609	6,430	3.48%	127,678	69.2%
611	6,546	3.55%	134,224	72.7%
613	6,652	3.60%	140,876	76.3%
615	6,643	3.60%	147,519	79.9%
617	6,631	3.59%	154,150	83.5%
619	6,567	3.56%	160,717	87.1%
621	610	0.33%	161,327	87.4%
622	5,668	3.07%	166,995	90.5%
624	523	0.28%	167,518	90.8%
626	5,441	2.95%	172,959	93.7%
628	365	0.20%	173,324	93.9%
631	4,898	2.65%	178,222	96.6%
633	306	0.17%	178,528	96.7%
640	3,687	2.00%	182,215	98.7%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
642	167	0.09%	182,382	98.8%
645	2,194	1.19%	184,576	100%

Table Q.32. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
521	9	0.00%	9	0.00%
526	20	0.01%	29	0.02%
527	1	0.00%	30	0.02%
530	52	0.03%	82	0.04%
531	6	0.00%	88	0.05%
535	144	0.08%	232	0.12%
536	20	0.01%	252	0.13%
539	404	0.21%	656	0.35%
540	59	0.03%	715	0.38%
544	792	0.42%	1,507	0.80%
545	108	0.06%	1,615	0.86%
549	1,348	0.72%	2,963	1.57%
550	186	0.10%	3,149	1.67%
553	1,902	1.01%	5,051	2.68%
554	247	0.13%	5,298	2.82%
558	2,455	1.30%	7,753	4.12%
559	285	0.15%	8,038	4.27%
564	2,851	1.52%	10,889	5.79%
565	376	0.20%	11,265	5.99%
568	3,005	1.60%	14,270	7.58%
569	396	0.21%	14,666	7.80%
571	3,143	1.67%	17,809	9.47%
572	413	0.22%	18,222	9.69%
574	3,158	1.68%	21,380	11.4%
575	385	0.20%	21,765	11.6%
577	3,110	1.65%	24,875	13.2%
578	366	0.19%	25,241	13.4%
579	3,127	1.66%	28,368	15.1%
580	473	0.25%	28,841	15.3%
581	3,313	1.76%	32,154	17.1%
582	415	0.22%	32,569	17.3%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
583	3,457	1.84%	36,026	19.1%
584	3,846	2.04%	39,872	21.2%
585	505	0.27%	40,377	21.5%
586	3,373	1.79%	43,750	23.3%
587	458	0.24%	44,208	23.5%
588	3,514	1.87%	47,722	25.4%
589	4,211	2.24%	51,933	27.6%
590	4,189	2.23%	56,122	29.8%
591	4,382	2.33%	60,504	32.2%
592	557	0.30%	61,061	32.5%
593	3,836	2.04%	64,897	34.5%
594	4,555	2.42%	69,452	36.9%
595	4,629	2.46%	74,081	39.4%
596	592	0.31%	74,673	39.7%
597	4,207	2.24%	78,880	41.9%
598	4,921	2.62%	83,801	44.5%
599	5,024	2.67%	88,825	47.2%
600	5,144	2.73%	93,969	49.9%
601	700	0.37%	94,669	50.3%
602	4,624	2.46%	99,293	52.8%
603	5,276	2.80%	104,569	55.6%
604	5,547	2.95%	110,116	58.5%
605	660	0.35%	110,776	58.9%
606	4,913	2.61%	115,689	61.5%
607	5,673	3.02%	121,362	64.5%
608	691	0.37%	122,053	64.9%
609	5,175	2.75%	127,228	67.6%
610	6,089	3.24%	133,317	70.9%
611	690	0.37%	134,007	71.2%
612	5,254	2.79%	139,261	74.0%
613	718	0.38%	139,979	74.4%
614	5,525	2.94%	145,504	77.3%
615	686	0.36%	146,190	77.7%
616	5,625	2.99%	151,815	80.7%
617	705	0.37%	152,520	81.1%
618	5,594	2.97%	158,114	84.0%
619	686	0.36%	158,800	84.4%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
621	5,597	2.97%	164,397	87.4%
622	635	0.34%	165,032	87.7%
624	5,653	3.00%	170,685	90.7%
625	542	0.29%	171,227	91.0%
628	5,422	2.88%	176,649	93.9%
629	441	0.23%	177,090	94.1%
633	4,795	2.55%	181,885	96.7%
634	368	0.20%	182,253	96.9%
643	3,654	1.94%	185,907	98.8%
644	185	0.10%	186,092	98.9%
648	2,051	1.09%	188,143	100%

Table Q.33. Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
527	15	0.01%	15	0.01%
529	1	0.00%	16	0.01%
532	59	0.03%	75	0.04%
534	17	0.01%	92	0.05%
536	171	0.09%	263	0.14%
538	34	0.02%	297	0.16%
541	441	0.24%	738	0.41%
543	85	0.05%	823	0.45%
546	900	0.50%	1,723	0.95%
548	150	0.08%	1,873	1.03%
550	1,534	0.84%	3,407	1.87%
552	271	0.15%	3,678	2.02%
555	2,030	1.12%	5,708	3.14%
557	323	0.18%	6,031	3.32%
559	2,602	1.43%	8,633	4.75%
561	389	0.21%	9,022	4.96%
565	2,925	1.61%	11,947	6.57%
567	452	0.25%	12,399	6.82%
569	3,040	1.67%	15,439	8.49%
571	453	0.25%	15,892	8.74%
572	3,240	1.78%	19,132	10.5%
574	472	0.26%	19,604	10.8%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
575	3,276	1.80%	22,880	12.6%
577	3,818	2.10%	26,698	14.7%
579	463	0.25%	27,161	14.9%
580	3,325	1.83%	30,486	16.8%
581	3,440	1.89%	33,926	18.7%
582	493	0.27%	34,419	18.9%
583	3,958	2.18%	38,377	21.1%
585	4,052	2.23%	42,429	23.3%
586	3,592	1.98%	46,021	25.3%
587	584	0.32%	46,605	25.6%
588	4,215	2.32%	50,820	28.0%
589	3,635	2.00%	54,455	30.0%
590	626	0.34%	55,081	30.3%
591	4,341	2.39%	59,422	32.7%
592	3,808	2.09%	63,230	34.8%
593	4,390	2.42%	67,620	37.2%
594	615	0.34%	68,235	37.5%
595	4,292	2.36%	72,527	39.9%
596	3,783	2.08%	76,310	42.0%
597	4,358	2.40%	80,668	44.4%
598	4,446	2.45%	85,114	46.8%
599	583	0.32%	85,697	47.1%
600	4,379	2.41%	90,076	49.6%
601	3,827	2.11%	93,903	51.7%
602	4,487	2.47%	98,390	54.1%
603	614	0.34%	99,004	54.5%
604	4,552	2.50%	103,556	57.0%
605	3,998	2.20%	107,554	59.2%
606	4,816	2.65%	112,370	61.8%
607	4,809	2.65%	117,179	64.5%
608	667	0.37%	117,846	64.8%
609	4,832	2.66%	122,678	67.5%
610	4,353	2.39%	127,031	69.9%
611	618	0.34%	127,649	70.2%
612	4,910	2.70%	132,559	72.9%
613	4,531	2.49%	137,090	75.4%
614	643	0.35%	137,733	75.8%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
615	597	0.33%	138,330	76.1%
616	4,677	2.57%	143,007	78.7%
617	4,875	2.68%	147,882	81.4%
618	671	0.37%	148,553	81.7%
619	5,662	3.11%	154,215	84.8%
621	5,536	3.05%	159,751	87.9%
623	588	0.32%	160,339	88.2%
624	5,033	2.77%	165,372	91.0%
626	498	0.27%	165,870	91.3%
628	4,859	2.67%	170,729	93.9%
630	431	0.24%	171,160	94.2%
634	4,454	2.45%	175,614	96.6%
636	348	0.19%	175,962	96.8%
644	3,537	1.95%	179,499	98.8%
646	230	0.13%	179,729	98.9%
649	2,042	1.12%	181,771	100%

Table Q.34. Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
529	19	0.01%	19	0.01%
531	1	0.00%	20	0.01%
533	41	0.02%	61	0.03%
535	2	0.00%	63	0.04%
538	135	0.08%	198	0.11%
540	16	0.01%	214	0.12%
542	366	0.20%	580	0.32%
544	56	0.03%	636	0.35%
547	878	0.49%	1,514	0.84%
549	155	0.09%	1,669	0.93%
552	1,659	0.92%	3,328	1.85%
554	257	0.14%	3,585	2.00%
556	2,568	1.43%	6,153	3.43%
558	418	0.23%	6,571	3.66%
561	3,498	1.95%	10,069	5.61%
563	566	0.32%	10,635	5.92%
565	4,344	2.42%	14,979	8.34%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
567	679	0.38%	15,658	8.72%
573	4,700	2.62%	20,358	11.3%
575	762	0.42%	21,120	11.8%
578	4,696	2.61%	25,816	14.4%
580	744	0.41%	26,560	14.8%
581	4,487	2.50%	31,047	17.3%
583	740	0.41%	31,787	17.7%
584	4,171	2.32%	35,958	20.0%
586	4,697	2.62%	40,655	22.6%
588	4,497	2.50%	45,152	25.1%
589	3,678	2.05%	48,830	27.2%
590	637	0.35%	49,467	27.5%
591	4,240	2.36%	53,707	29.9%
592	3,453	1.92%	57,160	31.8%
593	4,012	2.23%	61,172	34.1%
594	3,887	2.16%	65,059	36.2%
595	593	0.33%	65,652	36.6%
596	3,890	2.17%	69,542	38.7%
597	3,233	1.80%	72,775	40.5%
598	3,831	2.13%	76,606	42.7%
599	3,871	2.16%	80,477	44.8%
600	3,863	2.15%	84,340	47.0%
601	3,894	2.17%	88,234	49.1%
602	3,924	2.18%	92,158	51.3%
603	3,973	2.21%	96,131	53.5%
604	3,903	2.17%	100,034	55.7%
605	4,025	2.24%	104,059	57.9%
606	4,207	2.34%	108,266	60.3%
607	4,047	2.25%	112,313	62.5%
608	4,155	2.31%	116,468	64.8%
609	4,389	2.44%	120,857	67.3%
610	4,288	2.39%	125,145	69.7%
611	654	0.36%	125,799	70.0%
612	4,282	2.38%	130,081	72.4%
613	3,728	2.08%	133,809	74.5%
614	4,461	2.48%	138,270	77.0%
615	652	0.36%	138,922	77.3%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
616	4,348	2.42%	143,270	79.8%
617	3,756	2.09%	147,026	81.9%
618	635	0.35%	147,661	82.2%
619	4,385	2.44%	152,046	84.7%
621	4,448	2.48%	156,494	87.1%
623	4,296	2.39%	160,790	89.5%
625	4,255	2.37%	165,045	91.9%
627	470	0.26%	165,515	92.2%
628	3,519	1.96%	169,034	94.1%
630	394	0.22%	169,428	94.3%
631	3,313	1.84%	172,741	96.2%
633	328	0.18%	173,069	96.4%
636	2,791	1.55%	175,860	97.9%
638	231	0.13%	176,091	98.0%
644	2,189	1.22%	178,280	99.3%
646	122	0.07%	178,402	99.3%
649	1,209	0.67%	179,611	100%

Table Q.35. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
520	29	0.02%	29	0.02%
522	1	0.00%	30	0.02%
525	30	0.02%	60	0.04%
527	2	0.00%	62	0.04%
529	54	0.03%	116	0.07%
531	12	0.01%	128	0.08%
534	197	0.12%	325	0.19%
536	22	0.01%	347	0.21%
538	418	0.25%	765	0.45%
540	58	0.03%	823	0.49%
543	888	0.53%	1,711	1.01%
545	159	0.09%	1,870	1.11%
548	1,457	0.86%	3,327	1.97%
550	270	0.16%	3,597	2.13%
552	2,333	1.38%	5,930	3.51%
554	371	0.22%	6,301	3.73%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
557	3,070	1.82%	9,371	5.55%
559	530	0.31%	9,901	5.86%
569	3,759	2.23%	13,660	8.09%
571	604	0.36%	14,264	8.44%
575	4,116	2.44%	18,380	10.9%
577	685	0.41%	19,065	11.3%
579	4,338	2.57%	23,403	13.9%
581	707	0.42%	24,110	14.3%
582	4,367	2.59%	28,477	16.9%
584	4,904	2.90%	33,381	19.8%
586	4,741	2.81%	38,122	22.6%
588	4,471	2.65%	42,593	25.2%
590	4,192	2.48%	46,785	27.7%
591	3,473	2.06%	50,258	29.8%
592	616	0.36%	50,874	30.1%
593	3,946	2.34%	54,820	32.5%
594	3,186	1.89%	58,006	34.3%
595	3,624	2.15%	61,630	36.5%
596	3,573	2.12%	65,203	38.6%
597	3,467	2.05%	68,670	40.7%
598	3,208	1.90%	71,878	42.6%
599	3,258	1.93%	75,136	44.5%
600	3,111	1.84%	78,247	46.3%
601	3,114	1.84%	81,361	48.2%
602	3,135	1.86%	84,496	50.0%
603	3,053	1.81%	87,549	51.8%
604	5,549	3.29%	93,098	55.1%
605	3,132	1.85%	96,230	57.0%
606	3,650	2.16%	99,880	59.1%
607	3,154	1.87%	103,034	61.0%
608	3,080	1.82%	106,114	62.8%
609	3,207	1.90%	109,321	64.7%
610	3,152	1.87%	112,473	66.6%
611	3,359	1.99%	115,832	68.6%
612	3,460	2.05%	119,292	70.6%
613	3,486	2.06%	122,778	72.7%
614	3,630	2.15%	126,408	74.8%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
615	3,797	2.25%	130,205	77.1%
616	3,990	2.36%	134,195	79.4%
617	548	0.32%	134,743	79.8%
618	3,965	2.35%	138,708	82.1%
620	4,238	2.51%	142,946	84.6%
621	3,848	2.28%	146,794	86.9%
622	550	0.33%	147,344	87.2%
623	521	0.31%	147,865	87.5%
624	3,996	2.37%	151,861	89.9%
626	505	0.30%	152,366	90.2%
627	4,130	2.45%	156,496	92.7%
629	437	0.26%	156,933	92.9%
631	4,217	2.50%	161,150	95.4%
633	389	0.23%	161,539	95.6%
639	4,008	2.37%	165,547	98.0%
641	279	0.17%	165,826	98.2%
643	3,083	1.83%	168,909	100%

Table Q.36. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
520	14	0.01%	14	0.01%
522	4	0.00%	18	0.02%
525	26	0.02%	44	0.04%
527	4	0.00%	48	0.04%
530	59	0.05%	107	0.09%
532	6	0.01%	113	0.10%
534	168	0.14%	281	0.24%
536	22	0.02%	303	0.26%
539	372	0.32%	675	0.58%
541	56	0.05%	731	0.63%
543	769	0.66%	1,500	1.29%
545	143	0.12%	1,643	1.42%
548	1,311	1.13%	2,954	2.55%
550	275	0.24%	3,229	2.79%
552	1,987	1.71%	5,216	4.50%
554	342	0.30%	5,558	4.80%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
557	2,737	2.36%	8,295	7.16%
559	470	0.41%	8,765	7.56%
569	3,104	2.68%	11,869	10.2%
571	586	0.51%	12,455	10.7%
575	3,534	3.05%	15,989	13.8%
577	588	0.51%	16,577	14.3%
580	3,578	3.09%	20,155	17.4%
582	616	0.53%	20,771	17.9%
583	3,516	3.03%	24,287	21.0%
585	559	0.48%	24,846	21.4%
586	3,354	2.89%	28,200	24.3%
588	3,624	3.13%	31,824	27.5%
590	3,352	2.89%	35,176	30.4%
592	3,204	2.76%	38,380	33.1%
593	2,596	2.24%	40,976	35.4%
594	493	0.43%	41,469	35.8%
595	2,928	2.53%	44,397	38.3%
596	2,404	2.07%	46,801	40.4%
597	2,801	2.42%	49,602	42.8%
598	2,713	2.34%	52,315	45.1%
599	373	0.32%	52,688	45.5%
600	2,613	2.25%	55,301	47.7%
601	2,269	1.96%	57,570	49.7%
602	2,589	2.23%	60,159	51.9%
603	2,445	2.11%	62,604	54.0%
604	2,493	2.15%	65,097	56.2%
605	2,459	2.12%	67,556	58.3%
606	2,406	2.08%	69,962	60.4%
607	2,497	2.15%	72,459	62.5%
608	2,399	2.07%	74,858	64.6%
609	2,392	2.06%	77,250	66.7%
610	2,359	2.04%	79,609	68.7%
611	2,353	2.03%	81,962	70.7%
612	2,332	2.01%	84,294	72.7%
613	2,389	2.06%	86,683	74.8%
614	2,368	2.04%	89,051	76.8%
615	2,275	1.96%	91,326	78.8%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
616	2,254	1.95%	93,580	80.8%
617	2,132	1.84%	95,712	82.6%
618	2,195	1.89%	97,907	84.5%
619	232	0.20%	98,139	84.7%
620	2,101	1.81%	100,240	86.5%
622	2,098	1.81%	102,338	88.3%
623	1,933	1.67%	104,271	90.0%
624	167	0.14%	104,438	90.1%
625	1,971	1.70%	106,409	91.8%
627	1,941	1.67%	108,350	93.5%
629	112	0.10%	108,462	93.6%
630	1,693	1.46%	110,155	95.1%
632	95	0.08%	110,250	95.1%
634	1,749	1.51%	111,999	96.6%
636	90	0.08%	112,089	96.7%
639	1,522	1.31%	113,611	98.0%
641	55	0.05%	113,666	98.1%
648	1,272	1.10%	114,938	99.2%
650	27	0.02%	114,965	99.2%
653	921	0.79%	115,886	100%

Appendix R: Study of Operational Test Mode Comparability

Section 1: Introduction

1.1. Overview

Following the 2017 and 2018 administration, the New York State Education Department (NYSED) continued to offer its operational test (OP) in a computer-based testing (CBT) environment for the Grades 3–8 English Language Arts (ELA) and Mathematics tests in 2019. The schools had the option to administer the tests via paper-based testing (PBT) or computer-based testing (CBT). This study is to evaluate differences in test-level student performance that may be attributable to the mode in which a student tested.

In 2019, all the operational items administered in both PBT and CBT modes. The number of operational items and score points are summarized in Table 1.1.

Table 1.1. Operational Items Administered in Both CBT and PBT Modes

Grade	ELA				Math			
	MC Items	CR Items		Total Points	MC Items	CR Items		Total Points
		2-Point	4-Point			2-Point	3-Point	
3	18	6	1	34	27	6	1	42
4	18	6	1	34	31	6	1	46
5	28	6	1	44	31	6	1	46
6	28	6	1	44	31	7	1	48
7	28	7	1	46	33	7	1	50
8	28	7	1	46	33	7	1	50

The current study consists of two steps:

1. A propensity score matching approach was conducted to generate the CBT and PBT samples that were comparable on selected covariates that may affect student performance, aside from the test mode itself.
2. The difference in students' test scores were computed between the matched CBT and PBT samples to evaluate the test-level mode comparability.

Section 2: Method

2.1. Preparing Balanced Samples

2.1.1. *Overview*

While the ideal conditions under which to investigate test mode comparability would necessitate random assignment of schools to test in either the CBT or PBT modes, the practical constraints and resources of individual districts and schools preclude such designs. The next best solution is often referred to as a quasi-experimental design. Given that the student population was not randomly equivalent between test modes, the propensity score matching (PSM) methodology (Austin, 2011a; Rosenbaum, 2010) was applied to draw matched samples of PBT and CBT students who were considered comparable on average in their test performance. In other words, effective propensity score matching produces samples of PBT and CBT students that are on average otherwise comparable, with the only observed difference being that each sample tested in differing modes.

Table 1.2 shows the number of students in the clean datasets by test mode prior to propensity score matching. This study used the same data-cleaning procedures that have been used for operational psychometric analyses, with the following additional rules. In 2019, the CBT participation rates ranged from 15-20% for ELA and 12-15% for Math.

- For Grades 4–8, students without prior year scale scores in the same subject on the adjacent lower grade were removed.
- Because of sample size concerns and concerns about effects unrelated to test mode interfering with the study's inferences, students testing with the Braille or large print forms were dropped. Students who used a non-English language translation (i.e., Chinese, Korean, Haitian-Creole, Russian, Spanish) of a Mathematics form were also removed.

Table 1.2. Sample Sizes Before Matching by Test Mode

Grade	ELA				Math			
	PBT_N	CBT_N	PBT %	CBT %	PBT_N	CBT_N	PBT %	CBT %
3	150,388	27,115	84.72	15.28	159,274	22,128	87.80	12.20
4	152,010	28,299	84.31	15.69	161,135	21,198	88.37	11.63
5	141,544	33,839	80.71	19.29	148,983	21,975	87.15	12.85
6	141,634	34,548	80.39	19.61	144,992	24,435	85.58	14.42
7	128,552	31,360	80.39	19.61	136,764	23,151	85.52	14.48
8	126,088	25,878	82.97	17.03	93,095	13,994	86.93	13.07

Note. Sample sizes indicate the number of students who took at least one item administered in the test after the data cleaning used for operational psychometric analyses.

2.1.2. *Propensity Score Models and Matching*

In discussion with New York State's Assessment Technical Advisory Committee (TAC), the decision was made to model the propensity score at the student level for CBT testing. The decision to adopt CBT was a school-level decision and modeling it at the student level violates one part of the assumption of strong ignorability (Rosenbaum and Rubin, 1983), meaning that some students had probabilities of assignment to CBT that equaled zero or one. By conditioning on student-level and school-level covariates, Questar was able to best approximate the selection

process one might observe if students were able to self-select and therefore treat school assignment as something that was ignorable.

The propensity score matching process used a within-caliper matching approach; the caliper width was defined as 0.02 times the standard deviation of the propensity scores. This fine caliper was chosen because it did not cause a reduction in the number of matches while it provided a good balance between the matched samples. The matching procedure was a one-to-one match without replacement (Austin 2011a).

2.1.3. *Matching Covariates*

The propensity scores were calculated using logistic regression based on a list of selected matching covariates. The covariates included students' prior year scale score, which is the most predictive of current year's test performance. In addition, some key student-, school-, and district-level variables were selected.

The following covariates were used for estimating the propensity scores, and directly balanced throughout the process of propensity score matching:

- student prior year (grade $n - 1$) scale score;
- student gender;
- student racial/ethnic category;
- student English Language Learner/Multilingual Learner (ELL/MLL) status;
- student disability (SWD) status;
- school-type (i.e., public, charter, non-public) *;
- district-level needs/resource capacity (NRC) code; and
- district-level region as specified by the joint management team definitions (JMT).

**Note*: The school-type variable was dropped in the logistic regression analysis due to its high correlation with NRC code

2.1.4. *Judging Covariate Balance*

The covariate balance between the matched PBT and CBT samples was evaluated after propensity score matching. The standardized difference (d) for each covariate between matched CBT and PBT groups was computed to evaluate the balance and effectiveness of propensity score matching. To the extent that the standardized differences approach zero, balance can be said to be reasonably achieved on the selected covariates.

The formulae of d are different for continuous and discrete variables; there are minor modifications for estimating covariate balance before and after matching samples (Rosenbaum, 2010). The traditional experimental design is still a useful framework for this comparability study, so CBT can be considered the "treatment" and PBT can be considered the "control" condition. The analysis of covariate balance for discrete variables differs in that it uses the unbiased variance estimator for a proportion. (See page 174 of Austin (2011a) for examples of a similar but not identical formula.)

For variable k being treated as **continuous**:

1. Estimate the means and variances for the treatment ($\bar{x}_{tk}$ and s^2_{tk}) and control groups ($\bar{x}_{ck}$ and s^2_{ck}) before matching.
2. Estimate the means only for the treatment ($\bar{x}_{tmk}$) and control groups ($\bar{x}_{cmk}$) after matching.
3. Estimate the standardized difference for variable *k* **before** matching as:

$$d_{bk} = (\bar{x}_{tk} - \bar{x}_{ck}) / \sqrt{(s^2_{tk} + s^2_{ck})/2} \quad (1)$$

4. Estimate the standardized difference for variable *k*—note the use of the pre-matched pooled standard deviation in the denominator—**after** matching as:

$$d_{mk} = (\bar{x}_{tmk} - \bar{x}_{cmk}) / \sqrt{(s^2_{tk} + s^2_{ck})/2} \quad (2)$$

For variable *k* being treated as **discrete**:

1. Estimate the proportions for the treatment (p_{tk}) and control groups (p_{ck}) before matching.
2. Estimate the proportions for the treatment (p_{tmk}) and control groups (p_{cmk}) after matching.
3. Estimate the standardized difference for variable *k* **before** matching as:

$$d_{bk} = (p_{tk} - p_{ck}) / \sqrt{[p_{tk}(1 - p_{tk}) + p_{ck}(1 - p_{ck})]/2} \quad (3)$$

4. Estimate the standardized difference for variable *k*—note the use of the pre-matched pooled standard deviation in the denominator—**after** matching as:

$$d_{mk} = (p_{tmk} - p_{cmk}) / \sqrt{[p_{tk}(1 - p_{tk}) + p_{ck}(1 - p_{ck})]/2} \quad (4)$$

2.2. Evaluating Test-level Mode Comparability

In order to evaluate and detect test-level mode effects, two things were examined after propensity score matching CBT and PBT students. Given that the items were the same between the test modes, the distribution of raw scores for the matched PBT and CBT samples was first reviewed. This enabled a direct means of detecting possible mode effects.

Next, the distribution of scale scores for the matched PBT and CBT samples were reviewed. The scale scores were derived using the single operational raw-score-to-scale-score (RSSS) conversion table, which was estimated based on all students in the operational calibration sample. The mode treatment effect was simply calculated as the difference in scale score means for the matched PBT and CBT samples.

Section 3: Results

3.1. Propensity Score Matching

3.1.1. *Covariate Balance*

This study summarizes the covariate balance before and after matching. The covariate balance was summarized in Tables 3.1–3.10.

Very few covariates were flagged for having a standardized difference greater than 0.05. The standardized difference for the covariate of prior year scale score between matched samples were all at or below 0.05 across the grades and subjects. In general, the propensity score matching generated well-matched PBT and CBT groups.

Table 3.1. Covariate Balance Before and After Matching: ELA Grade 4

		Before Matching				After Matching			
variable	value	CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	598.81	601.24	-2.43	-0.13	599.13	599.19	-0.06	0.00
Disability (%)	No	87.75	83.76	3.98	0.11	87.72	87.66	0.06	0.00
	Yes	12.25	16.24	-3.99	-0.11	12.28	12.34	-0.06	-0.00
ELL (%)	No	97.32	89.01	8.32	0.33	97.34	96.88	0.46	0.02
	Yes	2.68	11	-8.32	-0.33	2.66	3.12	-0.46	-0.02
Ethnicity (%)	Missing	0.55	0	0.55	0.10	0	0	0	0.00
	Asian	4.25	11.28	-7.03	-0.26	4.69	4.46	0.23	0.01
	Black or African American (not Hispanic origin)	7.86	19.43	-11.57	-0.34	8.46	9.31	-0.85	-0.03
	Hispanic or Latino	13.47	31.39	-17.92	-0.44	14.35	13.36	0.99	0.02
	American Indian or Alaska Native	0.55	0.69	-0.14	-0.02	0.47	0.48	0	-0.00
	Multi-racial (not Hispanic origin)	4.24	2.69	1.55	0.08	4.57	3.63	0.94	0.05
	Native Hawaiian or Other Pacific Islander	0.07	0.27	-0.2	-0.05	0.07	0.06	0.02	0.00
	White (not Hispanic origin)	69	34.25	34.75	0.74	67.39	68.71	-1.32	-0.03
Gender (%)	Female	49.55	49.29	0.26	0.01	49.76	49.46	0.3	0.01
	Male	50.45	50.71	-0.26	-0.01	50.24	50.54	-0.3	-0.01
Joint Management Team Region (JMT)	New York City	0.8	49.75	-48.95	-1.36	0.93	1.04	-0.11	-0.00
	Long Island	5.43	11.54	-6.1	-0.22	5.94	6.65	-0.71	-0.03
	Lower Hudson Valley	9.63	6.69	2.94	0.11	11.28	9.26	2.02	0.07
	Mid-Hudson	10.22	3.29	6.93	0.28	11.51	11.56	-0.05	-0.00
	Capital District / North Country	19.19	5.21	13.98	0.44	21.12	20.01	1.11	0.03
	Central Region	2.08	2.68	-0.59	-0.04	2.24	1.52	0.72	0.05
	Mid-State	18.14	1.97	16.17	0.56	8.92	10.21	-1.29	-0.04
	Mid-South	6.07	2.34	3.73	0.19	7.23	6.51	0.71	0.04
	Mid-West	15.66	4.31	11.36	0.39	18.08	17.32	0.76	0.03
	West	9.99	6.23	3.76	0.14	11.9	14.96	-3.06	-0.11
Needs/Resource Category (NRC) (%)	New York	0.01	43.7	-43.68	-1.25	0.02	0.02	0	0.00
	Big 4 Cities	8.81	3.46	5.34	0.22	10.07	8.86	1.21	0.05
	Urban/Suburban	6.4	8.1	-1.7	-0.07	7.49	7.06	0.43	0.02
	High Needs Rural	14.13	3.8	10.33	0.37	14.36	15.02	-0.66	-0.02
	Average Needs	58.98	16.53	42.45	0.97	56.91	57.62	-0.71	-0.02
	Low Needs	7.75	10.89	-3.15	-0.11	8.99	8.92	0.06	0.00
	Charter School	1.06	7.4	-6.35	-0.32	1.23	1.39	-0.15	-0.01
	Religious and Independent Schools	2.86	6.11	-3.25	-0.16	0.94	1.11	-0.18	-0.01
School Type (%)	Public	96.08	86.49	9.59	0.35	97.83	97.5	0.33	0.01
	Charter	1.06	7.4	-6.35	-0.32	1.23	1.39	-0.15	-0.01
	Non-Public	2.86	6.11	-3.25	-0.16	0.94	1.11	-0.18	-0.01

Table 3.2 Covariate Balance Before and After Matching: ELA Grade 5

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	599.13	601.05	-1.92	-0.10	599.17	599.85	-0.68	-0.04
Disability (%)	No	87.99	82.92	5.07	0.14	88.22	87.85	0.36	0.01
	Yes	12.01	17.09	-5.08	-0.14	11.78	12.15	-0.36	-0.01
ELL (%)	No	97.41	89.95	7.46	0.31	97.43	97.93	-0.5	-0.02
	Yes	2.59	10.06	-7.47	-0.31	2.57	2.07	0.5	0.02
Ethnicity (%)	Missing	0.56	0	0.56	0.11	0	0	0	0.00
	Asian	4.66	11.95	-7.28	-0.27	5.29	5.83	-0.55	-0.02
	Black or African American (not Hispanic origin)	8.95	20.02	-11.07	-0.32	10.12	8.92	1.2	0.03
	Hispanic or Latino	15.54	31.33	-15.8	-0.38	16.16	12.92	3.24	0.08
	American Indian or Alaska Native	0.52	0.71	-0.19	-0.02	0.55	0.42	0.13	0.02
	Multi-racial (not Hispanic origin)	4.1	2.39	1.71	0.10	4.07	3.85	0.22	0.01
	Native Hawaiian or Other Pacific Islander	0.11	0.3	-0.19	-0.04	0.11	0.1	0.01	0.00
	White (not Hispanic origin)	65.57	33.32	32.26	0.68	63.7	67.96	-4.25	-0.09
Gender (%)	Female	49.19	49.11	0.08	0.00	49.44	49.08	0.36	0.01
	Male	50.81	50.9	-0.09	-0.00	50.56	50.92	-0.36	-0.01
Joint Management Team Region (JMT)	New York City	2.45	52.57	-50.12	-1.36	2.91	2.92	-0.01	-0.00
	Long Island	5.9	11.55	-5.65	-0.20	6.41	6.6	-0.19	-0.01
	Lower Hudson Valley	9.27	6.91	2.36	0.09	10.99	6.57	4.41	0.16
	Mid-Hudson	8.13	3.54	4.59	0.20	9.4	7.84	1.56	0.07
	Capital District / North Country	16.94	5.31	11.64	0.38	16.72	17.59	-0.88	-0.03
	Central Region	3.1	2.48	0.62	0.04	3.3	3.52	-0.22	-0.01
	Mid-State	16.08	1.87	14.21	0.51	8.38	7.48	0.9	0.03
	Mid-South	5.01	2.55	2.46	0.13	6.1	7.71	-1.61	-0.08
	Mid-West	14.18	4.07	10.11	0.36	13.63	15.85	-2.22	-0.08
	West	9.81	5.86	3.95	0.15	11.69	13.76	-2.07	-0.08
	Missing	9.11	3.29	5.82	0.24	10.47	10.17	0.31	0.01
Needs/Resource Category (NRC) (%)	New York	2.02	45.67	-43.65	-1.19	2.39	2.52	-0.13	-0.00
	Big 4 Cities	8.11	3.43	4.68	0.20	9.49	7.26	2.23	0.10
	Urban/Suburban	6.06	8.21	-2.15	-0.08	7.1	6.55	0.54	0.02
	High Needs Rural	13.09	3.68	9.4	0.34	13.47	13.7	-0.23	-0.01
	Average Needs	52.19	16.39	35.8	0.81	45.96	48.52	-2.56	-0.06
	Low Needs	8.75	10.97	-2.22	-0.07	10.31	10.5	-0.19	-0.01
	Charter School	0.67	8.37	-7.7	-0.38	0.81	0.78	0.03	0.00
	Religious and Independent Schools	9.11	3.29	5.82	0.24	10.47	10.17	0.31	0.01
School Type (%)	Public	90.22	88.35	1.86	0.06	88.72	89.05	-0.33	-0.01
	Charter	0.67	8.37	-7.7	-0.38	0.81	0.78	0.03	0.00
	Non-Public	9.11	3.29	5.82	0.24	10.47	10.17	0.31	0.01

Table 3.3 Covariate Balance Before and After Matching: ELA Grade 6

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	600.18	600.84	-0.65	-0.03	599.97	600.16	-0.19	-0.01
Disability (%)	No	87.52	83.43	4.09	0.12	87.3	87.67	-0.37	-0.01
	Yes	12.48	16.59	-4.1	-0.12	12.7	12.33	0.37	0.01
ELL (%)	No	97.33	90.27	7.06	0.30	96.93	97.07	-0.14	-0.01
	Yes	2.67	9.75	-7.07	-0.30	3.07	2.93	0.14	0.01
Ethnicity (%)	Missing	0.45	0	0.45	0.09	0	0	0	0.00
	Asian	4.52	12.23	-7.72	-0.28	5.71	6.37	-0.66	-0.02
	Black or African American (not Hispanic origin)	7.69	20.65	-12.95	-0.38	9.06	9.12	-0.06	-0.00
	Hispanic or Latino	15.01	31.49	-16.47	-0.40	16.28	15.57	0.7	0.02
	American Indian or Alaska Native	0.73	0.71	0.02	0.00	0.53	0.52	0.02	0.00
	Multi-racial (not Hispanic origin)	3.81	2.18	1.64	0.10	3.7	3.65	0.05	0.00
	Native Hawaiian or Other Pacific Islander	0.07	0.36	-0.29	-0.06	0.09	0.1	-0.02	-0.00
	White (not Hispanic origin)	67.72	32.41	35.31	0.75	64.64	64.66	-0.02	-0.00
Gender (%)	Female	48.86	49.29	-0.43	-0.01	49.27	48.79	0.48	0.01
	Male	51.14	50.73	0.42	0.01	50.73	51.21	-0.48	-0.01
Joint Management Team Region (JMT)	New York City	0.35	53.56	-53.21	-1.50	0.44	0.41	0.02	0.00
	Long Island	8.29	10.19	-1.9	-0.07	10	12.2	-2.2	-0.08
	Lower Hudson Valley	12.41	5.79	6.62	0.23	15.5	13.34	2.16	0.08
	Mid-Hudson	8.9	3.08	5.81	0.25	10.94	10.28	0.66	0.03
	Capital District / North Country	18.65	4.25	14.4	0.46	17.08	16.73	0.35	0.01
	Central Region	3.05	2.29	0.75	0.05	3.76	3.38	0.38	0.02
	Mid-State	13.65	2.09	11.57	0.44	7.49	10.12	-2.63	-0.10
	Mid-South	7.45	1.92	5.53	0.26	9.51	7.92	1.59	0.08
	Mid-West	15.07	3.41	11.66	0.41	12.23	12.39	-0.16	-0.01
	West	8.96	5.66	3.31	0.13	11.48	11.6	-0.12	-0.00
	Missing	3.23	7.78	-4.56	-0.20	1.56	1.61	-0.05	-0.00
Needs/Resource Category (NRC) (%)	New York	0.01	46.4	-46.4	-1.32	0	0	0	0.00
	Big 4 Cities	7.1	3.51	3.59	0.16	8.71	7.34	1.37	0.06
	Urban/Suburban	7.89	7.01	0.88	0.03	9.5	7.71	1.79	0.07
	High Needs Rural	12.76	3.44	9.31	0.35	15.06	13.94	1.11	0.04
	Average Needs	57.11	13.28	43.83	1.03	49.54	51.74	-2.2	-0.05
	Low Needs	11.34	9.99	1.35	0.04	14.93	16.89	-1.96	-0.06
	Charter School	0.57	8.59	-8.02	-0.39	0.71	0.76	-0.05	-0.00
	Religious and Independent Schools	3.23	7.78	-4.56	-0.20	1.56	1.61	-0.05	-0.00
School Type (%)	Public	96.2	83.65	12.55	0.43	97.74	97.63	0.1	0.00
	Charter	0.57	8.59	-8.02	-0.39	0.71	0.76	-0.05	-0.00
	Non-Public	3.23	7.78	-4.56	-0.20	1.56	1.61	-0.05	-0.00

Table 3.4 Covariate Balance Before and After Matching: ELA Grade 7

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	600.05	600.85	-0.80	-0.04	599.78	599.59	0.19	0.01
Disability (%)	No	87.13	83.57	3.56	0.10	86.9	87.09	-0.18	-0.01
	Yes	12.87	16.44	-3.57	-0.10	13.1	12.91	0.18	0.01
ELL (%)	No	97.35	91.26	6.09	0.27	97.27	97.32	-0.06	-0.00
	Yes	2.65	8.74	-6.1	-0.27	2.73	2.68	0.06	0.00
Ethnicity (%)	Missing	0.43	0	0.43	0.09	0	0	0	0.00
	Asian	4.81	12.84	-8.03	-0.29	5.89	5.64	0.25	0.01
	Black or African American (not Hispanic origin)	8.61	19.26	-10.65	-0.31	9.61	10.31	-0.71	-0.02
	Hispanic or Latino	15.08	30.62	-15.54	-0.38	14.78	14.95	-0.17	-0.00
	American Indian or Alaska Native	0.7	0.7	0	-0.00	0.43	0.45	-0.02	-0.00
	Multi-racial (not Hispanic origin)	3.61	1.95	1.67	0.10	3.38	3.48	-0.09	-0.01
	Native Hawaiian or Other Pacific Islander	0.11	0.42	-0.32	-0.06	0.11	0.11	0	0.00
	White (not Hispanic origin)	66.65	34.22	32.44	0.69	65.79	65.06	0.73	0.02
Gender (%)	Female	48.66	48.7	-0.04	-0.00	48.57	48.93	-0.36	-0.01
	Male	51.34	51.3	0.03	0.00	51.43	51.07	0.36	0.01
Joint Management Team Region (JMT)	New York City	0.19	51.97	-51.78	-1.46	0.24	0.33	-0.09	-0.00
	Long Island	8.71	10.07	-1.36	-0.05	10.29	11.57	-1.28	-0.04
	Lower Hudson Valley	11.78	6.28	5.5	0.19	12.04	12.05	-0.01	-0.00
	Mid-Hudson	10.93	2.86	8.07	0.32	12.36	10.04	2.32	0.09
	Capital District / North Country	17.34	5.16	12.18	0.39	18.02	17.71	0.32	0.01
	Central Region	3.08	2.37	0.7	0.04	3.45	2.87	0.58	0.04
	Mid-State	12.95	2.47	10.48	0.40	9.6	10.24	-0.64	-0.02
	Mid-South	7.49	2.04	5.45	0.26	8.14	7.84	0.3	0.01
	Mid-West	17.29	3.23	14.06	0.48	13.06	13.62	-0.57	-0.02
	West	8.78	5.91	2.86	0.11	10.86	11.82	-0.95	-0.04
	Missing	1.48	7.66	-6.18	-0.30	1.94	1.92	0.02	0.00
Needs/Resource Category (NRC) (%)	New York	0.01	50.15	-50.14	-1.42	0	0	0	0.00
	Big 4 Cities	8.64	3.12	5.52	0.24	8.14	8.91	-0.77	-0.03
	Urban/Suburban	6.96	7.56	-0.6	-0.02	8.59	8.11	0.49	0.02
	High Needs Rural	13.69	3.68	10.01	0.36	14.47	14.61	-0.14	-0.01
	Average Needs	58.05	13.31	44.73	1.06	52.73	53.08	-0.35	-0.01
	Low Needs	10.63	11.3	-0.68	-0.02	13.42	12.48	0.94	0.03
	Charter School	0.55	3.23	-2.68	-0.20	0.7	0.88	-0.18	-0.01
	Religious and Independent Schools	1.48	7.66	-6.18	-0.30	1.94	1.92	0.02	0.00
School Type (%)	Public	97.97	89.12	8.85	0.37	97.36	97.2	0.16	0.01
	Charter	0.55	3.23	-2.68	-0.20	0.7	0.88	-0.18	-0.01
	Non-Public	1.48	7.66	-6.18	-0.30	1.94	1.92	0.02	0.00

Table 3.5 Covariate Balance Before and After Matching: ELA Grade 8

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	599.29	601.24	-1.95	-0.10	599.08	599.02	0.07	0.00
Disability (%)	No	86.66	83.2	3.46	0.10	87.17	86.97	0.2	0.01
	Yes	13.34	16.8	-3.47	-0.10	12.83	13.03	-0.2	-0.01
ELL (%)	No	97.48	91.91	5.56	0.25	97.69	97.78	-0.09	-0.00
	Yes	2.52	8.09	-5.57	-0.25	2.31	2.22	0.09	0.00
Ethnicity (%)	Missing	0.37	0	0.37	0.09	0	0	0	0.00
	Asian	5	12.58	-7.58	-0.27	5.88	5.82	0.06	0.00
	Black or African American (not Hispanic origin)	8.66	21.61	-12.95	-0.37	9.39	8.78	0.61	0.02
	Hispanic or Latino	14.64	31.07	-16.43	-0.40	14.85	14.46	0.4	0.01
	American Indian or Alaska Native	0.58	0.84	-0.26	-0.03	0.55	0.64	-0.09	-0.01
	Multi-racial (not Hispanic origin)	2.93	1.69	1.25	0.08	2.73	2.77	-0.04	-0.00
	Native Hawaiian or Other Pacific Islander	0.1	0.33	-0.24	-0.05	0.1	0.1	0	0.00
	White (not Hispanic origin)	67.71	31.89	35.82	0.77	66.49	67.43	-0.94	-0.02
Gender (%)	Female	47.45	48.64	-1.19	-0.02	47.77	47.76	0.02	0.00
	Male	52.55	51.36	1.19	0.02	52.23	52.24	-0.02	-0.00
Joint Management Team Region (JMT)	New York City	0.59	56.5	-55.91	-1.58	0.74	0.78	-0.05	-0.00
	Long Island	8.44	8.61	-0.17	-0.01	8.91	9.2	-0.29	-0.01
	Lower Hudson Valley	11.75	6.06	5.69	0.20	13.41	11.67	1.75	0.06
	Mid-Hudson	9.95	2.67	7.28	0.30	9.98	10.46	-0.48	-0.02
	Capital District / North Country	18.53	5.25	13.28	0.42	19.78	20.81	-1.03	-0.03
	Central Region	3.2	2.12	1.08	0.07	3.24	2.99	0.25	0.02
	Mid-State	11.52	2.93	8.59	0.34	11.72	11.86	-0.14	-0.01
	Mid-South	8.17	2.11	6.05	0.28	8.53	9.01	-0.48	-0.02
	Mid-West	18.39	2.94	15.45	0.52	13.13	13.28	-0.15	-0.01
	West	7.9	5.75	2.15	0.09	8.72	7.6	1.13	0.04
	Missing	1.56	5.07	-3.51	-0.20	1.83	2.33	-0.5	-0.03
Needs/Resource Category (NRC) (%)	New York	0.09	50.15	-50.06	-1.41	0.12	0.12	0	0.00
	Big 4 Cities	9.2	3.09	6.11	0.26	10.02	8.26	1.77	0.07
	Urban/Suburban	6.5	6.96	-0.46	-0.02	7.12	8.03	-0.91	-0.04
	High Needs Rural	13.83	3.68	10.15	0.37	14.63	15.45	-0.82	-0.03
	Average Needs	56.73	13.17	43.56	1.03	52.42	52.62	-0.2	-0.00
	Low Needs	11.14	10.16	0.99	0.03	12.72	11.93	0.79	0.03
	Charter School	0.94	7.73	-6.79	-0.34	1.14	1.27	-0.13	-0.01
	Religious and Independent Schools	1.56	5.07	-3.51	-0.20	1.83	2.33	-0.5	-0.03
School Type (%)	Public	97.5	87.21	10.3	0.39	97.03	96.39	0.63	0.02
	Charter	0.94	7.73	-6.79	-0.34	1.14	1.27	-0.13	-0.01
	Non-Public	1.56	5.07	-3.51	-0.20	1.83	2.33	-0.5	-0.03

Table 3.6 Covariate Balance Before and After Matching: Mathematics Grade 4

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	598.97	600.88	-1.91	-0.10	598.88	599.27	-0.38	-0.02
Disability (%)	No	87.47	84.82	2.64	0.08	87.32	86.31	1.01	0.03
	Yes	12.53	15.18	-2.65	-0.08	12.68	13.69	-1.01	-0.03
ELL (%)	No	96.17	88.24	7.93	0.30	95.96	96.2	-0.25	-0.01
	Yes	3.83	11.76	-7.94	-0.30	4.04	3.8	0.25	0.01
Ethnicity (%)	Missing	1	0	1	0.14	0	0	0	0.00
	Asian	3.24	11.19	-7.95	-0.31	3.42	3.4	0.02	0.00
	Black or African American (not Hispanic origin)	7.15	18.56	-11.41	-0.35	7.78	6.2	1.59	0.05
	Hispanic or Latino	14.13	30.27	-16.14	-0.40	13.58	13.58	0.01	0.00
	American Indian or Alaska Native	0.75	0.65	0.1	0.01	0.35	0.54	-0.19	-0.02
	Multi-racial (not Hispanic origin)	3.75	2.78	0.96	0.05	3.59	3.86	-0.27	-0.01
	Native Hawaiian or Other Pacific Islander	0.06	0.27	-0.2	-0.05	0.07	0.05	0.01	0.00
	White (not Hispanic origin)	69.93	36.28	33.65	0.72	71.21	72.38	-1.17	-0.02
Gender (%)	Female	49	49.37	-0.37	-0.01	49.36	48.83	0.54	0.01
	Male	51	50.63	0.37	0.01	50.64	51.17	-0.54	-0.01
Joint Management Team Region (JMT)	New York City	0.84	46.8	-45.96	-1.28	0.93	1.11	-0.19	-0.01
	Long Island	5.53	11.43	-5.9	-0.21	5.66	5.44	0.22	0.01
	Lower Hudson Valley	11.09	6.5	4.6	0.16	10.32	9.1	1.22	0.04
	Mid-Hudson	10.32	3.57	6.75	0.27	11.27	13.35	-2.08	-0.08
	Capital District / North Country	19.75	5.72	14.03	0.43	22.19	21.58	0.61	0.02
	Central Region	1.1	2.75	-1.65	-0.12	1.25	1.62	-0.36	-0.03
	Mid-State	20.23	2.4	17.83	0.59	14.18	13.74	0.44	0.01
	Mid-South	6.53	2.46	4.07	0.20	7.62	8.21	-0.59	-0.03
	Mid-West	13.77	5.13	8.64	0.30	16.02	15.27	0.75	0.03
	West	8.13	6.57	1.56	0.06	9.42	8.93	0.49	0.02
	Missing	2.7	6.67	-3.97	-0.19	1.14	1.65	-0.5	-0.02
Needs/Resource Category (NRC) (%)	New York	0.02	41.24	-41.23	-1.18	0.02	0.02	0	0.00
	Big 4 Cities	11.01	3.42	7.59	0.30	10.15	7.64	2.51	0.10
	Urban/Suburban	5.98	8.04	-2.06	-0.08	6.53	5.62	0.91	0.04
	High Needs Rural	14.13	4.28	9.85	0.35	15.08	15.91	-0.83	-0.03
	Average Needs	59.84	18.38	41.46	0.94	59.94	62.34	-2.4	-0.05
	Low Needs	5.5	11.12	-5.63	-0.20	6.23	5.73	0.5	0.02
	Charter School	0.82	6.85	-6.03	-0.32	0.91	1.1	-0.19	-0.01
	Religious and Independent Schools	2.7	6.67	-3.97	-0.19	1.14	1.65	-0.5	-0.02
School Type (%)	Public	96.48	86.49	9.99	0.36	97.95	97.26	0.69	0.03
	Charter	0.82	6.85	-6.03	-0.32	0.91	1.1	-0.19	-0.01
	Non-Public	2.7	6.67	-3.97	-0.19	1.14	1.65	-0.5	-0.02

Table 3.7 Covariate Balance Before and After Matching: Mathematics Grade 5

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	600.07	601.51	-1.44	-0.08	600.63	600.45	0.18	0.01
Disability (%)	No	88.86	86.44	2.42	0.07	90.61	90.02	0.59	0.02
	Yes	11.14	13.58	-2.43	-0.07	9.39	9.98	-0.59	-0.02
ELL (%)	No	96.05	89.52	6.53	0.25	96.25	96.48	-0.23	-0.01
	Yes	3.95	10.49	-6.54	-0.25	3.75	3.52	0.23	0.01
Ethnicity (%)	Missing	1.08	0	1.08	0.15	0	0	0	0.00
	Asian	3.25	12.2	-8.95	-0.34	3.56	3.09	0.46	0.02
	Black or African American (not Hispanic origin)	7.21	19.26	-12.05	-0.36	7.59	6.74	0.86	0.03
	Hispanic or Latino	15.88	30.37	-14.49	-0.35	13.56	14.26	-0.7	-0.02
	American Indian or Alaska Native	0.47	0.72	-0.25	-0.03	0.49	0.52	-0.03	-0.00
	Multi-racial (not Hispanic origin)	3.71	2.47	1.24	0.07	3.56	3.11	0.45	0.03
	Native Hawaiian or Other Pacific Islander	0.1	0.3	-0.2	-0.04	0.09	0.07	0.02	0.00
	White (not Hispanic origin)	68.3	34.7	33.6	0.71	71.15	72.2	-1.05	-0.02
Gender (%)	Female	48.62	49.1	-0.48	-0.01	49.04	48.56	0.48	0.01
	Male	51.38	50.91	0.47	0.01	50.96	51.44	-0.48	-0.01
Joint Management Team Region (JMT)	New York City	0.62	50.56	-49.94	-1.40	0.68	0.86	-0.18	-0.01
	Long Island	6.76	11.15	-4.39	-0.15	6.66	6.9	-0.24	-0.01
	Lower Hudson Valley	11.55	6.67	4.88	0.17	9.65	8.81	0.85	0.03
	Mid-Hudson	11.04	3.33	7.71	0.30	11.45	10.92	0.53	0.02
	Capital District / North Country	16.2	5.9	10.3	0.33	17.71	19.75	-2.04	-0.07
	Central Region	3.21	2.26	0.95	0.06	3.31	3.79	-0.48	-0.03
	Mid-State	19.59	2.29	17.3	0.58	16.04	14.11	1.93	0.06
	Mid-South	6.26	2.4	3.86	0.19	7.13	5.93	1.2	0.06
	Mid-West	15.29	4.69	10.61	0.36	16.89	17.97	-1.09	-0.04
	West	8.2	6	2.2	0.09	9.07	9.13	-0.06	-0.00
	Missing	1.27	4.76	-3.49	-0.20	1.42	1.84	-0.42	-0.02
Needs/Resource Category (NRC) (%)	New York	0	44.2	-44.2	-1.26	0	0.01	0	0.00
	Big 4 Cities	12.25	2.93	9.32	0.36	9.76	9.43	0.32	0.01
	Urban/Suburban	6.25	7.59	-1.34	-0.05	6.64	4.98	1.66	0.07
	High Needs Rural	13.34	4.02	9.32	0.34	14.03	15.55	-1.52	-0.05
	Average Needs	59.77	17.47	42.3	0.96	60.44	60.88	-0.44	-0.01
	Low Needs	6.49	11.33	-4.83	-0.17	7.04	6.46	0.58	0.02
	Charter School	0.62	7.72	-7.09	-0.36	0.69	0.86	-0.17	-0.01
	Religious and Independent Schools	1.27	4.76	-3.49	-0.20	1.42	1.84	-0.42	-0.02
School Type (%)	Public	98.1	87.54	10.57	0.42	97.9	97.3	0.59	0.02
	Charter	0.62	7.72	-7.09	-0.36	0.69	0.86	-0.17	-0.01
	Non-Public	1.27	4.76	-3.49	-0.20	1.42	1.84	-0.42	-0.02

Table 3.8 Covariate Balance Before and After Matching: Mathematics Grade 6

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	599.85	600.94	-1.09	-0.06	600.07	600.30	-0.22	-0.01
Disability (%)	No	89.18	86.49	2.68	0.08	89.8	90.41	-0.61	-0.02
	Yes	10.82	13.52	-2.7	-0.08	10.2	9.59	0.61	0.02
ELL (%)	No	97.05	90.07	6.97	0.29	97.1	97.33	-0.23	-0.01
	Yes	2.95	9.94	-6.99	-0.29	2.9	2.67	0.23	0.01
Ethnicity (%)	Missing	0.85	0	0.85	0.13	0	0	0	0.00
	Asian	3.79	12.57	-8.79	-0.32	4.27	3.78	0.5	0.02
	Black or African American (not Hispanic origin)	8.03	19.64	-11.61	-0.34	8.34	7.95	0.39	0.01
	Hispanic or Latino	16	30.62	-14.62	-0.35	15.64	15.18	0.46	0.01
	American Indian or Alaska Native	0.6	0.72	-0.12	-0.01	0.53	0.45	0.08	0.01
	Multi-racial (not Hispanic origin)	3.6	2.24	1.36	0.08	3.62	3.53	0.09	0.01
	Native Hawaiian or Other Pacific Islander	0.07	0.35	-0.28	-0.06	0.08	0.08	0.01	0.00
	White (not Hispanic origin)	67.07	33.88	33.19	0.70	67.52	69.03	-1.52	-0.03
Gender (%)	Female	48.18	49.31	-1.13	-0.02	48.21	48.59	-0.38	-0.01
	Male	51.82	50.7	1.12	0.02	51.79	51.41	0.38	0.01
Joint Management Team Region (JMT)	New York City	0.46	51.66	-51.19	-1.44	0.49	0.51	-0.02	-0.00
	Long Island	8.53	10.15	-1.62	-0.06	8.42	8.8	-0.38	-0.01
	Lower Hudson Valley	12.24	6.23	6	0.21	12.78	11.33	1.45	0.05
	Mid-Hudson	10.07	3.14	6.93	0.28	10.35	10.21	0.14	0.01
	Capital District / North Country	20.18	4.73	15.44	0.48	20.22	21.58	-1.36	-0.04
	Central Region	2.42	2.24	0.18	0.01	2.42	2.37	0.05	0.00
	Mid-State	15.06	2.47	12.59	0.46	13.62	13.43	0.2	0.01
	Mid-South	7.79	2.09	5.7	0.27	8.39	8.49	-0.11	-0.00
	Mid-West	13.5	4.39	9.11	0.32	14.72	14.37	0.35	0.01
	West	6.82	5.98	0.84	0.03	7.39	7.65	-0.26	-0.01
	Missing	2.93	6.94	-4.01	-0.19	1.2	1.27	-0.07	-0.00
Needs/Resource Category (NRC) (%)	New York	0	44.92	-44.92	-1.28	0	0	0	0.00
	Big 4 Cities	10.01	3.03	6.99	0.29	10.37	9.1	1.27	0.05
	Urban/Suburban	8.39	6.74	1.65	0.06	8.99	7.6	1.38	0.05
	High Needs Rural	11.91	3.94	7.96	0.30	12.31	13.61	-1.3	-0.05
	Average Needs	57.28	15.6	41.68	0.96	56.89	57.58	-0.69	-0.02
	Low Needs	9.03	10.78	-1.75	-0.06	9.76	10.34	-0.57	-0.02
	Charter School	0.46	8.08	-7.62	-0.38	0.49	0.51	-0.02	-0.00
	Religious and Independent Schools	2.93	6.94	-4.01	-0.19	1.2	1.27	-0.07	-0.00
School Type (%)	Public	96.61	85	11.62	0.41	98.31	98.23	0.09	0.00
	Charter	0.46	8.08	-7.62	-0.38	0.49	0.51	-0.02	-0.00
	Non-Public	2.93	6.94	-4.01	-0.19	1.2	1.27	-0.07	-0.00

Table 3.9 Covariate Balance Before and After Matching: Mathematics Grade 7

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	602.10	600.79	1.32	0.07	602.05	602.15	-0.10	-0.01
Disability (%)	No	88.56	86.62	1.94	0.06	90.25	90.29	-0.04	-0.00
	Yes	11.44	13.39	-1.95	-0.06	9.75	9.71	0.04	0.00
ELL (%)	No	97.6	90.75	6.84	0.30	97.58	98.06	-0.48	-0.02
	Yes	2.4	9.25	-6.85	-0.30	2.42	1.94	0.48	0.02
Ethnicity (%)	Missing	0.79	0	0.79	0.13	0	0	0	0.00
	Asian	4.32	12.41	-8.09	-0.30	4.94	4.58	0.36	0.01
	Black or African American (not Hispanic origin)	8.21	19.64	-11.43	-0.33	8.65	8.31	0.34	0.01
	Hispanic or Latino	15.81	30.24	-14.43	-0.35	14.57	14.19	0.37	0.01
	American Indian or Alaska Native	0.71	0.68	0.03	0.00	0.37	0.34	0.04	0.00
	Multi-racial (not Hispanic origin)	3.36	1.96	1.41	0.09	3.38	3.31	0.07	0.00
	Native Hawaiian or Other Pacific Islander	0.1	0.41	-0.32	-0.06	0.09	0.1	-0.01	-0.00
	White (not Hispanic origin)	66.69	34.66	32.03	0.68	67.99	69.17	-1.18	-0.02
Gender (%)	Female	48.26	49.31	-1.05	-0.02	48.59	48.44	0.15	0.00
	Male	51.74	50.7	1.04	0.02	51.41	51.56	-0.15	-0.00
Joint Management Team Region (JMT)	New York City	0.24	52.48	-52.24	-1.47	0.28	0.43	-0.15	-0.00
	Long Island	7.16	9.67	-2.51	-0.09	7.33	7.84	-0.5	-0.02
	Lower Hudson Valley	13.57	5.91	7.66	0.26	12.57	11.48	1.09	0.04
	Mid-Hudson	11.84	2.85	8.99	0.35	11.72	11.27	0.45	0.02
	Capital District / North Country	21.15	4.52	16.63	0.51	19.14	20.32	-1.18	-0.04
	Central Region	2.22	2.24	-0.02	-0.00	2.24	2.35	-0.1	-0.01
	Mid-State	13.96	2.72	11.24	0.42	14.1	13.45	0.65	0.02
	Mid-South	7.11	2.06	5.05	0.24	7.88	7.62	0.26	0.01
	Mid-West	14.86	3.98	10.88	0.38	16.13	16.4	-0.27	-0.01
	West	6.52	5.89	0.63	0.03	7.03	7.34	-0.31	-0.01
	Missing	1.38	7.7	-6.31	-0.31	1.56	1.5	0.06	0.00
Needs/Resource Category (NRC) (%)	New York	0	46.21	-46.21	-1.31	0	0.01	0	0.00
	Big 4 Cities	11.03	2.6	8.43	0.34	9.39	9.02	0.37	0.01
	Urban/Suburban	7.03	6.81	0.22	0.01	7.4	6.95	0.45	0.02
	High Needs Rural	13.45	3.72	9.73	0.35	13.63	13.93	-0.3	-0.01
	Average Needs	57.67	14.37	43.3	1.01	57.67	58.01	-0.35	-0.01
	Low Needs	9.04	11.03	-1.99	-0.07	9.91	10.06	-0.15	-0.01
	Charter School	0.4	7.57	-7.17	-0.37	0.44	0.52	-0.07	-0.00
	Religious and Independent Schools	1.38	7.7	-6.31	-0.31	1.56	1.5	0.06	0.00
School Type (%)	Public	98.22	84.74	13.48	0.50	98	97.99	0.01	0.00
	Charter	0.4	7.57	-7.17	-0.37	0.44	0.52	-0.07	-0.00
	Non-Public	1.38	7.7	-6.31	-0.31	1.56	1.5	0.06	0.00

Table 3.10 Covariate Balance Before and After Matching: Mathematics Grade 8

variable	value	Before Matching				After Matching			
		CBT	PBT	Delta	*d*	CBT	PBT	Delta	*d*
2018 OP Scale Score	(Mean)	593.92	597.22	-3.30	-0.18	594.99	595.08	-0.09	0.00
Disability (%)	No	84.61	83.52	1.1	0.03	86.16	86.99	-0.83	-0.02
	Yes	15.39	16.49	-1.11	-0.03	13.84	13.01	0.83	0.02
ELL (%)	No	96.63	89.68	6.95	0.28	96.43	96.9	-0.47	-0.02
	Yes	3.37	10.32	-6.96	-0.28	3.57	3.1	0.47	0.02
Ethnicity (%)	Missing	1.31	0	1.31	0.16	0	0	0	0.00
	Asian	3.45	10.65	-7.2	-0.28	3.73	3.96	-0.23	-0.01
	Black or African American (not Hispanic origin)	10.97	22.17	-11.2	-0.30	9.78	10.38	-0.6	-0.02
	Hispanic or Latino	17.73	32.7	-14.97	-0.35	12.69	14.47	-1.78	-0.04
	American Indian or Alaska Native	0.6	0.86	-0.26	-0.03	0.56	0.43	0.13	0.02
	Multi-racial (not Hispanic origin)	2.76	1.65	1.11	0.08	2.61	2.51	0.1	0.01
	Native Hawaiian or Other Pacific Islander	0.09	0.35	-0.26	-0.05	0.09	0.07	0.02	0.00
	White (not Hispanic origin)	63.09	31.63	31.46	0.66	70.54	68.18	2.36	0.05
Gender (%)	Female	46.34	47.49	-1.15	-0.02	46.13	45.49	0.63	0.01
	Male	53.66	52.52	1.14	0.02	53.87	54.51	-0.63	-0.01
Joint Management Team Region (JMT)	New York City	0.64	57.06	-56.42	-1.59	0.86	1.02	-0.16	-0.00
	Long Island	4.21	5.11	-0.9	-0.04	4.52	4.51	0.01	0.00
	Lower Hudson Valley	16.99	6.03	10.97	0.35	8.82	12.86	-4.04	-0.13
	Mid-Hudson	9.94	3.04	6.9	0.28	10.48	10.43	0.05	0.00
	Capital District / North Country	21.32	5.43	15.9	0.48	25	21.83	3.17	0.10
	Central Region	2.47	2.41	0.06	0.00	2.5	2.48	0.02	0.00
	Mid-State	14.78	2.77	12.01	0.43	13.73	12.97	0.76	0.03
	Mid-South	8.11	2.41	5.7	0.26	9.52	8.88	0.64	0.03
	Mid-West	14.06	3.52	10.54	0.38	15.75	15.63	0.12	0.00
	West	5.68	6.86	-1.17	-0.05	6.53	7	-0.46	-0.02
	Missing	1.79	5.37	-3.59	-0.19	2.29	2.4	-0.11	-0.01
Needs/Resource Category (NRC) (%)	New York	0.16	51.53	-51.37	-1.45	0.22	0.22	0	0.00
	Big 4 Cities	17.38	2.75	14.63	0.50	6.01	8.98	-2.97	-0.10
	Urban/Suburban	6.95	6.76	0.19	0.01	7.96	7.92	0.04	0.00
	High Needs Rural	13.79	4.61	9.18	0.32	15.66	14.89	0.77	0.03
	Average Needs	52.33	14.12	38.21	0.89	58.92	55.52	3.39	0.08
	Low Needs	6.92	7.87	-0.95	-0.04	8.02	8.92	-0.9	-0.03
	Charter School	0.69	6.99	-6.3	-0.33	0.92	1.14	-0.23	-0.01
	Religious and Independent Schools	1.79	5.37	-3.59	-0.19	2.29	2.4	-0.11	-0.01
School Type (%)	Public	97.52	87.64	9.88	0.38	96.79	96.46	0.34	0.01
	Charter	0.69	6.99	-6.3	-0.33	0.92	1.14	-0.23	-0.01
	Non-Public	1.79	5.37	-3.59	-0.19	2.29	2.4	-0.11	-0.01

3.2. Test-level Mode Comparability

After having achieved a reasonably good covariate balance between the matched CBT and PBT samples, the test-level mode comparability was evaluated. Questar calculated the sample means for each matched sample and their standardized differences before and after matching for the following variables:

- 2018 Scale Score: the prior year (grade n – 1) scale score, which was the proxy for prior ability that was entered as a key predictor into the propensity score model
- 2019 Raw Score: the current year operational raw score
- 2019 Scale Score: the current year scale score

The test-level performance before and after matching is summarized in Tables 3.11–3.14.

Table 3.11 Test-level Performance between Test Modes Before Matching – ELA

		PBT			CBT		
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
ELA4	2018 Scale Score	132,457	601.24	20.04	25,035	598.81	18.67
	2019 Raw Score	152,010	20.72	6.77	28,299	19.31	6.44
	2019 Scale Score	152,010	599.17	19.56	28,299	595.09	18.08
ELA5	2018 Scale Score	125,862	601.05	20.31	30,142	599.13	18.34
	2019 Raw Score	141,544	27.33	9.02	33,839	26.22	8.63
	2019 Scale Score	141,544	599.68	21.45	33,839	597.02	19.89
ELA6	2018 Scale Score	122,333	600.84	20.30	29,607	600.18	18.54
	2019 Raw Score	141,634	28.61	8.65	34,548	28.10	8.35
	2019 Scale Score	141,634	598.45	22.51	34,548	597.06	21.19
ELA7	2018 Scale Score	114,055	600.85	20.38	26,808	600.05	18.55
	2019 Raw Score	128,552	29.96	9.26	31,360	28.92	8.77
	2019 Scale Score	128,552	600.19	21.33	31,360	597.52	19.36
ELA8	2018 Scale Score	113,222	601.24	19.82	22,258	599.29	18.64
	2019 Raw Score	126,088	31.67	8.98	25,878	30.27	8.89
	2019 Scale Score	126,088	600.48	19.85	25,878	597.22	18.92

Table 3.12. Test-level Performance between Test Modes After Matching – ELA

		PBT			CBT		
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
ELA4	2018 Scale Score	21,781	599.19	19.5	21,781	599.13	18.71
	2019 Raw Score	21,781	20.41	6.43	21,781	19.54	6.41
	2019 Scale Score	21,781	598.18	18.19	21,781	595.72	17.98
ELA5	2018 Scale Score	25,210	599.85	19.49	25,210	599.17	18.43
	2019 Raw Score	25,210	27.5	8.67	25,210	26.53	8.6
	2019 Scale Score	25,210	600.01	20.41	25,210	597.76	19.85
ELA6	2018 Scale Score	23,181	600.16	19.66	23,181	599.97	18.82
	2019 Raw Score	23,181	28.76	8.51	23,181	28.22	8.37

		PBT			CBT		
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
ELA6	2019 Scale Score	23,181	598.67	21.96	23,181	597.35	21.26
	2018 Scale Score	21,807	599.59	19.82	21,807	599.78	18.8
ELA7	2019 Raw Score	21,807	29.49	9.05	21,807	29.2	8.77
	2019 Scale Score	21,807	598.93	20.3	21,807	598.16	19.42
	2018 Scale Score	19,875	599.02	19.83	19,875	599.08	18.72
ELA8	2019 Raw Score	19,875	31	9.18	19,875	30.63	8.74
	2019 Scale Score	19,875	598.94	20.01	19,875	597.99	18.61

Table 3.13 Test-level Performance between Test Modes Before Matching – Math

		PBT			CBT		
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
	2018 Scale Score	140,255	600.88	19.98	18,793	598.97	18.93
Math4	2019 Raw Score	161,135	28.46	11.12	21,198	27.28	10.52
	2019 Scale Score	161,135	600.58	20.42	21,198	598.10	18.57
	2018 Scale Score	132,758	601.51	19.83	19,668	600.07	18.23
Math5	2019 Raw Score	148,983	27.67	11.54	21,975	25.85	11.03
	2019 Scale Score	148,983	601.08	20.48	21,975	597.60	19.08
	2018 Scale Score	125,598	600.94	20.13	20,994	599.85	18.53
Math6	2019 Raw Score	144,992	26.29	12.40	24,435	25.08	11.64
	2019 Scale Score	144,992	601.20	20.00	24,435	599.21	18.38
	2018 Scale Score	121,675	600.79	19.97	20,101	602.10	18.17
Math7	2019 Raw Score	136,764	28.58	13.50	23,151	26.93	12.62
	2019 Scale Score	136,764	601.93	20.36	23,151	599.35	18.73
	2018 Scale Score	82,260	597.22	19.31	11,839	593.92	17.60
Math8	2019 Raw Score	93,095	25.89	12.67	13,994	22.56	11.16
	2019 Scale Score	93,095	601.02	21.32	13,994	595.62	19.68

Table 3.14 Test-level Performance between Test Modes After Matching – Math

		PBT			CBT		
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
	2018 Scale Score	16,520	599.27	19.38	16,520	598.88	18.97
Math4	2019 Raw Score	16,520	28.18	10.72	16,520	27.48	10.49
	2019 Scale Score	16,520	599.74	19.31	16,520	598.43	18.49
	2018 Scale Score	17,660	600.45	18.06	17,660	600.63	18.02
Math5	2019 Raw Score	17,660	27.62	10.91	17,660	26.62	10.84
	2019 Scale Score	17,660	600.67	18.74	17,660	598.98	18.53
	2018 Scale Score	19,582	600.30	18.75	19,582	600.07	18.47
Math6	2019 Raw Score	19,582	26.67	11.68	19,582	25.66	11.60
	2019 Scale Score	19,582	601.65	18.27	19,582	600.10	18.19

		PBT			CBT		
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Math7	2018 Scale Score	18,446	602.15	18.56	18,446	602.05	18.00
	2019 Raw Score	18,446	29.21	12.77	18,446	27.61	12.53
	2019 Scale Score	18,446	602.61	18.59	18,446	600.38	18.32
Math8	2018 Scale Score	10,133	595.08	17.56	10,133	594.99	16.90
	2019 Raw Score	10,133	24.81	11.49	10,133	23.69	10.94
	2019 Scale Score	10,133	599.27	19.22	10,133	597.74	18.54

Section 4: Discussion and Conclusions

4.1. Discussion

Based on the analyses described above, NYSED—in consultation with New York State's Assessment TAC and Questar—decided to apply an additive adjustment to CBT students' scale scores because it best balanced concerns about fairness, interpretability and face validity. NYSED also chose to set a ceiling above which the CBT students' scale scores would not be adjusted–namely the maximum observed scale score available to PBT students. In other words, the highest scale score on CBT was constrained to be equal to the highest scale score for PBT students.

The differences in the 2019 scale score means between the matched samples were computed. The differences were rounded to the nearest whole numbers, which were used as the uniform additive adjustment applied to the CBT students within each test.

For Grade 3 students who do not have prior year scores, there was no propensity score matched samples being generated. Alternatively, the average of mode adjustments of Grades 4 and 5 was computed and used as the adjustment for Grade 3 CBT students. The CBT adjustments in all grade/subject were summarized in Table 4.1.

Table 1.3 CBT Scale Score Adjustments after Matching

		PBT			*CBT*					CBT Adjustment		
Subject	**Grade**	***n***	***M***	***SD***	***n***	***M***	***SD***	**Δ**	***d***	**2019**	**2018**	**2017**
ELA	3	-	-	-	-	-	-	-2.36	-	+2	+1	+4
	4	21,781	598.18	18.19	21,781	595.72	17.98	-2.46	-0.14	+2	+1	+5
	5	25,210	600.01	20.41	25,210	597.76	19.85	-2.25	-0.11	+2	+1	+2
	6	23,181	598.67	21.96	23,181	597.35	21.26	-1.33	-0.06	+1	+1	+5
	7	21,807	598.93	20.30	21,807	598.16	19.42	-0.77	-0.04	+1	+1	+2
	8	19,875	598.94	20.01	19,875	597.99	18.61	-0.95	-0.05	+1	+1	0
Math	3	-	-	-	-	-	-	-1.50	-	+2	+1	+4
	4	16,520	599.74	19.31	16,520	598.43	18.49	-1.31	-0.07	+1	+1	+5
	5	17,660	600.67	18.74	17,660	598.98	18.53	-1.69	-0.09	+2	0	+2
	6	19,582	601.66	18.27	19,582	600.10	18.19	-1.55	-0.09	+2	+2	+3
	7	18,446	602.61	18.59	18,446	600.38	18.32	-2.24	-0.12	+2	0	0
	8	10,133	599.27	19.22	10,133	597.74	18.54	-1.53	-0.08	+2	+1	+8

Note. CBT scale scores were only adjusted up to the maximum observed PBT scale score value.

4.2. Conclusions

Following 2017 and 2018, two administration modes were offered in the Grades 3-8 ELA and Mathematics tests in Spring 2019. The decision to offer PBT vs. CBT was optional, and the population of students who tested via CBT were not assumed equivalent to the population of students who tested via PBT. The propensity score matching was conducted select a sample of PBT students that could be comparable to the population of CBT students. The test-level performance was then evaluated between the matched CBT and PBT samples. The results revealed small differences between CBT and PBT group across the tests, with mode effects slightly favoring PBT groups. The observed differences were applied as the adjustments to CBT students to ensure that students received comparable test scores regardless of the test mode.

Section 5: References

Austin, P. C. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424. doi: http://dx.doi.org/10.1080/00273171.2011.568786

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2), 150–161. doi: http://dx.doi.org/10.1002/pst.433

Cai, L. Thissen, D., & du Toit, S. (2017). *IRTPRO 4* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155. http://dx.doi.org/10.1037/0033-2909.112.1.155

Firth, D. (1993), Bias reduction of maximum likelihood estimates, *Biometrika*, 80, 27–38. http://dx.doi.org/10.1093/biomet/80.1.27

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 129–145). Hillsdale, NJ: Erlbaum.

Kim, S. & M. J. Kolen (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models* [Computer software]. Iowa City, IA: Iowa Testing Programs, The University of Iowa.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement, 16*(2), 159–176. http://dx.doi.org/10.1177/014662169201600206

Rosenbaum, P. R. (2010). *Design of observational studies*. New York: Springer. doi: http://dx.doi.org/10.1007/978-1-4419-1213-8

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. doi: http://dx.doi.org/10.1093/biomet/70.1.41

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. Applied Psychological Measurement, *7*(2), 201–210. doi: http://dx.doi.org/10.1177/014662168300700208xxxx

Thissen, D., Chen, W-H., & Bock, R. D. (2003). *MULTILOG 7 for Windows: Multiple-category item analysis and test scoring using item response theory* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Way, W. D., Davis, L. L., & Strain-Seymour, E. (2008). *The validity case for assessing direct writing by computer (A Pearson Assessments & Information White Paper)*. Pearson: Iowa City, IA.

Appendix S: Memo on Operational Test Mode Comparability

THE STATE EDUCATION DEPARTMENT / THE UNIVERSITY OF THE STATE OF NEW YORK / ALBANY, NY 12234

Assistant Commissioner
Office of State Assessment

August 2019

TO: District Superintendents
Superintendents of Schools
Principals of Public, Religious, and Independent Schools
Leaders of Charter Schools

FROM: Steven E. Katz

SUBJECT: Comparability of Spring 2019 Grades 3-8 English Language Arts and Mathematics Paper-based and Computer-based Tests

The purpose of this memorandum is to provide information about the results of the comparability study that was conducted for the Spring 2019 Grades 3-8 English Language Arts (ELA) and Mathematics paper-based and computer-based tests.

Background

In Spring 2019, the Department offered the Grades 3-8 ELA and Mathematics Tests in two administration modes: paper-based testing (PBT) and computer-based testing (CBT). Administering these tests via CBT was optional for schools and those schools that chose to offer CBT made this decision independently for each subject and grade. The Department provided readiness verification tools to help those schools selecting CBT ensure they were well equipped and prepared to provide a successful CBT experience for their students. Additionally, several CBT practice test sessions were made available to CBT schools to familiarize students and teachers with the CBT delivery system. Each of the CBT practice test sessions featured examples of all of the types of test questions included on the tests. This provided the opportunity for students to practice answering both multiple-choice and constructed-response questions on the computer devices they would be using for the actual test.

To further ensure fairness, the Department's contractor, Questar Assessment Inc., conducted a comparability study to identify whether or not there were any differences in student performance that could be attributed to the mode of test administration (i.e., PBT versus CBT). The comparability study methodology and results are summarized below. The findings of this study were used to ensure that students received a score that was representative of their knowledge and skills, regardless of whether they took the tests on paper or computer.

Comparability Study Methodology

Only some schools chose to administer the tests via CBT (representing approximately 19% of all ELA test takers and 14% of all Math test takers). Therefore, the population of students who tested via CBT were not assumed equivalent to the population of students who tested via PBT. In order to select a sample of students who tested via PBT that could be compared to those students who tested via CBT, a method called propensity score matching was employed. Propensity score matching allowed for the identification of groups of students who tested via

PBT that were similar to the groups of students who tested via CBT on a number of school and student characteristics, including achievement on the prior year's test.

Using these characteristics, Questar selected a group of PBT students that matched the group of CBT students for each grade and subject. This allowed for a direct comparison of student results between the two groups. For comparison, the mean scale scores were calculated for each grade and subject by mode of testing. The results are shown in the section below.

Results of Comparability Study

Table 1 shows the scale score means for the PBT and CBT groups on the 2019 English Language Arts Tests by grade as well as the differences in mean scale scores between the matched groups. Table 2 shows these same data for the 2019 Mathematics Tests.

Table 1. *PBT and CBT Means and Differences for Grades 3-8 ELA*

	PBT Scale Score Mean	CBT Scale Score Mean	Difference (Rounded to nearest whole number)
Grade 3	See footnote*		n/a
Grade 4	598.2	595.7	+2
Grade 5	600.0	597.8	+2
Grade 6	598.7	597.4	+1
Grade 7	598.9	598.2	+1
Grade 8	598.9	598.0	+1

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level.

Table 2. *PBT and CBT Means and Differences for Grades 3-8 Math*

	PBT Scale Score Mean	CBT Scale Score Mean	Difference (Rounded to nearest whole number)
Grade 3	See footnote*		n/a
Grade 4	599.7	598.4	+1
Grade 5	600.7	599.0	+2
Grade 6	601.7	600.1	+2
Grade 7	602.6	600.4	+2
Grade 8	599.3	597.7	+2

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level.

For those tests in which a difference in mean scale scores between the two comparable groups was observed, the scale scores for all students who took the test in that grade via CBT, (which was the lower scoring mode in all instances during this administration), were adjusted by adding the number of scale score points shown in the "Difference" columns of Tables 1 and 2 to the CBT students' scale scores, up to the maximum attainable scale score. Thus, the scale score adjustments for students who tested via CBT, shown in Table 3 below, reflect the differences between the PBT and CBT groups found in the comparability study. These slight adjustments ensured that students who demonstrated comparable proficiencies in their knowledge and skills received comparable scores whether they tested on paper or on computer.

Table 3. *Summary of Scale Score Adjustments for CBT*

	ELA Scale Score Adjustment	Math Scale Score Adjustment
Grade 3	+2*	+2*
Grade 4	+2	+1
Grade 5	+2	+2
Grade 6	+1	+2
Grade 7	+1	+2
Grade 8	+1	+2

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level. Instead, the mean adjustment for the other elementary grades for which a comparison was possible (i.e., Grades 4 & 5) was applied to the scores of Grade 3 students who tested via CBT.

For questions concerning the Grades 3-8 ELA or Mathematics Tests, please email the Office of State Assessment at call 518-474-5902. For questions concerning CBT, please email CBT Support.