New York State Testing Program 2022:
English Language Arts and Mathematics
Grades 3–8

Technical Report

Questar Assessment Inc.
2022

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 2022 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, test administration, scoring, scaling, and student performance.

1.2. Test Purpose

The 2022 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 2022, some religious and independent schools participated in the testing program across all grade levels. These schools were included in the data analyses. Public school and charter 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.nysed.gov/common/nysed/files/programs/state-assessment/38-sam-2022.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. *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*.

1.4.3. *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 2022 tests. In 2022, 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

| Grade | Total Subscore Points | |
	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

| Grade | Reporting Subscores and Total Subscore Points | | |
	Subscore 1	Subscore 2	Subscore 3
3	Operations and Algebraic Thinking 19	Number and Operations—Fractions 7	Measurement and Data 10
4	Operations and Algebraic Thinking 10	Numbers and Operations in Base 10 11	Number and Operations—Fractions 12
5	Numbers and Operations in Base 10 13	Number and Operations—Fractions 18	Measurement and Data 13
6	Ratios and Proportional Relationships 12	The Number System 9	Expressions and Equations 19

Grade	Reporting Subscores and Total Subscore Points		
	Subscore 1	Subscore 2	Subscore 3
7	Ratios and Proportional Relationships 14	The Number System 11	Expressions and Equations 14
8	Expressions and Equations 19	Functions 14	Geometry 13

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 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available online at http://www.nysed.gov/common/nysed/files/programs/state-assessment/38-sam-2022.pdf

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 Pearson. 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.

1.7. Test Translations

The NYSTP Grades 3–8 Mathematics Tests are translated into eight languages: Arabic, Bengali, Chinese (Simplified), 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: https://www.nysedregents.org/ei/translations.html.

English Language Learners (ELLs) 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 ELLs taking the ELA Tests: separate testing location and bilingual glossaries.

Section 2: Test Design and Development

2.1. Test Descriptions

The 2022 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, a six-day period for ELA computer-based tests, and a seven-day period for Mathematics computer-based tests from March to May of 2022. 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 online at http://www.nysed.gov/common/nysed/files/programs/state-assessment/38-sam-2022.pdf

2.1.1. *ELA Tests*

The 2022 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 2022 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 2022 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 2022 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, and 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 2022 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 2022, 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 2022 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.

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 item needs 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).

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 *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 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.

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 2018–2019 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 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 2022 test forms was conducted during the 2018–2019 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, 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 ELA and Mathematics, the educators used the following checklists to review each item.

ELA Item 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 Item 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 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. 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 a psychometric perspective.

There were two types of approaches used for field-testing in Spring 2019: embedded field-testing and stand-alone field-testing. With embedded field-testing, field test items were administered within the 2019 operational test forms for most multiple-choice items. With stand-alone field-testing, field test items were administered separately from the 2019 operational forms later in Spring 2019, which included CR items and a small 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 2019 Field Test Technical Report.

2.10. Rangefinding

Questar conducted rangefinding for items included on the operational 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, derived from Operational scoring training materials, 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 March to May of 2022. 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 2017 through 2019.

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 selected 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. Appendix F provides general psychometric guidelines for operational item selection. For example, one of the guidelines for building the NYSTP Grades 3–8 ELA and Mathematics Tests was that the point-biserial correlation 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. 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 October 2019 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 October 21–22, 2019 in Albany, New York, educators from around the State worked with NYSED and Questar to review the content of the proposed 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.

Historically, the Educator Form Construction Meeting takes place during the fall preceding the spring administration for which the form is being built. For example, the operational forms administered in the Spring of 2018 were taken to the Form Construction Meeting in the fall of 2017. However, because no exams were given in Spring 2020, and "reuse" forms were administered in Spring 2021, the planned 2020 forms taken to Form Construction in the fall 2019 were not administered until Spring 2022. The "reuse" forms administered in Spring 2021 consisted of a form selected from Spring 2018 or Spring 2019.

On October 21, 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 2020 administration. Because no tests were administered in Spring 2020 and "reuse" forms were administered in Spring 2021 (see explanation in Section 2.12 above), the forms taken to Final Eyes in Fall 2019 were not administered until Spring 2022.

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). NYSED uses various assessment data in reporting AYP. Specific to student-level outcomes, the NYSTP documents student performance in the area of ELA as defined by the New York State ELA Learning Standards, and in the area of Mathematics as defined by the New York State Mathematics Learning.

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 0.88 to 0.91. For all subgroups, the reliability coefficients were greater than or equal to 0.81. For the total population, the Mathematics reliability coefficients (Cronbach's alpha) ranged from 0.91 to 0.93. For all subgroups, the reliability coefficients were greater than or equal to 0.77. 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., ELA proficiency or Mathematics 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. 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 six times and eight 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 21% and 25% 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		
		Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	**6.98**	**27.91**	**27.91**
	2	1.52	6.08	33.98
	3	1.13	4.52	38.50
4	1	**6.79**	**27.16**	**27.16**
	2	1.34	5.35	32.51
5	1	**8.06**	**23.03**	**23.03**
	2	1.47	4.21	27.24
	3	1.25	3.57	30.82
6	1	**7.44**	**21.26**	**21.26**
	2	1.54	4.39	25.65
	3	1.15	3.28	28.93
	4	1.03	2.95	31.88
7	1	**8.62**	**23.95**	**23.95**
	2	1.72	4.79	28.74
	3	1.20	3.34	32.08
	4	1.04	2.88	34.97
	5	1.01	2.82	37.78
8	1	**8.91**	**24.74**	**24.74**
	2	1.96	5.45	30.19
	3	1.06	2.94	33.13
	4	1.01	2.80	35.93

Table 3.2. Mathematics Tests Factor Analysis

Grade	#	Extracted Factor		
		Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	**9.61**	**28.27**	**28.27**
	2	1.67	4.91	33.18
	3	1.19	3.50	36.69
	4	1.00	2.94	39.63
4	1	**10.93**	**28.76**	**28.76**
	2	1.59	4.19	32.95
	3	1.06	2.79	35.74
5	1	**11.48**	**30.21**	**30.21**
	2	1.37	3.60	33.82
	3	1.07	2.83	36.64
6	1	**10.92**	**28.01**	**28.01**
	2	1.33	3.42	31.43

Grade	Extracted Factor			
	#	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	3	1.11	2.85	34.28
	4	1.01	2.59	36.86
7	1	**10.64**	**25.94**	**25.94**
	2	1.27	3.09	29.03
	3	1.26	3.07	32.10
	4	1.03	2.52	34.63
8	1	**10.45**	**25.48**	**25.48**
	2	1.39	3.40	28.88
	3	1.20	2.92	31.80
	4	1.12	2.73	34.53
	5	1.04	2.53	37.05

As additional evidence for construct validity, the same factor analysis procedure was employed to assess the dimensionality of the ELA and Mathematics construct for selected subgroups of students in each grade: ELLs, students with disabilities (SWD), and students using test accommodations (SUA). Appendix L provides factor analysis results for these subgroup classifications. The results were comparable to the results obtained from the total population data. Evaluation of the magnitude of eigenvalue and proportion of variance explained by the main factor provide evidence of essential unidimensionality of the construct measured by the tests for these subgroups.

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 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. In each grade, for both ELA and Mathematics, few items were flagged for DIF. 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. Please refer to the 2022 Field Test Technical report for details about the DIF analysis.

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 (2022)* located here: http://www.nysed.gov/common/nysed/files/programs/state-assessment/38-scoring-leader-handbook-2022.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 2022. The PBT testing window was Tuesday, March 29–Thursday, March 31 for the Grades 3–8 ELA Tests and Tuesday, April 26–Thursday, April 28 for the Grades 3–8 Mathematics Tests. The CBT testing window was Tuesday, March 29–Tuesday, April 5 for the Grades 3–8 ELA Tests and Tuesday, April 26–Wednesday, May 4 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 and CBT was Friday, April 1–Friday, April 8 for the Grades 3–8 ELA Tests and Friday, April 29–Monday, May 9 for the Grades 3–8 Mathematics Tests.

4.2. Scoring Procedures of Operational Tests

Qualified teachers and administrators performed the scoring of the NYSTP 2022 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, districtwide, or schoolwide 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 schoolwide models. At the districtwide 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 schoolwide scoring, oversight was provided by the principal; otherwise, titles for the schoolwide model were the same as those for the districtwide 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 *2022 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 *2022 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 2021–2022 school year, schools and school districts were able to score Grades 3–8 ELA and/or Mathematics Tests regionally, multi-district, districtwide, or schoolwide, 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 2022 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 New York 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 25, found online at: http://www.nysed.gov/common/nysed/files/programs/state-assessment/38-scoring-leader-handbook-2022.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 scoring guidelines and policies were followed, approximately 5% of student results for both ELA and mathematics 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, at the end of May 2022. During Phase 2, "straggler files" were submitted to Questar in June 2022.

The "straggler files" contained fewer than 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 public, charter, 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 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 types and those with at least one entirely missing test session. Other deleted cases included students with incorrect or incomplete grade information, duplicate record cases, no-response record cases, and mismatched form codes.

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 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 2022 relative to 2019[1], but they were not used directly in the sampling down analyses.

---

[1] Because no exams were given in Spring 2020, and "reuse" forms were administered in Spring 2021, the demographic comparisons were made between 2022 and 2019.

Comparison results between the final 2022 sample and 2019 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 2019 and 2022 operational forms. There were some minor differences in the 2019 and 2022 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. This decision was made because the differences in student performance and demographic makeup reflect the changes observed in the testing population during the pandemic, and therefore it did not make sense to resample using metrics observed before the pandemic (i.e., 2019).

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	175,201
Wrong Subject	0	175,201
No Grade	0	175,201
Wrong Grade	36	175,165
Form Code Mismatch	421	174,744
School Type	2,938	171,806
Missing Entire Session	19,171	152,635
Invalid Score	0	152,635
Not Tested Reason	368	152,267
Out-of-Range CR Scores	0	152,267
Duplicated Record	8	152,259
Test Mode Discrepancy	0	152,259

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	178,949
Wrong Subject	0	178,949
No Grade	0	178,949
Wrong Grade	25	178,924
Form Code Mismatch	494	178,430
School Type	3,059	175,371
Missing Entire Session	17,061	158,310
Invalid Score	0	158,310
Not Tested Reason	4,851	153,459
Out-of-Range CR Scores	0	153,459

Exclusion Rule	# Deleted	# Cases Remain
Duplicated Record	22	153,437
Test Mode Discrepancy	0	153,437

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	187,367
Wrong Subject	0	187,367
No Grade	0	187,367
Wrong Grade	32	187,335
Form Code Mismatch	396	186,939
School Type	3,258	183,681
Missing Entire Session	17,911	165,770
Invalid Score	0	165,770
Not Tested Reason	6,249	159,521
Out-of-Range CR Scores	0	159,521
Duplicated Record	22	159,499
Test Mode Discrepancy	0	159,499

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	191,547
Wrong Subject	0	191,547
No Grade	0	191,547
Wrong Grade	36	191,511
Form Code Mismatch	455	191,056
School Type	3,770	187,286
Missing Entire Session	19,959	167,327
Invalid Score	0	167,327
Not Tested Reason	9,552	157,775
Out-of-Range CR Scores	0	157,775
Duplicated Record	25	157,750
Test Mode Discrepancy	0	157,750

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	183,977
Wrong Subject	0	183,977
No Grade	35	183,942
Wrong Grade	29	183,913
Form Code Mismatch	376	183,537
School Type	3,916	179,621
Missing Entire Session	23,635	155,986
Invalid Score	0	155,986
Not Tested Reason	11,913	144,073
Out-of-Range CR Scores	0	144,073
Duplicated Record	20	144,053
Test Mode Discrepancy	0	144,053

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	189,096
Wrong Subject	0	189,096
No Grade	0	189,096
Wrong Grade	63	189,033
Form Code Mismatch	361	188,672
School Type	4,407	184,265
Missing Entire Session	45,287	138,978
Invalid Score	0	138,978
Not Tested Reason	570	138,408
Out-of-Range CR Scores	0	138,408
Duplicated Record	28	138,380
Test Mode Discrepancy	0	138,380

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	183,484
Wrong Subject	0	183,484
No Grade	0	183,484
Wrong Grade	52	183,432
Form Code Mismatch	8,968	174,464
School Type	2,877	171,587

Exclusion Rule	# Deleted	# Cases Remain
Missing Entire Session	14,858	156,729
Invalid Score	0	156,729
Not Tested Reason	3,295	153,434
Out-of-Range CR Scores	0	153,434
Duplicated Record	10	153,424
Test Mode Discrepancy	0	153,424

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	190,016
Wrong Subject	0	190,016
No Grade	0	190,016
Wrong Grade	41	189,975
Form Code Mismatch	9,819	180,156
School Type	2,369	177,787
Missing Entire Session	16,742	161,045
Invalid Score	0	161,045
Not Tested Reason	4,231	156,814
Out-of-Range CR Scores	0	156,814
Duplicated Record	26	156,788
Test Mode Discrepancy	0	156,788

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	189,032
Wrong Subject	0	189,032
No Grade	0	189,032
Wrong Grade	47	188,985
Form Code Mismatch	8,957	180,028
School Type	3,181	176,847
Missing Entire Session	19,350	157,497
Invalid Score	0	157,497
Not Tested Reason	5,844	151,653
Out-of-Range CR Scores	0	151,653
Duplicated Record	20	151,633
Test Mode Discrepancy	0	151,633

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	193,010
Wrong Subject	0	193,010
No Grade	0	193,010
Wrong Grade	42	192,968
Form Code Mismatch	10,283	182,685
School Type	2,921	179,764
Missing Entire Session	22,396	157,368
Invalid Score	0	157,368
Not Tested Reason	10,023	147,345
Out-of-Range CR Scores	0	147,345
Duplicated Record	25	147,320
Test Mode Discrepancy	0	147,320

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	193,691
Wrong Subject	0	193,691
No Grade	0	193,691
Wrong Grade	26	193,665
Form Code Mismatch	10,293	183,372
School Type	3,162	180,210
Missing Entire Session	37,896	142,314
Invalid Score	0	142,314
Not Tested Reason	552	141,762
Out-of-Range CR Scores	0	141,762
Duplicated Record	32	141,730
Test Mode Discrepancy	0	141,730

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

Table 5.12. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	147,137
Wrong Subject	0	147,137
No Grade	0	147,137
Wrong Grade	27	147,109
Form Code Mismatch	7,230	139,879
School Type	3,417	136,462

Exclusion Rule	# Deleted	# Cases Remain
Missing Entire Session	46,257	90,205
Invalid Score	0	90,205
Not Tested Reason	439	89,766
Out-of-Range CR Scores	0	89,766
Duplicated Record	20	89,746
Test Mode Discrepancy	0	89,746

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 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 NRC is assigned at the district level and is an indicator of district and school socioeconomic status. The ethnicity and gender[2] 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	75,603	49.65
	Male	76,654	50.34
	Non-Binary	-	-
Ethnicity	Asian	16,654	10.95
	African American	20,645	13.57
	Hispanic	41,198	27.08
	American Indian	1,105	0.73
	Multiracial	5,680	3.73
	Pacific Islander	346	0.23
	White	66,515	43.72
NRC	New York	49,076	32.23
	Big 4 Cities	6,399	4.20
	Urban/Suburban	12,242	8.04
	Rural	9,112	5.98
	Average Needs	42,833	28.13
	Low Needs	19,308	12.68
	Charter School	6,417	4.21
	Religious and Independent	6,872	4.51
SWD	No	130,936	86.00

---

[2] In 2021–22, a new gender category was introduced, "nonbinary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

Demographic Category		N-Count	% of Total N-Count
	Yes	21,323	14.00
SUA	No	134,460	88.31
	Yes	17,799	11.69
ELL	No	135,208	88.80
	Yes	17,051	11.20
SWD/ SUA	No	136,472	89.63
	Yes	15,787	10.37
ELL/ SUA	No	149,740	98.35
	Yes	2,519	1.65

*The total n-count was 152,259.

Table 5.14. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,213	49.67
	Male	77,222	50.33
	Non-Binary	-	-
Ethnicity	Asian	17,975	11.72
	African American	21,452	13.99
	Hispanic	41,437	27.03
	American Indian	1,139	0.74
	Multiracial	5,317	3.47
	Pacific Islander	328	0.21
	White	65,678	42.84
NRC	New York	51,010	33.24
	Big 4 Cities	6,271	4.09
	Urban/Suburban	12,589	8.20
	Rural	9,102	5.93
	Average Needs	41,922	27.32
	Low Needs	19,225	12.53
	Charter School	6,173	4.02
	Religious and Independent	7,145	4.66
SWD	No	131,441	85.66
	Yes	21,996	14.34
SUA	No	133,042	86.71
	Yes	20,395	13.29
ELL	No	137,607	89.68
	Yes	15,830	10.32

Demographic Category		N-Count	% of Total N-Count
SWD/ SUA	No	135,735	88.46
	Yes	17,702	11.54
ELL/ SUA	No	150,618	98.16
	Yes	2,819	1.84

*The total n-count was 153,437.

Table 5.15. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	79,229	49.67
	Male	80,262	50.32
	Non-Binary	-	-
Ethnicity	Asian	16,967	10.65
	African American	25,980	16.30
	Hispanic	44,467	27.90
	American Indian	1,153	0.72
	Multiracial	5,200	3.26
	Pacific Islander	369	0.23
	White	65,243	40.94
NRC	New York	51,635	32.37
	Big 4 Cities	6,166	3.87
	Urban/Suburban	12,326	7.73
	Rural	9,037	5.67
	Average Needs	41,605	26.08
	Low Needs	18,660	11.70
	Charter School	13,198	8.27
	Religious and Independent	6,872	4.31
SWD	No	135,765	85.12
	Yes	23,734	14.88
SUA	No	137,155	85.99
	Yes	22,344	14.01
ELL	No	145,320	91.11
	Yes	14,179	8.89
SWD/ SUA	No	140,176	87.89
	Yes	19,323	12.11
ELL/ SUA	No	156,487	98.11
	Yes	3,012	1.89

*The total n-count was 159,499.

Table 5.16. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	77,772	49.30
	Male	79,968	50.69
	Non-Binary	-	-
Ethnicity	Asian	16,643	10.56
	African American	26,034	16.52
	Hispanic	45,394	28.8
	American Indian	1,097	0.70
	Multiracial	4,903	3.11
	Pacific Islander	327	0.21
	White	63,233	40.11
NRC	New York	50,457	31.99
	Big 4 Cities	6,279	3.98
	Urban/Suburban	12,234	7.76
	Rural	8,722	5.53
	Average Needs	39,151	24.82
	Low Needs	17,642	11.18
	Charter School	13,862	8.79
	Religious and Independent	9,403	5.96
SWD	No	134,673	85.37
	Yes	23,077	14.63
SUA	No	135,834	86.11
	Yes	21,916	13.89
ELL	No	144,486	91.59
	Yes	13,264	8.41
SWD/ SUA	No	139,169	88.22
	Yes	18,581	11.78
ELL/ SUA	No	154,665	98.04
	Yes	3,085	1.96

*The total n-count was 157,750.

Table 5.17. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	70,388	48.86
	Male	73,654	51.13
	Non-Binary	-	-

Demographic Category		N-Count	% of Total N-Count
Ethnicity	Asian	16,817	11.68
	African American	21,848	15.18
	Hispanic	39,724	27.60
	American Indian	1,018	0.71
	Multiracial	4,418	3.07
	Pacific Islander	321	0.22
	White	59,796	41.54
NRC	New York	52,798	36.65
	Big 4 Cities	6,256	4.34
	Urban/Suburban	11,730	8.14
	Rural	8,783	6.10
	Average Needs	36,922	25.63
	Low Needs	17,808	12.36
	Charter School	4,041	2.81
	Religious and Independent	5,715	3.97
SWD	No	123,226	85.54
	Yes	20,827	14.46
SUA	No	124,405	86.36
	Yes	19,648	13.64
ELL	No	133,188	92.46
	Yes	10,865	7.54
SWD/ SUA	No	127,407	88.44
	Yes	16,646	11.56
ELL/ SUA	No	141,517	98.24
	Yes	2,536	1.76

*The total n-count was 144,053.

Table 5.18. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	66,758	48.24
	Male	71,603	51.74
	Non-Binary	-	-
Ethnicity	Asian	16,834	12.17
	African American	21,692	15.69
	Hispanic	38,553	27.88
	American Indian	990	0.72

Demographic Category		N-Count	% of Total N-Count
	Multiracial	3,876	2.80
	Pacific Islander	346	0.25
	White	55,997	40.49
NRC	New York	52,175	37.7
	Big 4 Cities	6,243	4.51
	Urban/Suburban	11,379	8.22
	Rural	8,762	6.33
	Average Needs	33,960	24.54
	Low Needs	16,179	11.69
	Charter School	3,763	2.72
	Religious and Independent	5,919	4.28
SWD	No	118,513	85.64
	Yes	19,867	14.36
SUA	No	119,444	86.32
	Yes	18,936	13.68
ELL	No	128,190	92.64
	Yes	10,190	7.36
SWD/ SUA	No	122,348	88.41
	Yes	16,032	11.59
ELL/ SUA	No	136,146	98.39
	Yes	2,234	1.61

*The total n-count was 138,380.

Table 5.19. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,723	50.01
	Male	76,699	49.99
	Non-Binary	-	-
Ethnicity	Asian	15,794	10.3
	African American	22,902	14.94
	Hispanic	38,840	25.34
	American Indian	1,406	0.72
	Multiracial	5,609	3.66
	Pacific Islander	352	0.23
	White	68,691	44.81

Demographic Category		N-Count	% of Total N-Count
NRC	New York	44,325	28.89
	Big 4 Cities	6,277	4.09
	Urban/Suburban	11,257	7.34
	Rural	9,198	6.00
	Average Needs	41,459	27.02
	Low Needs	17,614	11.48
	Charter School	12,619	8.22
	Religious and Independent	10,675	6.96
SWD	No	134,848	87.89
	Yes	18,576	12.11
SUA	No	138,757	90.44
	Yes	14,667	9.56
ELL	No	135,203	88.12
	Yes	18,221	11.88
SWD/ SUA	No	140,682	91.69
	Yes	12,742	8.31
ELL/ SUA	No	151,588	98.80
	Yes	1,836	1.20

*The total n-count was 153,424.

Table 5.20. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,390	50.00
	Male	78,395	50.00
	Non-Binary	-	-
Ethnicity	Asian	17,235	11.00
	African American	23,883	15.25
	Hispanic	39,502	25.22
	American Indian	1,171	0.75
	Multiracial	5,375	3.43
	Pacific Islander	331	0.21
	White	69,161	44.15
NRC	New York	45,775	29.2
	Big 4 Cities	6,124	3.91
	Urban/Suburban	11,230	7.16
	Rural	9,333	5.95
	Average Needs	40,690	25.95

Demographic Category		N-Count	% of Total N-Count
	Low Needs	17,805	11.36
	Charter School	12,616	8.05
NRC	Religious and Independent	13,215	8.43
SWD	No	137,859	87.93
	Yes	18,929	12.07
SUA	No	140,071	89.34
	Yes	16,717	10.66
ELL	No	139,803	89.17
	Yes	16,985	10.83
SWD/ SUA	No	142,686	91.01
	Yes	14,102	8.99
ELL/ SUA	No	154,784	98.72
	Yes	2,004	1.28

*The total n-count was 156,788.

Table 5.21. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	75,536	49.82
	Male	76,087	50.18
	Non-Binary	-	-
Ethnicity	Asian	16,058	10.6
	African American	23,810	15.71
	Hispanic	38,464	25.39
	American Indian	1,072	0.71
	Multiracial	4,952	3.27
	Pacific Islander	347	0.23
	White	66,816	44.1
NRC	New York	45,810	30.21
	Big 4 Cities	5,972	3.94
	Urban/Suburban	10,799	7.12
	Rural	9,168	6.05
	Average Needs	39,588	26.11
	Low Needs	17,463	11.52
	Charter School	12,071	7.96
	Religious and Independent	10,762	7.10

Demographic Category		N-Count	% of Total N-Count
SWD	No	132,962	87.69
	Yes	18,671	12.31
SUA	No	134,755	88.87
	Yes	16,878	11.13
ELL	No	137,038	90.37
	Yes	14,595	9.63
SWD/ SUA	No	137,337	90.57
	Yes	14,296	9.43
ELL/ SUA	No	149,706	98.73
	Yes	1,927	1.27

*The total n-count was 151,633.

Table 5.22. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	72,766	49.39
	Male	74,543	50.60
	Non-Binary	-	-
Ethnicity	Asian	15,238	10.35
	African American	23,885	16.23
	Hispanic	38,168	25.93
	American Indian	1,002	0.68
	Multiracial	4,660	3.17
	Pacific Islander	304	0.21
	White	63,920	43.43
NRC	New York	43,612	29.60
	Big 4 Cities	6,081	4.13
	Urban/Suburban	10,426	7.08
	Rural	8,705	5.91
	Average Needs	36,451	24.74
	Low Needs	16,115	10.94
	Charter School	13,045	8.85
	Religious and Independent	12,885	8.75
SWD	No	129,389	87.83
	Yes	17,931	12.17
SUA	No	131,022	88.94
	Yes	16,298	11.06

Demographic Category		N-Count	% of Total N-Count
ELL	No	135,007	91.64
	Yes	12,313	8.36
SWD/ SUA	No	133,771	90.80
	Yes	13,549	9.20
ELL/ SUA	No	145,570	98.81
	Yes	1,750	1.19

*The total n-count was 147,320.

Table 5.23. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	69,942	49.35
	Male	71,772	50.64
	Non-Binary	-	-
Ethnicity	Asian	15,574	11.00
	African American	23,723	16.75
	Hispanic	36,585	25.83
	American Indian	976	0.69
	Multiracial	4,271	3.02
	Pacific Islander	319	0.23
	White	60,195	42.50
NRC	New York	45,398	32.03
	Big 4 Cities	5,926	4.18
	Urban/Suburban	9,601	6.77
	Rural	8,577	6.05
	Average Needs	33,879	23.90
	Low Needs	15,458	10.91
	Charter School	12,138	8.56
	Religious and Independent	10,753	7.59
SWD	No	124,489	87.84
	Yes	17,241	12.16
SUA	No	126,222	89.06
	Yes	15,508	10.94
ELL	No	131,846	93.03
	Yes	9,884	6.97
SWD/ SUA	No	128,851	90.91
	Yes	12,879	9.09

Demographic Category		N-Count	% of Total N-Count
ELL/ SUA	No	140,210	98.93
	Yes	1,520	1.07

*The total n-count was 141,730.

Table 5.24. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	43,300	48.25
	Male	46,428	51.73
	Non-Binary	-	-
Ethnicity	Asian	7,278	8.12
	African American	15,962	17.81
	Hispanic	24,336	27.15
	American Indian	624	0.70
	Multiracial	2,404	2.68
	Pacific Islander	189	0.21
	White	38,855	43.34
NRC	New York	26,541	29.57
	Big 4 Cities	4,929	5.49
	Urban/Suburban	6,536	7.28
	Rural	7,171	7.99
	Average Needs	22,475	25.04
	Low Needs	7,788	8.68
	Charter School	6,476	7.22
	Religious and Independent	7,830	8.72
SWD	No	75,963	84.64
	Yes	13,783	15.36
SUA	No	77,278	86.11
	Yes	12,468	13.89
ELL	No	81,871	91.23
	Yes	7,875	8.77
SWD/ SUA	No	79,207	88.26
	Yes	10,539	11.75
ELL/ SUA	No	88,646	98.77
	Yes	1,100	1.23

*The total n-count was 89,746.

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. Point-biserial statistics are used to examine item-test correlations, or item discrimination. Items are flagged for review by a subject matter expert according to the criteria listed in Table 5.25.

Table 5.25. Item Analysis Flagging Criteria

Item Type	p-value	Point-Biserial
MC	< 0.25 or > 0.90	< 0.20
CR	< 0.30 or > 0.85	< 0.40

The number of 2022 OP items flagged for each content area and grade are given in Table 5.26.

Table 5.26. Number of Flagged Items

Subject	Grade	# Items	# Flagged Items p-value	# Flagged Items Point-Biserial
ELA	3	25		
	4	25		
	5	35		2
	6	35		2
	7	36		2
	8	36		
Mathematics	3	34	3	1
	4	38	1	

			# Flagged Items	
Subject	Grade	# Items	p-value	Point-Biserial
	5	38	1	
	6	39	2	
	7	41	0	
	8	41	6	

If an MC item is flagged, a subject matter expert reviews the item and intended key to verify that the item was scored correctly. Choices are checked to verify that one and only one correct answer exists. If a CR item is flagged, a subject matter expert reviews the item to ensure that all components are present (e.g., art not omitted) and to ensure the item is clearly worded.

If no defects are found in a flagged item, a subject matter expert may suggest a reason for the statistical flag, if apparent. Multiple flags may result in an item being removed from the pool of eligible items.

The summary statistics of p-value and point-biserial correlations for the operational items are show in Table 5.27 and Table 5.28, respectively. There was a reasonably wide range of item difficulties for each test. The mean item difficulties ranged from 0.61 to 0.67 for ELA and 0.50 to 0.60 for Mathematics.

Point-biserial correlations ranged from 0.11 to 0.69 for the ELA tests and 0.16 to 0.76 for the Mathematics tests. The mean point-biserial correlations ranged from 0.39 to 0.46 for ELA and 0.46 to 0.50 for mathematics.

Table 5.27. Item Difficulty Distribution

Subject	Grade	N-Count	Mean	SD	Min	Max
	3	152,259	0.61	0.15	0.31	0.90
	4	153,437	0.62	0.11	0.42	0.84
ELA	5	159,499	0.63	0.13	0.30	0.87
	6	157,750	0.66	0.12	0.44	0.86
	7	144,053	0.65	0.14	0.34	0.88
	8	138,380	0.67	0.11	0.43	0.84
	3	153,424	0.60	0.18	0.22	0.96
	4	156,788	0.60	0.16	0.25	0.90
Mathematics	5	151,633	0.55	0.15	0.24	0.89
	6	147,320	0.52	0.13	0.12	0.79
	7	141,730	0.56	0.13	0.30	0.81
	8	89,746	0.50	0.16	0.15	0.83

Table 5.28. Item Discrimination Distribution

Subject	Grade	N-Count	Mean	SD	Min	Max
	3	152,259	0.46	0.12	0.21	0.68
	4	153,437	0.45	0.12	0.21	0.69

Subject	Grade	N-Count	Mean	SD	Min	Max
ELA	5	159,499	0.42	0.12	0.17	0.66
	6	157,750	0.39	0.13	0.16	0.65
	7	144,053	0.42	0.14	0.11	0.68
	8	138,380	0.44	0.12	0.21	0.68
	3	153,424	0.48	0.12	0.16	0.69
	4	156,788	0.49	0.08	0.36	0.68
Mathematics	5	151,633	0.50	0.11	0.27	0.76
	6	147,320	0.49	0.09	0.32	0.64
	7	141,730	0.46	0.12	0.22	0.74
	8	89,746	0.46	0.10	0.27	0.68

In Appendix M, Tables M1–M12 illustrate classical test statistics for all items on each grade-level test.

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 were less than 1%.

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. DIF was evaluated for operational items using two methods: the Mantel-Haenszel Delta method (Dorans, & Holland, 1993) for MC items, and the standardized mean difference method (Dorans, Schmitt, and Bleistein, 1992) for CR items. Please refer to the *2022 Field-test Technical Report for details* about these DIF methods and item fagging criteria. 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, to identify potential systematic issues that could be addressed in future item writing.

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).

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 (Lord and Novick, 1968; Lord, 1980) 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

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

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 ability student.

For analysis of the CR items, the generalized partial credit (GPC) model (Muraki, 1992) was used. The probability of an examinee with proficiency θ obtaining a score k on item i is:

$$P_{ik}(\theta) = P(x_i = k) = \frac{e^{\sum_{h=0}^{k} Da_i(\theta - b_i + d_{ih})]}}{\sum_{c=0}^{m_i} e^{\sum_{h=0}^{c} Da_i(\theta - b_i + d_{ih})}},$$

where
      m_i is the maximum number of score points of item i,

      a_i is the discrimination parameter of item i,

      b_i is the location parameter of item i,

d_{ih} is the category parameter of item i on score h,

$d_{i0} = 0$,

$\sum_{k=1}^{m_i} d_{ik} = 0$, and

D is a scaling constant of 1.7.

Each item has one discrimination parameter, one location parameter, and m_i -1 independent category parameters.

6.2. Calibration Sample

The cleaned data were used for calibration of the NYSTP 2022 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 2022 operational test samples were generally comparable to 2019 populations[3] in terms of NRC, student gender[4], race and ethnicity, proportions of ELLs, proportions of students with disabilities, and proportions of students using testing accommodations.

Table 6.1. ELA Grades 3 and 4 Demographic Statistics

		Grade 3		Grade 4	
		2019	2022	2019	2022
Demographic Category		Population	Sample	Population	Sample
Gender	Female	49.43	49.65	49.33	49.69
	Male	50.57	50.34	50.67	50.33
	Non-Binary	-	-	-	-
Ethnicity	Asian	9.92	10.95	10.18	11.72
	African American	17.07	13.57	17.63	13.99
	Hispanic	28.69	27.08	28.60	27.03
	American Indian	0.68	0.73	0.67	0.74
	Multiracial	3.11	3.73	2.94	3.47
	Pacific Islander	0.22	0.23	0.24	0.21
	White	40.31	43.72	39.73	42.84
NRC	New York	36.22	32.23	36.84	33.24
	Big 4 Cities	4.17	4.20	4.30	4.09
	Urban/Suburban	8.15	8.04	7.83	8.20

---

[3] Because no exams were given in Spring 2020, and "reuse" forms were administered in Spring 2021, the 2022 sample were compared with the 2019 population.

[4] In 2021–22, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

			Grade 3		Grade 4	
			2019	2022	2019	2022
Demographic Category			**Population**	**Sample**	**Population**	**Sample**
NRC		Rural	5.46	5.98	5.43	5.93
		Average Needs	24.19	28.13	23.19	27.32
		Low Needs	10.75	12.68	10.40	12.53
		Charter School	7.03	4.21	6.41	4.02
		Religious and Independent	4.01	4.51	5.60	4.66
SWD		No	85.12	86.00	84.38	85.66
		Yes	14.88	14.00	15.62	14.34
SUA		No	86.85	88.31	85.12	86.71
		Yes	13.15	11.69	14.88	13.29
ELL		No	89.28	88.80	90.31	89.68
		Yes	10.72	11.20	9.69	10.32
SWD/ SUA		No	88.33	89.63	86.93	88.46
		Yes	11.67	10.37	13.07	11.54
ELL/ SUA		No	97.95	98.35	97.78	98.16
		Yes	2.05	1.65	2.22	1.84

Table 6.2. ELA Grades 5 and 6 Demographic Statistics

			Grade 5		Grade 6	
			2019	2022	2019	2022
Demographic Category			**Population**	**Sample**	**Population**	**Sample**
Gender		Female	49.12	49.67	49.20	49.23
		Male	50.88	50.32	50.80	50.69
		Non-Binary	-	-	-	-
Ethnicity		Asian	10.55	10.65	10.73	10.56
		African American	17.90	16.3	18.12	16.52
		Hispanic	28.31	27.9	28.28	28.8
		American Indian	0.67	0.72	0.71	0.70
		Multiracial	2.72	3.26	2.50	3.11
		Pacific Islander	0.26	0.23	0.30	0.21
		White	39.58	40.94	39.37	40.11
NRC		New York	37.24	32.37	37.30	31.99
		Big 4 Cities	4.33	3.87	4.21	3.98
		Urban/Suburban	7.80	7.73	7.19	7.76
		Rural	5.50	5.67	5.27	5.53
		Average Needs	23.30	26.08	21.88	24.82

		Grade 5		Grade 6	
		2019	**2022**	**2019**	**2022**
Demographic Category		**Population**	**Sample**	**Population**	**Sample**
	Low Needs	10.54	11.7	10.25	11.18
	Charter School	6.88	8.27	7.01	8.79
	Religious and Independent	4.42	4.31	6.89	5.96
SWD	No	83.89	85.12	84.22	85.37
	Yes	16.11	14.88	15.78	14.63
SUA	No	84.45	85.99	85.02	86.11
	Yes	15.55	14.01	14.98	13.89
ELL	No	91.38	91.11	91.64	91.59
	Yes	8.62	8.89	8.36	8.41
SWD/ SUA	No	86.40	87.89	87.06	88.22
	Yes	13.60	12.11	12.94	11.78
ELL/ SUA	No	97.91	98.11	98.02	98.04
	Yes	2.09	1.89	1.98	1.96

Table 6.3. ELA Grades 7 and 8 Demographic Statistics

		Grade 7		Grade 8	
		2019	**2022**	**2019**	**2022**
Demographic Category		**Population**	**Sample**	**Population**	**Sample**
Gender	Female	48.69	48.86	48.44	48.24
	Male	51.31	51.13	51.56	51.74
	Non-Binary	-	-	-	-
Ethnicity	Asian	11.27	11.68	11.30	12.17
	African American	17.19	15.18	19.42	15.69
	Hispanic	27.59	27.6	28.29	27.88
	American Indian	0.70	0.71	0.80	0.72
	Multiracial	2.28	3.07	1.90	2.80
	Pacific Islander	0.36	0.22	0.29	0.25
	White	40.61	41.54	38.01	40.49
NRC	New York	40.31	36.65	41.62	37.7
	Big 4 Cities	4.20	4.34	4.14	4.51
	Urban/Suburban	7.44	8.14	6.88	8.22
	Rural	5.64	6.10	5.40	6.33
	Average Needs	22.08	25.63	20.59	24.54
	Low Needs	11.17	12.36	10.32	11.69
	Charter School	2.71	2.81	6.57	2.72

Demographic Category		Grade 7		Grade 8	
		2019	**2022**	**2019**	**2022**
		Population	**Sample**	**Population**	**Sample**
	Religious and Independent	6.45	3.97	4.47	4.28
SWD	No	84.27	85.54	83.79	85.64
	Yes	15.73	14.46	16.21	14.36
SUA	No	84.77	86.36	84.50	86.32
	Yes	15.23	13.64	15.50	13.68
ELL	No	92.45	92.46	92.86	92.64
	Yes	7.55	7.54	7.14	7.36
SWD/ SUA	No	87.01	88.44	86.70	88.41
	Yes	12.99	11.56	13.30	11.59
ELL/ SUA	No	98.21	98.24	98.38	98.39
	Yes	1.79	1.76	1.62	1.61

Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2019	**2022**	**2019**	**2022**
		Population	**Sample**	**Population**	**Sample**
Gender	Female	49.31	50.01	49.33	50.00
	Male	50.69	49.99	50.67	50.00
	Non-Binary	-	-	-	-
Ethnicity	Asian	9.92	10.30	10.28	11.00
	African American	16.58	14.94	17.26	15.25
	Hispanic	28.31	25.34	28.43	25.22
	American Indian	0.67	0.72	0.66	0.75
	Multiracial	3.02	3.66	2.90	3.43
	Pacific Islander	0.22	0.23	0.24	0.21
	White	41.28	44.81	40.24	44.15
NRC	New York	35.65	28.89	36.45	29.2
	Big 4 Cities	4.16	4.09	4.30	3.91
	Urban/Suburban	8.10	7.34	7.80	7.16
	Rural	5.43	6.00	5.43	5.95
	Average Needs	23.83	27.02	23.20	25.95
	Low Needs	10.64	11.48	10.47	11.36
	Charter School	6.64	8.22	6.15	8.05
	Religious and Independent	5.56	6.96	6.21	8.43

Demographic Category		Grade 3		Grade 4	
		2019 Population	2022 Sample	2019 Population	2022 Sample
SWD	No	85.72	87.89	85.13	87.93
	Yes	14.28	12.11	14.87	12.07
SUA	No	87.98	90.44	86.56	89.34
	Yes	12.02	9.56	13.44	10.66
ELL	No	87.96	88.12	89.16	89.17
	Yes	12.04	11.88	10.84	10.83
SWD/ SUA	No	89.36	91.69	88.13	91.01
	Yes	10.64	8.31	11.87	8.99
ELL/ SUA	No	98.03	98.80	97.89	98.72
	Yes	1.97	1.20	2.11	1.28

Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2019 Population	2022 Sample	2019 Population	2022 Sample
Gender	Female	49.03	49.82	49.14	49.39
	Male	50.97	50.18	50.86	50.60
	Non-Binary	-	-	-	-
Ethnicity	Asian	11.06	10.60	11.32	10.35
	African American	17.73	15.71	17.98	16.23
	Hispanic	28.54	25.39	28.54	25.93
	American Indian	0.69	0.71	0.71	0.68
	Multiracial	2.64	3.27	2.44	3.17
	Pacific Islander	0.27	0.23	0.31	0.21
	White	39.07	44.1	38.71	43.43
NRC	New York	38.51	30.21	38.43	29.6
	Big 4 Cities	4.12	3.94	4.03	4.13
	Urban/Suburban	7.42	7.12	6.97	7.08
	Rural	5.22	6.05	5.09	5.91
	Average Needs	22.91	26.11	21.61	24.74
	Low Needs	10.70	11.52	10.52	10.94
	Charter School	6.80	7.96	6.98	8.85
	Religious and Independent	4.31	7.10	6.36	8.75
SWD	No	86.74	87.69	86.87	87.86
	Yes	13.26	12.31	13.13	12.17

Demographic Category		Grade 5		Grade 6	
		2019	**2022**	**2019**	**2022**
		Population	**Sample**	**Population**	**Sample**
SUA	No	88.20	88.87	88.66	88.94
	Yes	11.80	11.13	11.34	11.06
ELL	No	90.35	90.37	91.07	91.64
	Yes	9.65	9.63	8.93	8.36
SWD/ SUA	No	89.78	90.57	90.22	90.80
	Yes	10.22	9.43	9.78	9.20
ELL/ SUA	No	98.23	98.73	98.34	98.81
	Yes	1.77	1.27	1.66	1.19

Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2019	**2022**	**2019**	**2022**
		Population	**Sample**	**Population**	**Sample**
Gender	Female	49.15	49.35	47.34	48.25
	Male	50.85	50.64	52.66	51.73
	Non-Binary	-	-	-	-
Ethnicity	Asian	11.25	11.00	9.73	8.12
	African American	18.00	16.75	20.74	17.81
	Hispanic	28.18	25.83	30.80	27.15
	American Indian	0.68	0.69	0.82	0.70
	Multiracial	2.16	3.02	1.79	2.68
	Pacific Islander	0.37	0.23	0.32	0.21
	White	39.34	42.5	35.80	43.34
NRC	New York	39.52	32.03	44.81	29.57
	Big 4 Cities	3.82	4.18	4.66	5.49
	Urban/Suburban	6.84	6.77	6.78	7.28
	Rural	5.13	6.05	5.81	7.99
	Average Needs	20.64	23.9	19.12	25.04
	Low Needs	10.74	10.91	7.74	8.68
	Charter School	6.53	8.56	6.17	7.22
	Religious and Independent	6.78	7.59	4.90	8.72
SWD	No	86.89	87.84	83.65	84.64
	Yes	13.11	12.16	16.35	15.36

		Grade 7		Grade 8	
		2019	**2022**	**2019**	**2022**
Demographic Category		**Population**	**Sample**	**Population**	**Sample**
SUA	No	88.36	89.06	85.59	86.11
	Yes	11.64	10.94	14.41	13.89
ELL	No	91.74	93.03	90.59	91.23
	Yes	8.26	6.97	9.41	8.77
SWD/ SUA	No	90.08	90.91	87.71	88.26
	Yes	9.92	9.09	12.29	11.74
ELL/ SUA	No	98.54	98.93	98.42	98.77
	Yes	1.46	1.07	1.58	1.23

6.2.1. *Calibration Process*

The item parameters were estimated using Scientific Software International (SSI) Inc.'s IRTPRO Version 5.2 (Cai, Thissen, & du Toit, 2011) package. MC and CR items were calibrated simultaneously, using marginal maximum likelihood procedures.

The calibration of NYSTP 2022 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 N, in Tables N1–N12. The parameter estimates are expressed on the theta metric and are defined below:

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

- CR items:
    - a is an item discrimination parameter
    - b is an item-location parameter
    - d_k is item category parameter for category k

Table 6.7. ELA Calibration Results

Grade	N-Count	Range of a-Parameters		Range of b-Parameters	
3	152,259	0.27	1.41	-1.78	0.89
4	153,437	0.41	1.27	-1.35	1.43
5	159,499	0.24	1.47	-2.60	1.61

Grade	N-Count	Range of *a*-Parameters		Range of *b*-Parameters	
6	157,750	0.20	1.20	-2.49	1.07
7	144,053	0.19	1.53	-1.42	2.78
8	138,380	0.29	1.74	-1.33	1.27

Table 6.8. Mathematics Calibration Results

Grade	N-Count	Range of *a*-Parameters		Range of *b*-Parameters	
3	153,424	0.44	1.90	-2.43	1.46
4	156,788	0.52	1.64	-1.65	0.94
5	151,633	0.51	1.96	-1.72	1.12
6	147,320	0.58	1.87	-1.05	1.38
7	141,730	0.43	2.21	-1.18	1.33
8	89,746	0.41	1.80	-1.20	1.74

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. All items showed adequate model-data fit. The fact that the items in the NYSTP 2022 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 their response to another item. In other words, when a student's proficiency is accounted for, their 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

$$d_{ij} = u_{ij} - p_j(\hat{\theta}_i)$$

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

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

The generalization to items with multiple response categories uses

$$d_{ij} = x_{ij} - E_{ij},$$

where

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

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 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
An IRT based equating procedure was conducted for all ELA and mathematics tests to place the 2022 calibrated item parameter estimates on the base scale established in 2018.

Once the item calibrations were completed, a set of common items between the current (new) and the previous (reference) administrations was used to conduct the equating of the form onto the base scale using a common-item nonequivalent groups design. The anchor items were a pre-selected set of eligible MC items that were placed throughout the test and contributed to the students' scores. Most of the anchor items were embedded as field test items in 2019, except for seven anchor items in the ELA Grade 5 test that were field tested in 2018. Anchor items consisted of 28–35% of the total score points and 34–48% of the operational items across subjects and grade levels. The banked IRT parameters of the anchor items were used as reference parameters in the 2022 post-equating process.

The Stocking-Lord (1983) procedure, as implemented by the STUIRT computer program (Kim & Kolen, 2004), was used to obtain the equating coefficients needed for the linear transformation of each new form onto the base scale.

The stability of the common items used in the Stocking-Lord procedure was assessed using the weighted, squared deviations between the item characteristic curves (ICCs) of the items for the new and reference administrations. For a given item i the deviation called "d-squared" was calculated by

$$d_i^2 = \sum_k \{[\mathrm{Pr}_{i,new}(\theta_k) - \mathrm{Pr}_{i,ref}(\theta_k)]^2 \cdot g(\theta_k)\}$$

where i indexes anchor items, k indexes quadrature points for θ, $\mathrm{Pr}_{i,ref}(\cdot)$ is the probability of a correct response to item i under the previous administrations, while $\mathrm{Pr}_{i,new}(\cdot)$ is the same quantity based on the current administration's on-scale item parameters, and $g(\theta_k)$ are weights for each of the K quadrature points.

Anchor item with $d_i^2 > 0.01$ would be flagged and considered for removal from an anchor set. Based on the iterative process of the equating analysis with the anchor item with the largest d^2 being removed with each iteration along with impact data analysis, Questar identified three items to be dropped from the anchor set, one anchor item each from grades 5, 7, and 8 of ELA.

Tables 6.9 and 6.10 present the resulting equating coefficients. The following parameters were equated using the formula below:

$$a_i^E = \frac{a_i^C}{M_1^E},$$

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

$$d_{ij}^E = d_{ij}^C M_1^E,$$

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	0.9597	-0.1201
4	1.0210	-0.2523
5	0.9476	0.0551
6	0.9897	0.1329

Grade	M_1^E	M_2^E
7	0.9429	0.1817
8	1.0552	-0.0284

Table 6.10. Mathematics Equating Coefficients

Grade	M_1^E	M_2^E
3	1.0243	-0.2027
4	1.1320	-0.2322
5	1.1044	-0.2496
6	1.0211	-0.1682
7	0.9783	-0.1245
8	1.0450	-0.2395

6.6. Scaling and 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 their 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 Test Characteristic Curve (TCC) method (see the details provided below). The obtained scaling transformation intercept and slope (M_1^E and M_2^E) 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 points. 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:

$$\sum_{i=1}^{n} v_i x_i = \sum_{i=1}^{n} v_i E(X_i \mid \tilde{\theta}),$$

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.

The raw-to-theta score conversion tables were built up using above-described inversed TCC approach, based on which each student receives a theta score estimate corresponding to their raw score. 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:

o   At the lower end of the scale, for any theta estimates that were lower than -2.50, 0.25 was subtracted from the preceding adjusted theta value that was within the range.

o   At the higher end of the scale, for any theta estimates that were higher than 3.00, 0.25 was added to the previous theta value that was within the range.

o   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

The scaling coefficients, M_1^E and M_2^E, in Table 6.11 were then applied to derive the scale score of each student from their theta score estimate as follows. Note that the scaling coefficients were determined in the 2018 standard review and has been used in subsequent years of test administration.

$$\text{Scale Score} = (M_1^E \theta) + M_2^E$$

Table 6.11. Operational Scaling Coefficients

Grade	Slope (M_1^E)	Intercept (M_2^E)
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

Grade	Slope (M_1^E)	Intercept (M_2^E)
6	18.191784	600.432302
7	18.559827	600.499091
8	18.115200	600.640639

6.6.1. *Raw Score-to-Scale Score Conversion Tables, CSEM and Performance Levels*

The scale score is the basic score for the NYSTP. The PBT raw score-to-scale score (RSSS) conversion tables were based on the total number correct are presented in Appendix P, Tables P1–P12. NYSED made a policy decision to add two scale score points for each grade and content area for students who tested on CBT, which resulted in separate RSSS conversion tables for CBT. This decision was made as a result of advice provided by the Technical Advisory Committee.

The conditional standard error of measurement (CSEM) of a scale score is calculated as follows to include in the RSSS table:

$$\text{CSEM (Scale Score)} = M_1^E \frac{1}{\sqrt{I(\hat{\theta})}},$$

where $\hat{\theta}$ is the theta estimate corresponding to the scale score, $I(\hat{\theta})$ is the value of the test information function (TIF) at $\hat{\theta}$ and M_1^E is the scaling coefficient defined in Table 6.11.

The final element of the raw-score-to-scale-score tables is the application of the performance level cut scores. Scale score cuts were set in the summer of 2018 through standard review and can be applied to the 2022 scale scores because the 2022 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 and 6.6. 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	535–582	583–601	602–628	629–656
4	528–583	584–602	603–618	619–654
5	518–593	594–608	609–621	622–658
6	517–589	590–601	602–613	614–662
7	518–590	591–606	607–622	623–656
8	502–583	584–602	603–616	617–657

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	522–586	587–599	600–614	615–649
4	517–587	588–601	602–613	614–649
5	523–591	592–603	604–615	616-648
6	529–591	592–603	604–615	616-650
7	518–592	593–605	606-617	618-647
8	524–595	596–609	610-621	622-656

6.7. Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2022 TCCs were generated using final item parameters for all reporting test items administered in Spring 2022. 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 6.1–6.24.
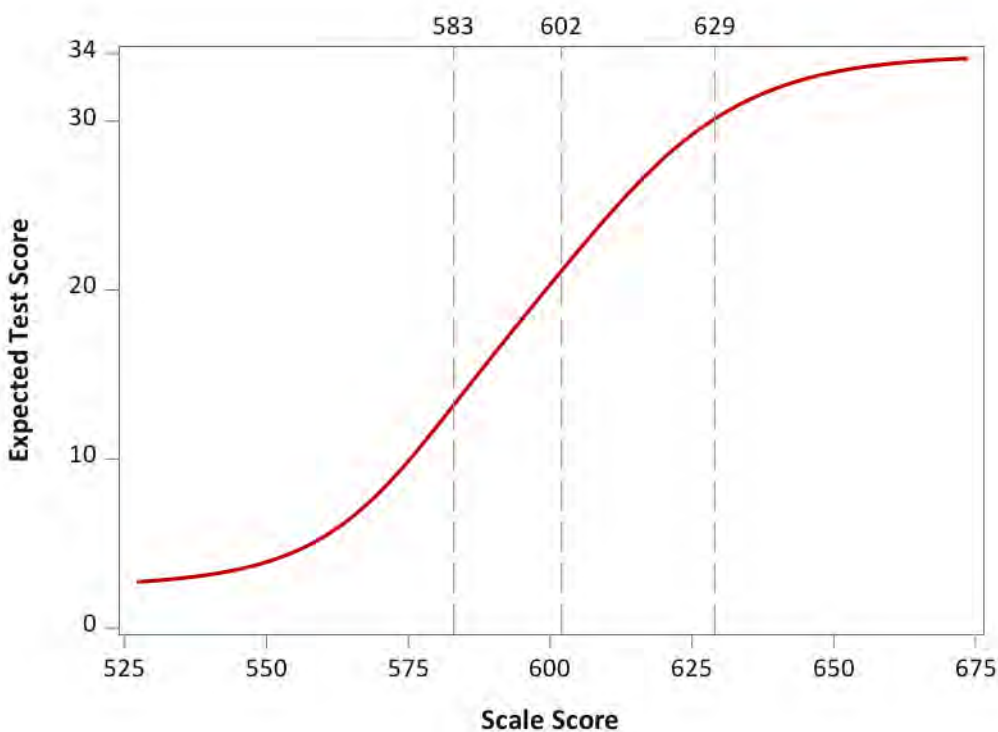
Figure 6.1. ELA Grade 3 TCC
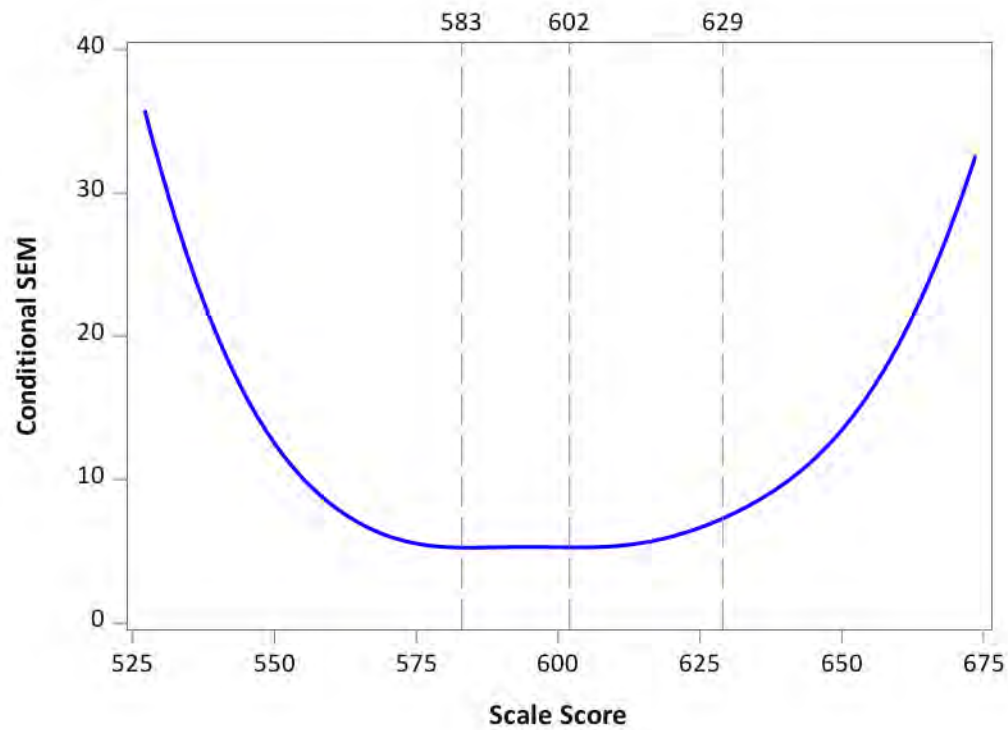
Figure 6.2. ELA Grade 3 CSEM Curve
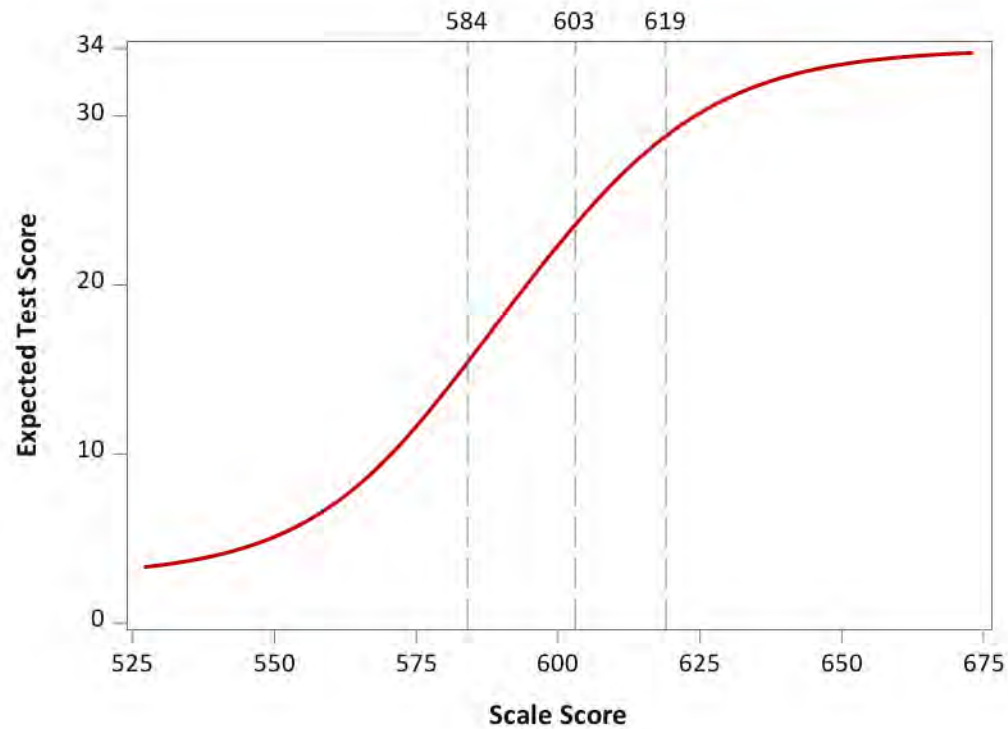
Figure 6.3. ELA Grade 4 TCC
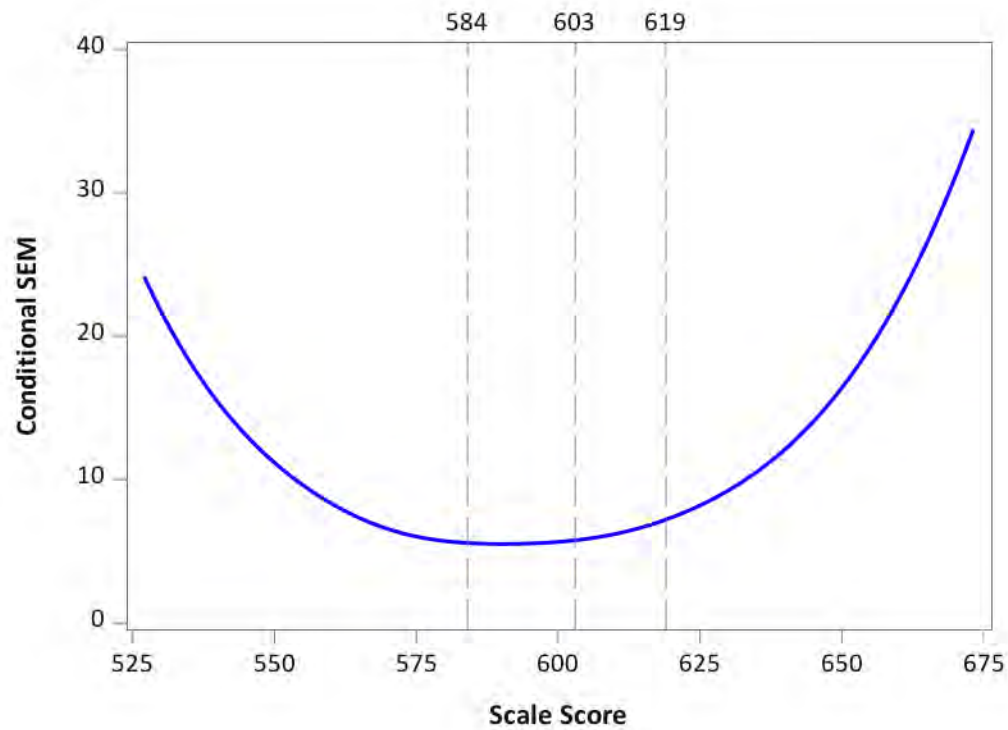
Figure 6.4. ELA Grade 4 CSEM Curve
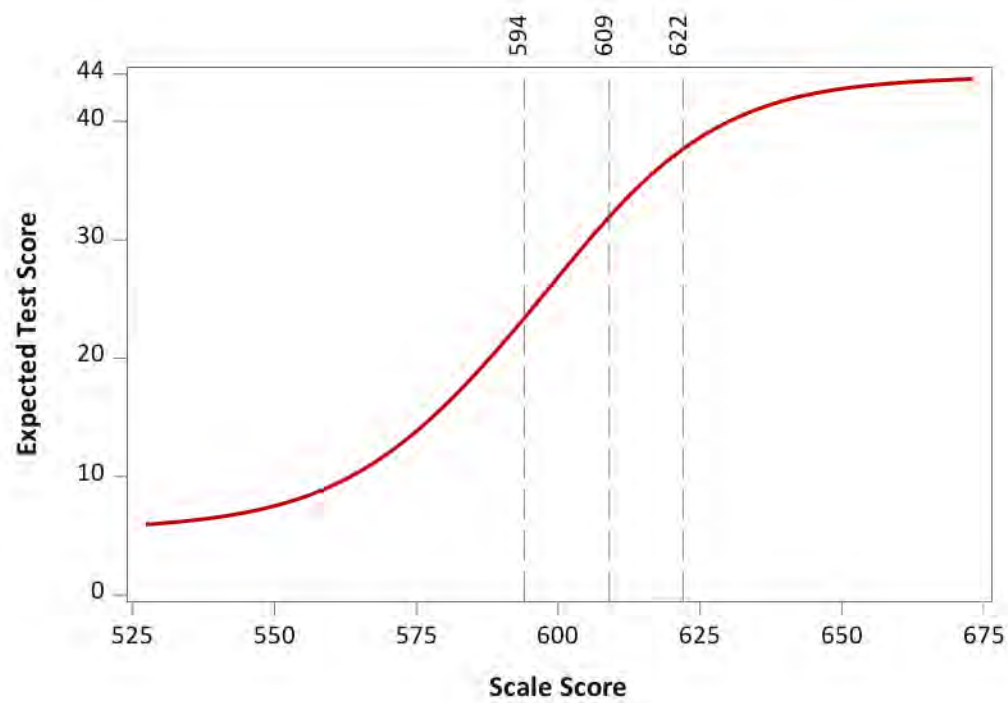
Figure 6.5. ELA Grade 5 TCC
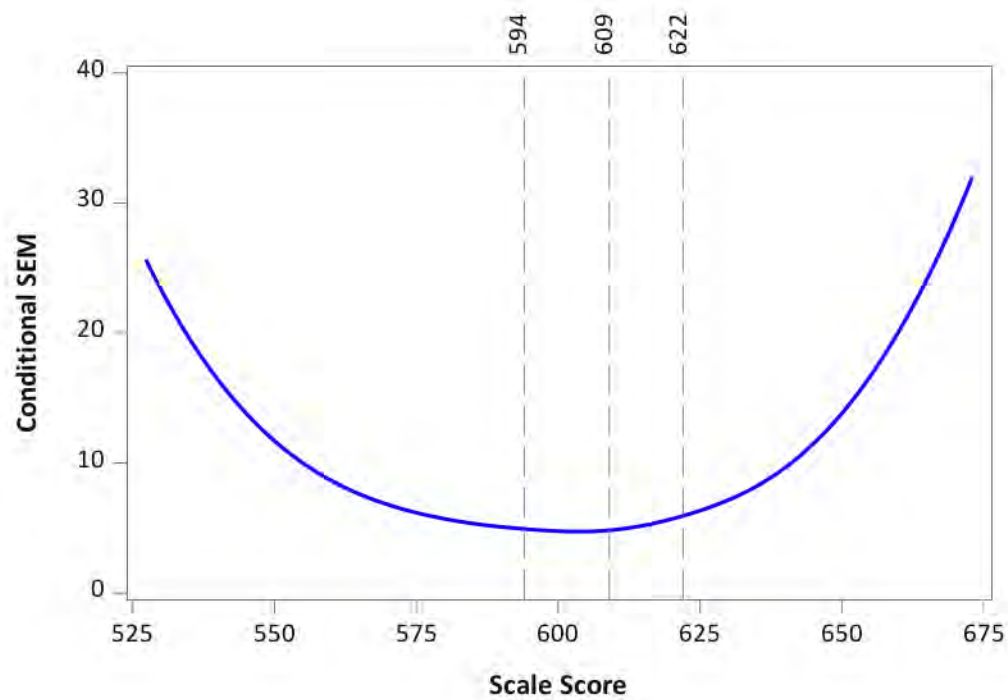
Figure 6.6. ELA Grade 5 CSEM Curve
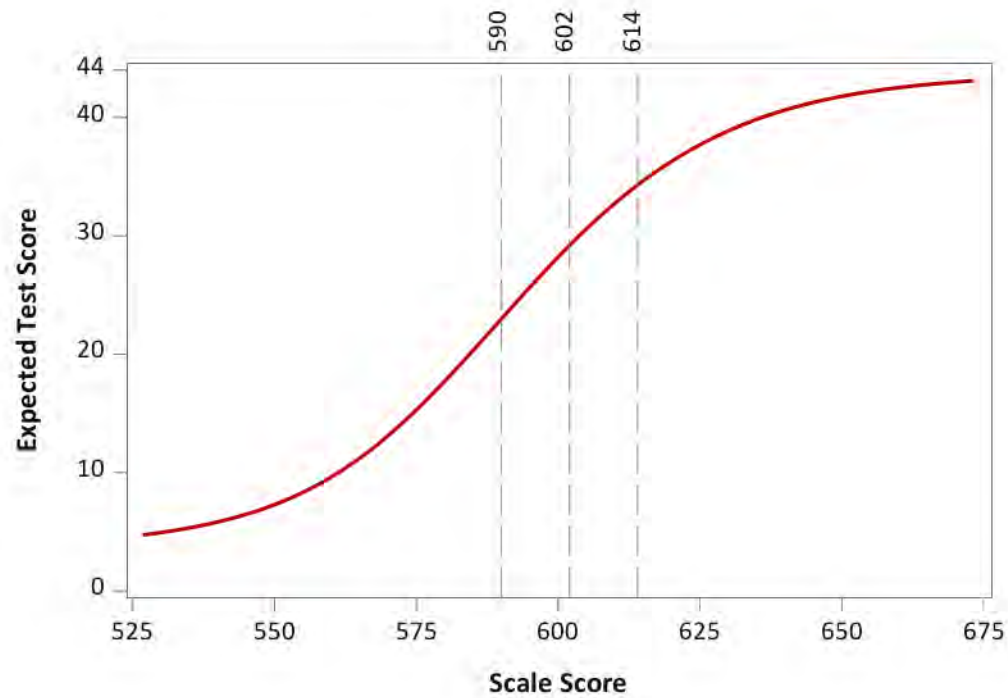
Figure 6.7. ELA Grade 6 TCC
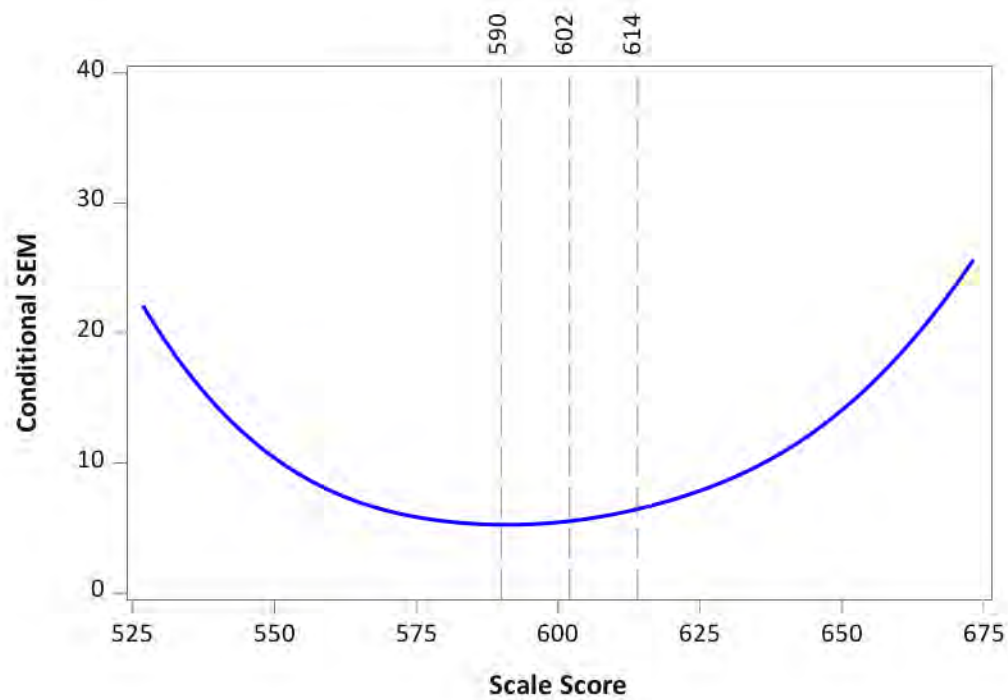
Figure 6.8. ELA Grade 6 CSEM Curve
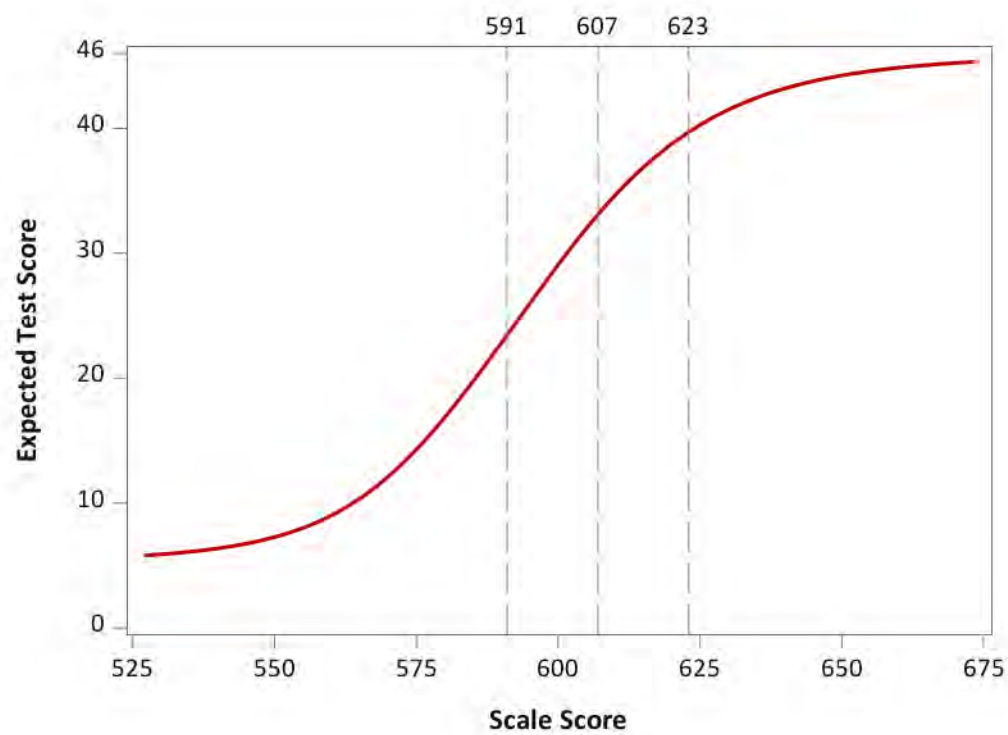
Figure 6.9. ELA Grade 7 TCC
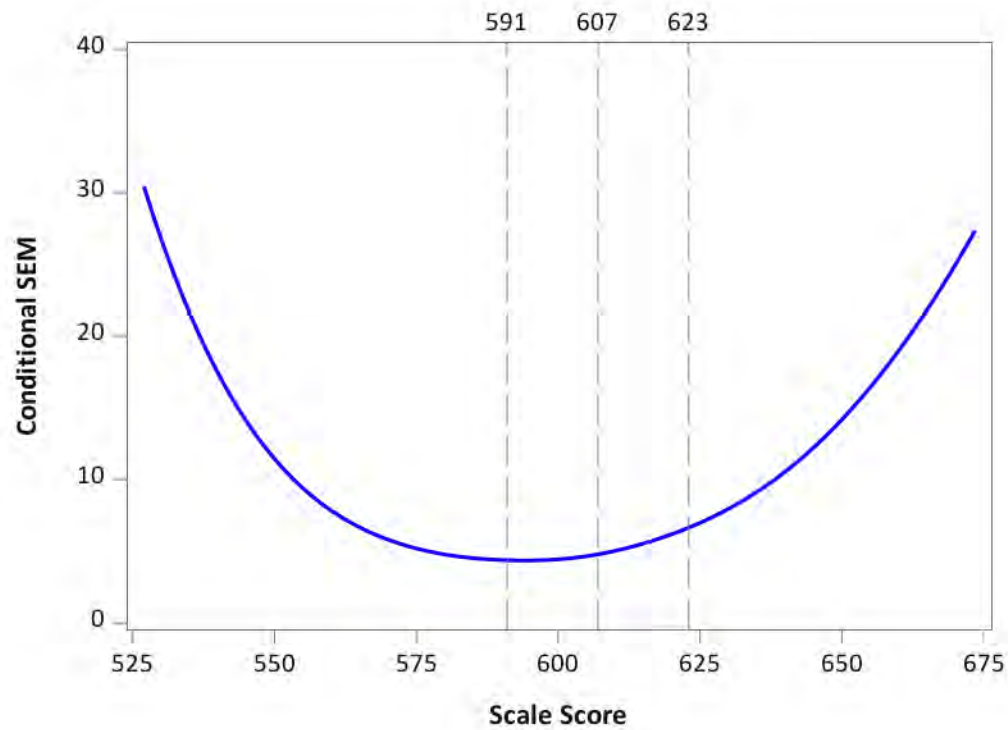
Figure 6.10. ELA Grade 7 CSEM Curve
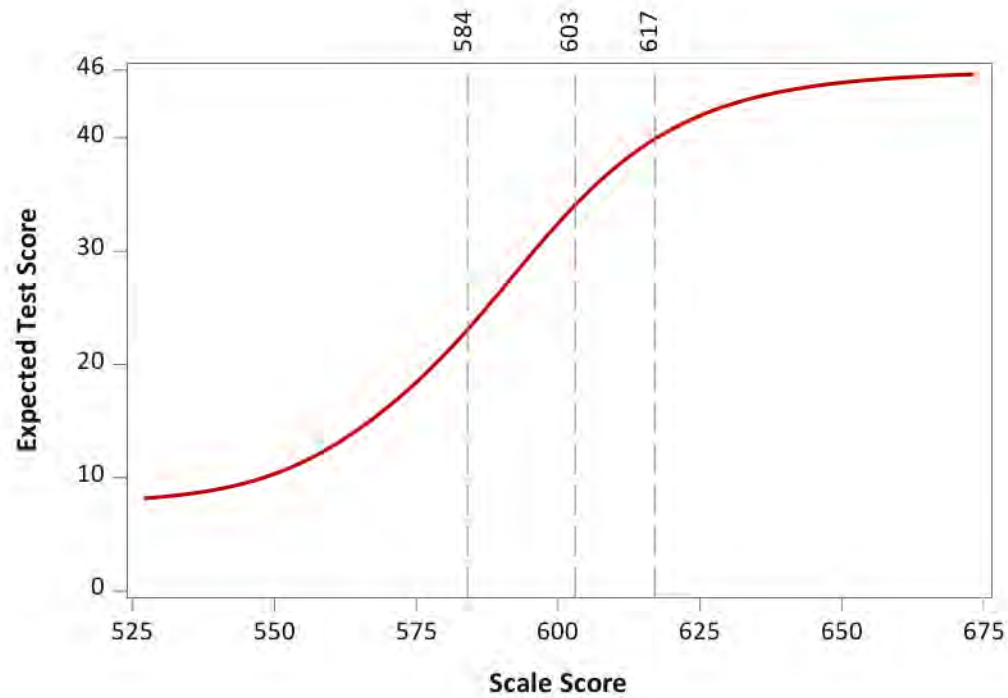
Figure 6.11. ELA Grade 8 TCC
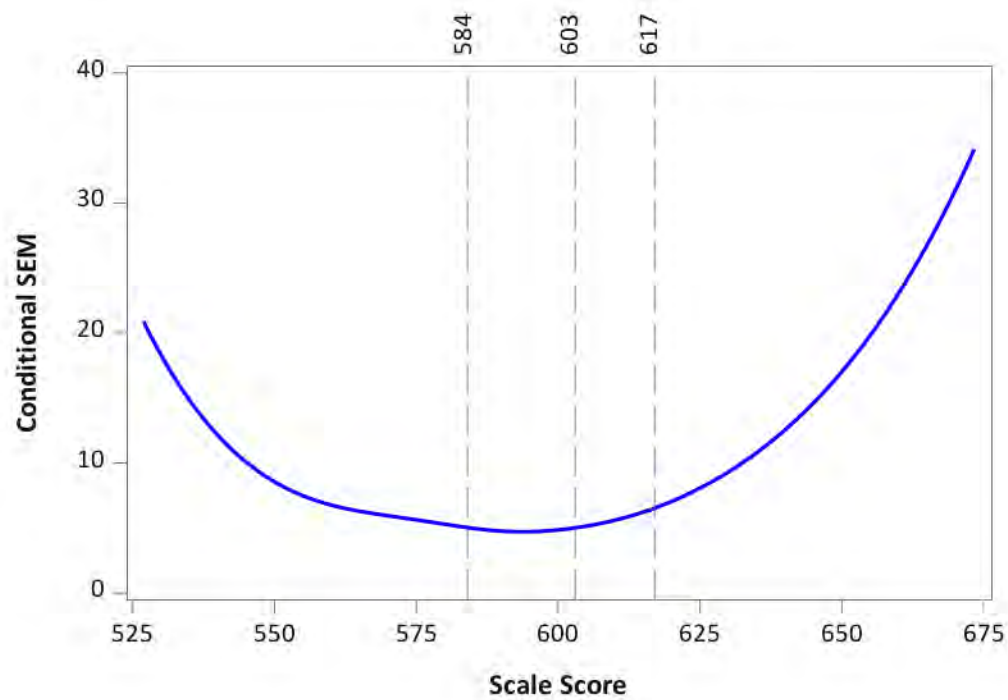
Figure 6.12. ELA Grade 8 CSEM Curve
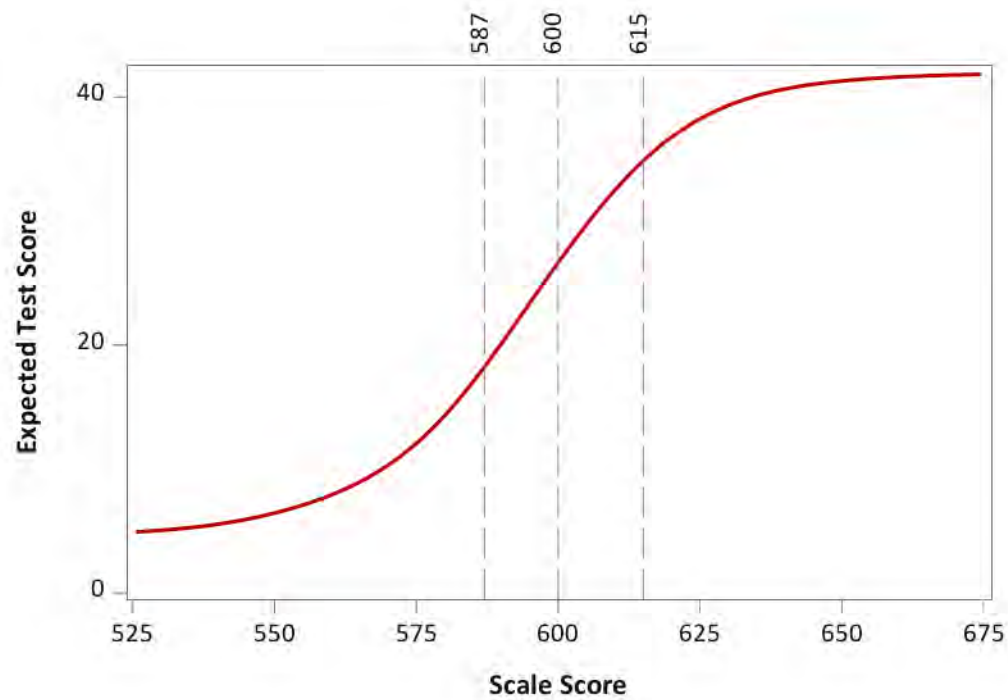
Figure 6.13. Mathematics Grade 3 TCC
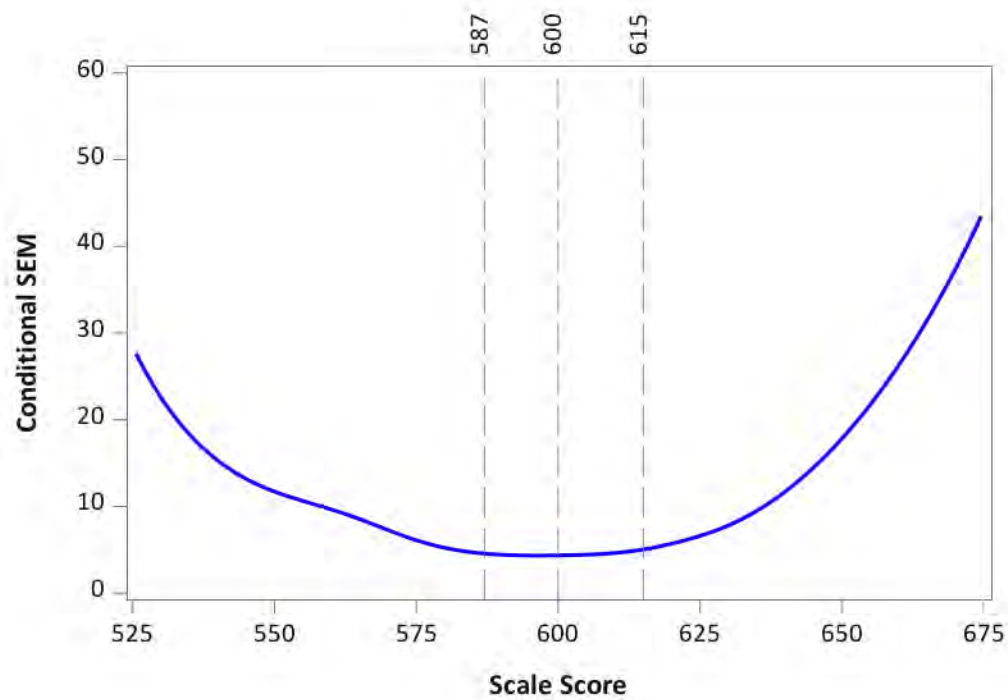
Figure 6.14. Mathematics Grade 3 CSEM Curve
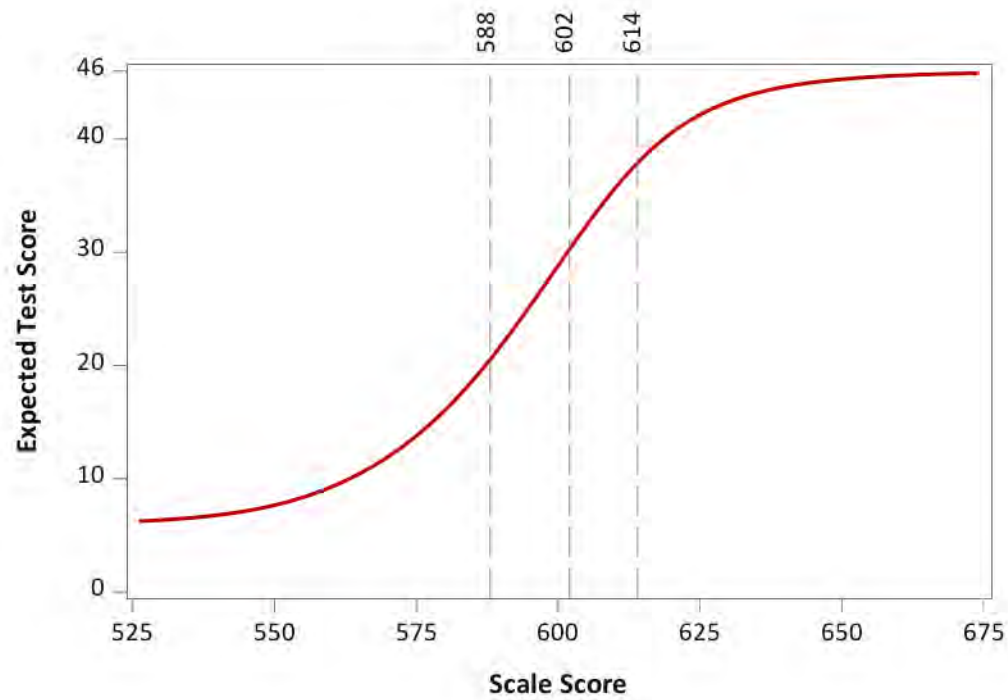
Figure 6.15. Mathematics Grade 4 TCC
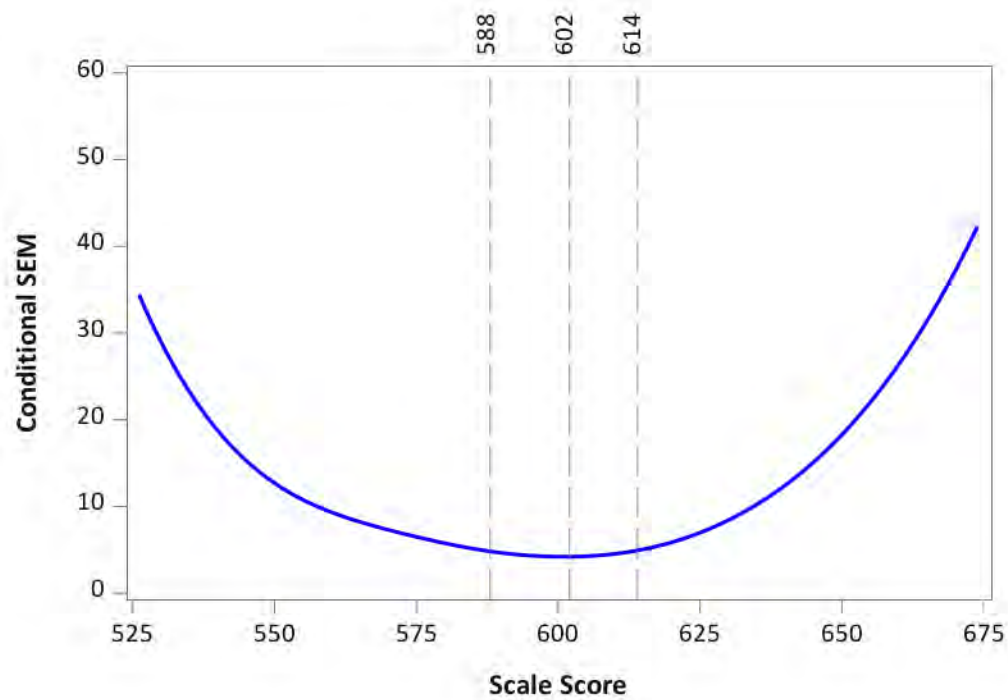
Figure 6.16. Mathematics Grade 4 CSEM Curve
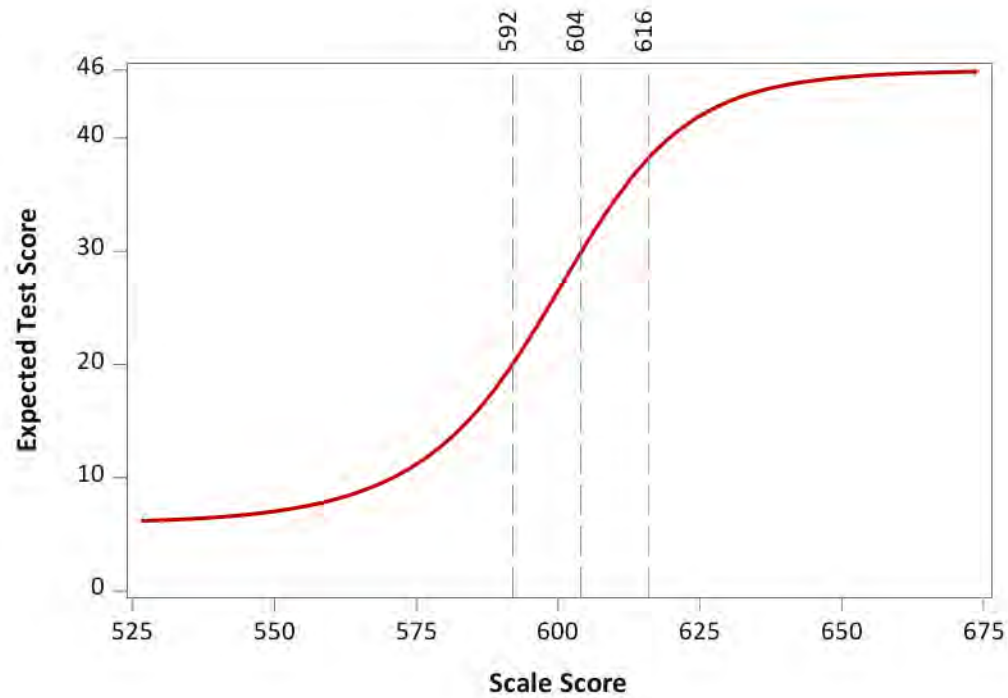
Figure 6.17. Mathematics Grade 5 TCC
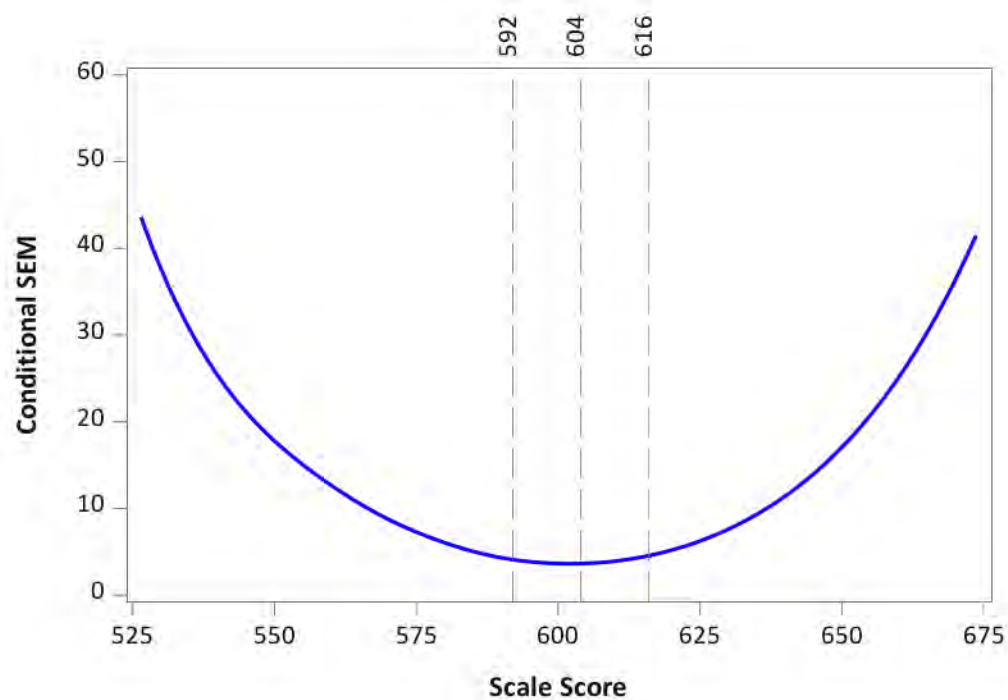
Figure 6.18. Mathematics Grade 5 CSEM Curve
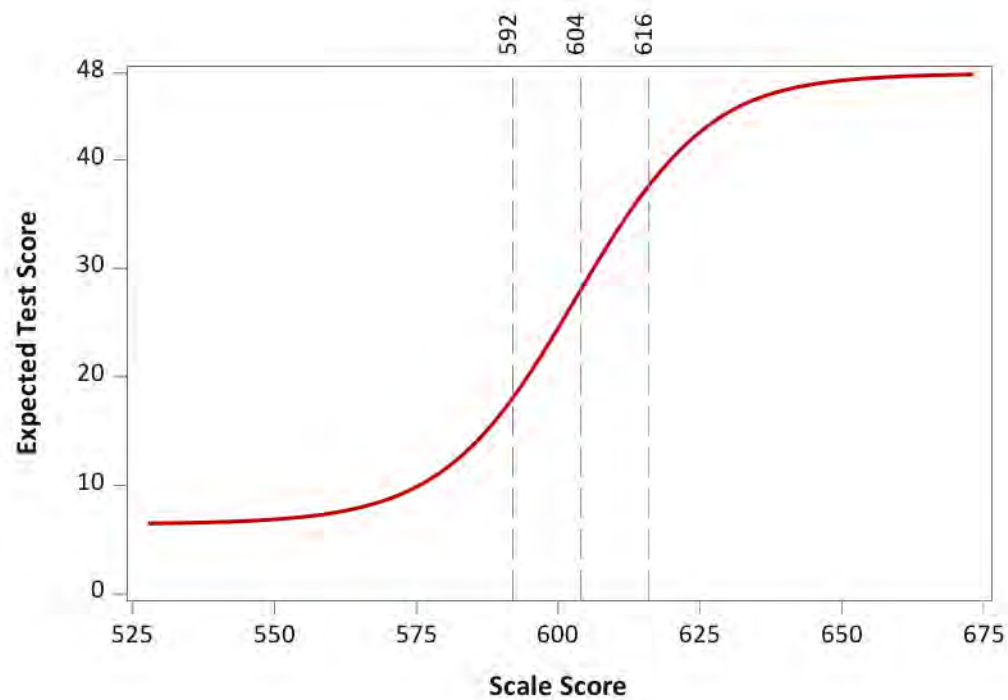
Figure 6.19. Mathematics Grade 6 TCC
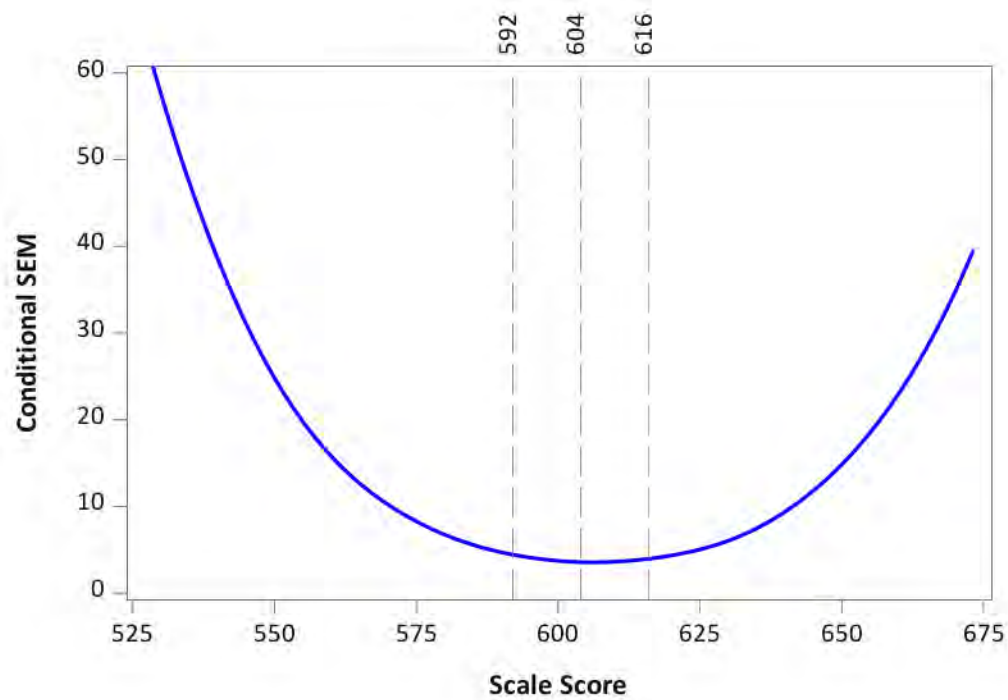
Figure 6.20. Mathematics Grade 6 CSEM Curve
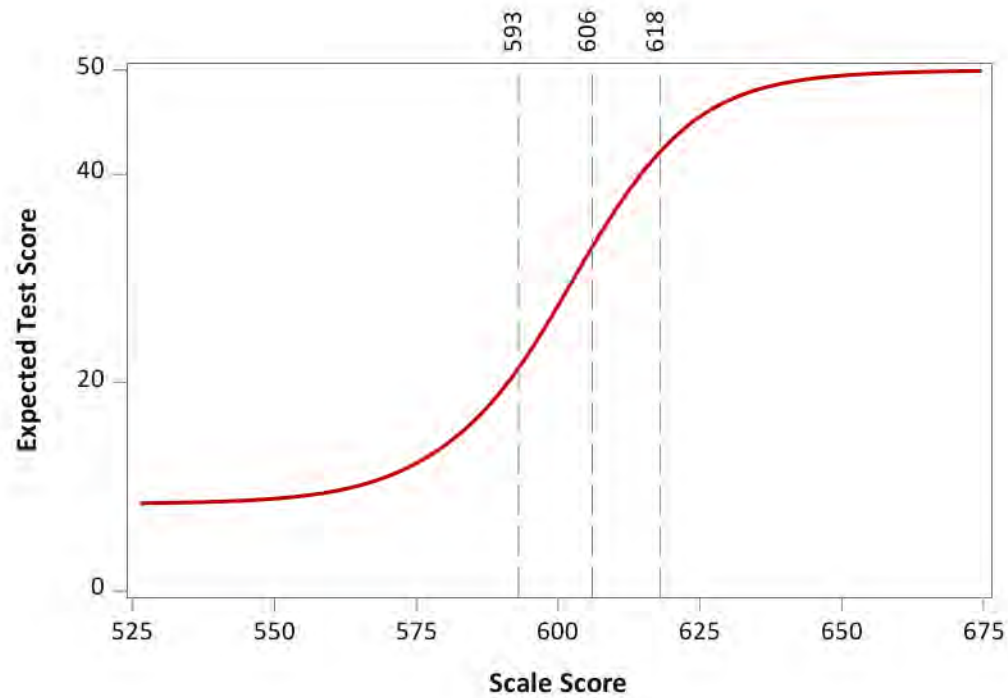
Figure 6.21. Mathematics Grade 7 TCC
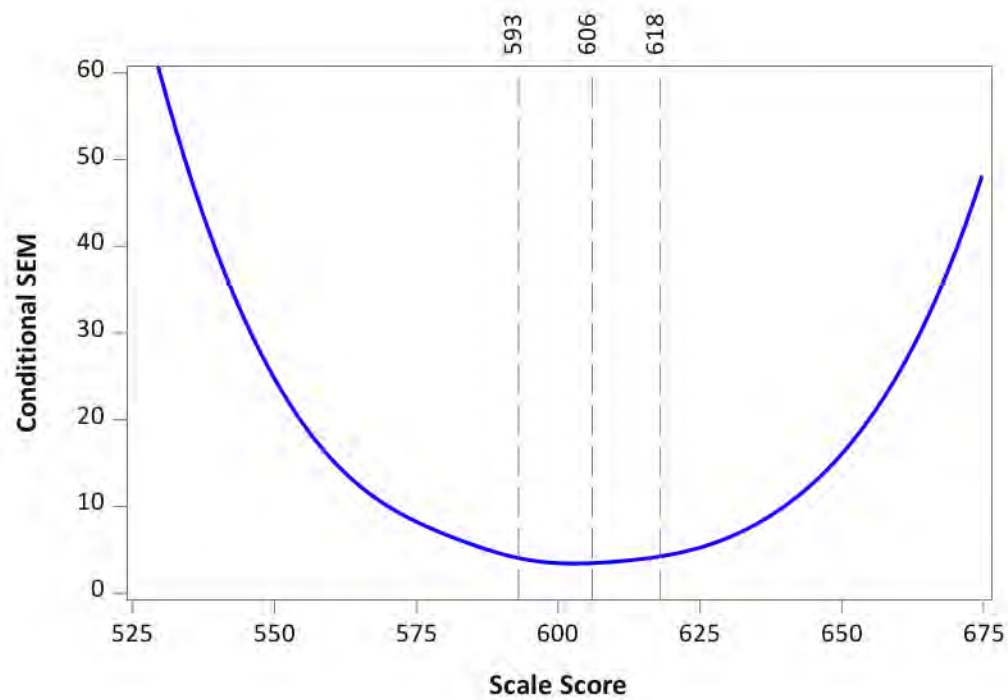
Figure 6.22. Mathematics Grade 7 CSEM Curve
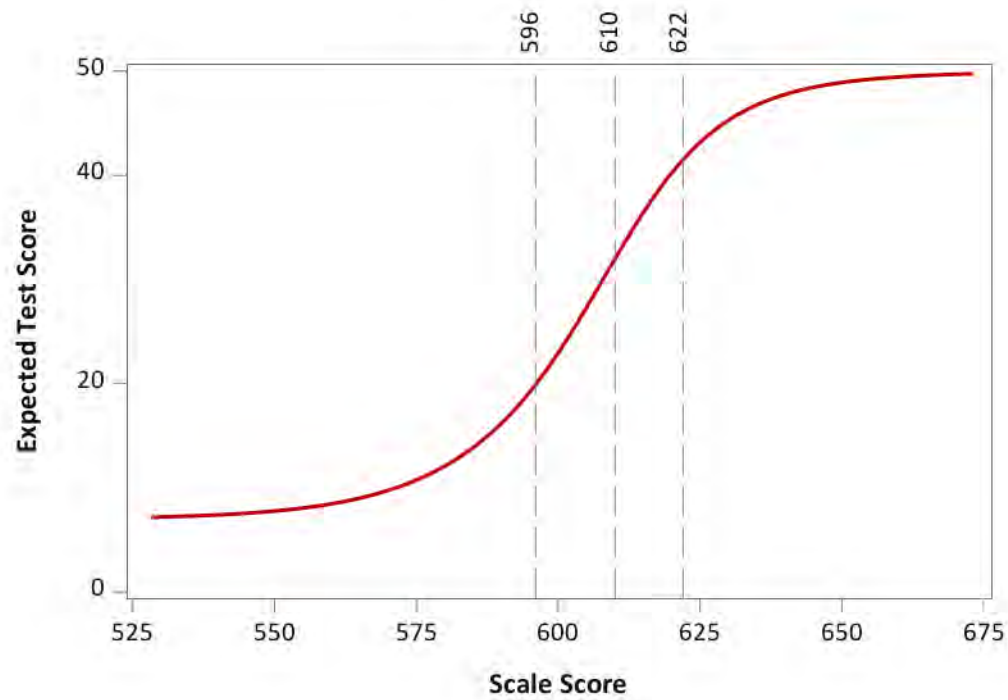
Figure 6.23. Mathematics Grade 8 TCC
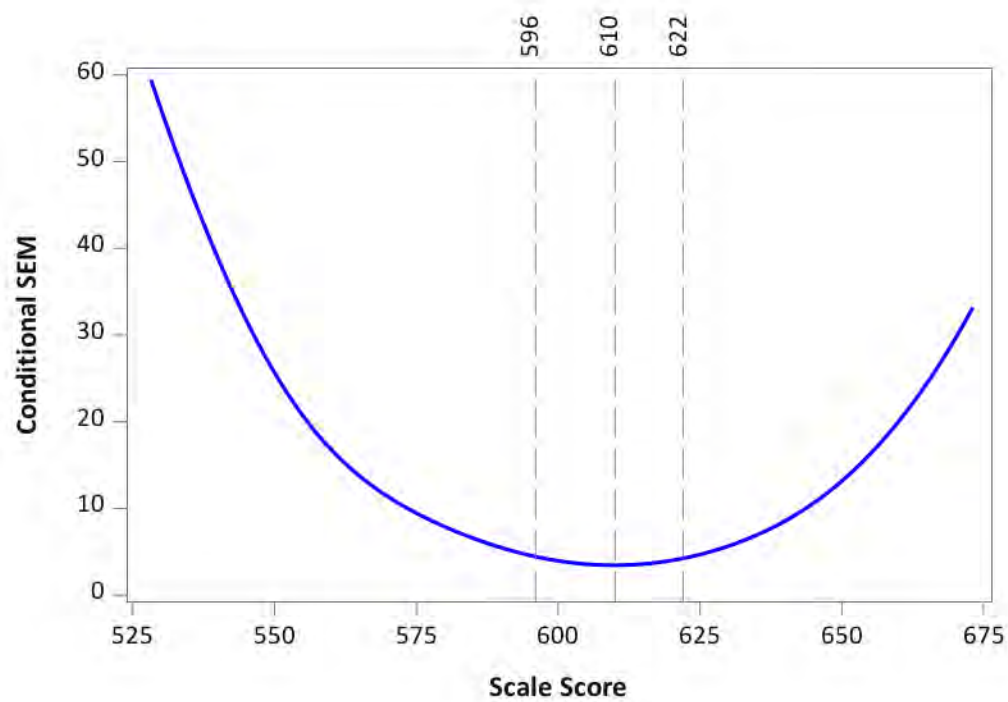
Figure 6.24. Mathematics Grade 8 CSEM Curve

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 0.88 to 0.92. Grades 3–8 Mathematics reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.91 to 0.94. The reliabilities were similar across grades and slightly higher for the Mathematics tests than for the ELA tests. All reliabilities were at least 0.88 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	0.61	0.31	0.90	152,259	34	19.23	7.03
4	0.62	0.42	0.84	153,437	34	19.96	7.29
5	0.63	0.30	0.87	159,499	44	27.12	8.86
6	0.66	0.44	0.86	157,750	44	28.29	8.43
7	0.65	0.34	0.88	144,053	46	29.92	9.19
8	0.67	0.43	0.84	138,380	46	31.05	9.39

Table 7.2. ELA Test 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	152,259	25	34	0.89	2.38	0.90	2.27
4	153,437	25	34	0.88	2.50	0.89	2.36
5	159,499	35	44	0.90	2.86	0.90	2.74
6	157,750	35	44	0.88	2.88	0.89	2.74
7	144,053	36	46	0.90	2.90	0.91	2.74
8	138,380	36	46	0.91	2.87	0.92	2.73

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	0.60	0.22	0.96	153,424	42	23.80	9.94
4	0.60	0.25	0.90	156,788	46	25.84	11.32
5	0.55	0.24	0.89	151,633	46	23.71	11.78
6	0.52	0.12	0.79	147,320	48	23.17	11.84
7	0.56	0.30	0.81	141,730	50	26.43	12.08
8	0.50	0.15	0.83	89,746	50	22.24	11.51

Table 7.4. Mathematics Test 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	153,424	34	42	0.91	2.91	0.92	2.73
4	156,788	38	46	0.93	3.05	0.93	2.89
5	151,633	38	46	0.93	3.05	0.94	2.87
6	147,320	39	48	0.93	3.18	0.93	3.04
7	141,730	41	50	0.92	3.33	0.93	3.15
8	89,746	41	50	0.93	3.15	0.93	3.04

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

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	152,259	18	0.81	1.73	0.81	1.72
4	153,437	18	0.80	1.79	0.80	1.78
5	159,499	28	0.84	2.24	0.85	2.23
6	157,750	28	0.82	2.24	0.82	2.24
7	144,053	28	0.84	2.22	0.84	2.22
8	138,380	28	0.86	2.24	0.86	2.23

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

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	153,424	27	0.88	2.06	0.89	2.04
4	156,788	31	0.91	2.24	0.91	2.23
5	151,633	31	0.91	2.31	0.91	2.29
6	147,320	31	0.90	2.38	0.90	2.37
7	141,730	33	0.89	2.49	0.89	2.48
8	89,746	33	0.89	2.48	0.89	2.48

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

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	152,259	7	16	0.86	1.38	0.87	1.34
4	153,437	7	16	0.85	1.50	0.86	1.44
5	159,499	7	16	0.85	1.48	0.86	1.43
6	157,750	7	16	0.85	1.48	0.86	1.43
7	144,053	8	18	0.89	1.49	0.90	1.42
8	138,380	8	18	0.89	1.43	0.90	1.35

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

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

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	153,424	7	15	0.82	1.82	0.83	1.78
4	156,788	7	15	0.82	1.85	0.83	1.80
5	151,633	7	15	0.85	1.74	0.86	1.69
6	147,320	8	17	0.84	1.91	0.85	1.85
7	141,730	8	17	0.86	1.92	0.86	1.89
8	89,746	8	17	0.86	1.73	0.86	1.72

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

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, 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 ELLs using accommodations specific to their ELL status (ELL/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 ELLs 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[5] 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	152,259	0.89	2.36	0.90	2.26
Gender	Female	75,603	0.88	2.36	0.89	2.25
	Male	76,654	0.89	2.36	0.90	2.26
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	16,654	0.88	2.31	0.89	2.18
	African American	20,645	0.89	2.39	0.90	2.29
	Hispanic	41,198	0.88	2.38	0.89	2.30
	American Indian	1,105	0.88	2.40	0.89	2.29
	Multiracial	5,680	0.89	2.34	0.90	2.22
	Pacific Islander	346	0.88	2.42	0.89	2.30
	White	66,515	0.87	2.33	0.88	2.23
NRC	New York	49,076	0.90	2.38	0.91	2.25
	Big 4 Cities	6,399	0.90	2.39	0.90	2.29
	Urban/Suburban	12,242	0.88	2.38	0.88	2.31
	Rural	9,112	0.86	2.33	0.87	2.28
	Average Needs	42,833	0.86	2.32	0.87	2.25
	Low Needs	19,308	0.85	2.26	0.86	2.16
	Charter School	6,417	0.88	2.36	0.89	2.26
	Religious and Independent	6,872	0.87	2.37	0.88	2.26

---

[5] In 2021–22, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SWD	All Codes	21,323	0.87	2.38	0.87	2.32
SUA	All Codes	17,799	0.86	2.37	0.87	2.32
ELL	ELL=Y	17,051	0.86	2.43	0.87	2.35
SWD/SUA	SWD & SUA codes	15,787	0.85	2.37	0.86	2.32
ELL/SUA	SUA & ELL codes	2,519	0.83	2.38	0.84	2.33

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	153,437	0.88	2.48	0.89	2.36
Gender	Female	76,213	0.88	2.46	0.89	2.34
	Male	77,222	0.88	2.49	0.89	2.37
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	17,975	0.88	2.35	0.89	2.22
	African American	21,452	0.88	2.52	0.89	2.40
	Hispanic	41,437	0.87	2.53	0.89	2.41
	American Indian	1,139	0.88	2.52	0.89	2.39
	Multiracial	5,317	0.89	2.45	0.90	2.32
	Pacific Islander	328	0.88	2.47	0.89	2.33
	White	65,678	0.87	2.44	0.88	2.33
NRC	New York	51,010	0.89	2.48	0.91	2.32
	Big 4 Cities	6,271	0.89	2.50	0.90	2.38
	Urban/Suburban	12,589	0.87	2.53	0.88	2.43
	Rural	9,102	0.86	2.50	0.87	2.42
	Average Needs	41,922	0.86	2.46	0.87	2.37
	Low Needs	19,225	0.84	2.32	0.86	2.24
	Charter School	6,173	0.87	2.45	0.88	2.35
	Religious and Independent	7,145	0.87	2.48	0.89	2.36
SWD	All Codes	21,996	0.86	2.51	0.87	2.42
SUA	All Codes	20,395	0.86	2.50	0.87	2.42
ELL	ELL=Y	15,830	0.84	2.56	0.85	2.45
SWD/SUA	SWD & SUA codes	17,702	0.85	2.49	0.86	2.41
ELL/SUA	SUA & ELL codes	2,819	0.81	2.49	0.82	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	159,499	0.90	2.84	0.90	2.73
Gender	Female	79,229	0.89	2.80	0.90	2.70
	Male	80,262	0.90	2.86	0.91	2.75
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	16,967	0.89	2.62	0.90	2.52
	African American	25,980	0.89	2.91	0.90	2.79
	Hispanic	44,467	0.89	2.91	0.90	2.80
	American Indian	1,153	0.89	2.87	0.90	2.75
	Multiracial	5,200	0.90	2.80	0.91	2.69
	Pacific Islander	369	0.89	2.80	0.90	2.69
	White	65,243	0.89	2.79	0.90	2.69
NRC	New York	51,635	0.90	2.83	0.91	2.70
	Big 4 Cities	6,166	0.90	2.93	0.91	2.81
	Urban/Suburban	12,326	0.88	2.91	0.89	2.82
	Rural	9,037	0.88	2.89	0.89	2.80
	Average Needs	41,605	0.88	2.82	0.89	2.74
	Low Needs	18,660	0.86	2.66	0.87	2.59
	Charter School	13,198	0.88	2.80	0.89	2.72
	Religious and Independent	6,872	0.89	2.87	0.90	2.75
SWD	All Codes	23,734	0.87	2.95	0.88	2.86
SUA	All Codes	22,344	0.87	2.95	0.88	2.86
ELL	ELL=Y	14,179	0.85	3.00	0.86	2.90
SWD/SUA	SWD & SUA codes	19,323	0.86	2.95	0.87	2.87
ELL/SUA	SUA & ELL codes	3,012	0.82	2.96	0.83	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	157,750	0.88	2.85	0.89	2.73
Gender	Female	77,772	0.88	2.80	0.89	2.70
	Male	79,968	0.89	2.88	0.89	2.76
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	16,643	0.86	2.64	0.87	2.54
	African American	26,034	0.88	2.92	0.89	2.81
	Hispanic	45,394	0.88	2.91	0.89	2.80
	American Indian	1,097	0.88	2.87	0.89	2.77
	Multiracial	4,903	0.89	2.82	0.90	2.70
	Pacific Islander	327	0.86	2.80	0.87	2.70
	White	63,233	0.87	2.81	0.88	2.70

	Demographic Category	N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	New York	50,457	0.89	2.84	0.90	2.70
	Big 4 Cities	6,279	0.89	2.97	0.90	2.84
	Urban/Suburban	12,234	0.88	2.92	0.89	2.83
	Rural	8,722	0.88	2.89	0.88	2.80
	Average Needs	39,151	0.87	2.84	0.88	2.74
	Low Needs	17,642	0.84	2.69	0.85	2.61
	Charter School	13,862	0.86	2.82	0.87	2.74
	Religious and Independent	9,403	0.88	2.86	0.89	2.74
SWD	All Codes	23,077	0.87	2.96	0.87	2.88
SUA	All Codes	21,916	0.87	2.96	0.87	2.88
ELL	ELL=Y	13,264	0.84	2.97	0.84	2.89
SWD/SUA	SWD & SUA codes	18,581	0.86	2.96	0.87	2.88
ELL/SUA	SUA & ELL codes	3,085	0.81	2.94	0.81	2.88

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	144,053	0.90	2.86	0.91	2.73
Gender	Female	70,388	0.89	2.80	0.90	2.69
	Male	73,654	0.90	2.89	0.91	2.75
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	16,817	0.88	2.55	0.89	2.46
	African American	21,848	0.90	2.95	0.91	2.82
	Hispanic	39,724	0.90	2.93	0.91	2.80
	American Indian	1,018	0.89	2.91	0.90	2.78
	Multiracial	4,418	0.91	2.85	0.91	2.70
	Pacific Islander	321	0.90	2.74	0.91	2.60
	White	59,796	0.89	2.83	0.90	2.70
NRC	New York	52,798	0.90	2.78	0.91	2.65
	Big 4 Cities	6,256	0.90	3.01	0.91	2.85
	Urban/Suburban	11,730	0.90	2.98	0.91	2.86
	Rural	8,783	0.89	2.97	0.90	2.84
	Average Needs	36,922	0.89	2.87	0.90	2.75
	Low Needs	17,808	0.87	2.68	0.88	2.59
	Charter School	4,041	0.89	2.78	0.90	2.70
	Religious and Independent	5,715	0.90	2.92	0.91	2.76
SWD	All Codes	20,827	0.88	3.03	0.89	2.91
SUA	All Codes	19,648	0.88	3.03	0.89	2.91
ELL	ELL=Y	10,865	0.85	3.05	0.86	2.92
SWD/SUA	SWD & SUA codes	16,646	0.87	3.03	0.88	2.91

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL/SUA	SUA & ELL codes	2,536	0.82	3.02	0.83	2.91

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	138,380	0.91	2.83	0.91	2.72
Gender	Female	66,758	0.90	2.74	0.90	2.65
	Male	71,603	0.91	2.89	0.92	2.77
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	16,834	0.90	2.49	0.90	2.41
	African American	21,692	0.90	2.94	0.91	2.83
	Hispanic	38,553	0.90	2.91	0.91	2.8
	American Indian	990	0.90	2.87	0.90	2.76
	Multiracial	3,876	0.91	2.82	0.92	2.69
	Pacific Islander	346	0.89	2.76	0.90	2.67
	White	55,997	0.90	2.80	0.91	2.69
NRC	New York	52,175	0.90	2.74	0.91	2.65
	Big 4 Cities	6,243	0.91	3.03	0.92	2.88
	Urban/Suburban	11,379	0.90	3.01	0.91	2.89
	Rural	8,762	0.90	2.92	0.91	2.82
	Average Needs	33,960	0.90	2.86	0.91	2.75
	Low Needs	16,179	0.89	2.61	0.89	2.53
	Charter School	3,763	0.90	2.74	0.91	2.67
	Religious and Independent	5,919	0.91	2.92	0.92	2.74
SWD	All Codes	19,867	0.88	3.06	0.89	2.95
SUA	All Codes	18,936	0.88	3.05	0.89	2.94
ELL	ELL=Y	10,190	0.85	3.12	0.86	2.97
SWD/SUA	SWD & SUA codes	16,032	0.87	3.06	0.88	2.95
ELL/SUA	SUA & ELL codes	2,234	0.82	3.07	0.83	2.96

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	153,424	0.91	2.88	0.92	2.72
Gender	Female	76,723	0.91	2.90	0.92	2.74
	Male	76,699	0.92	2.86	0.93	2.69
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	15,794	0.91	2.64	0.92	2.49
	African American	22,902	0.91	2.90	0.92	2.74
	Hispanic	38,840	0.90	2.92	0.91	2.77

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	American Indian	1,106	0.91	2.88	0.92	2.72
	Multiracial	5,609	0.92	2.85	0.93	2.67
	Pacific Islander	352	0.91	2.83	0.92	2.67
	White	68,691	0.91	2.88	0.92	2.72
NRC	New York	44,325	0.92	2.85	0.93	2.68
	Big 4 Cities	6,277	0.91	2.83	0.92	2.67
	Urban/Suburban	11,257	0.90	2.91	0.91	2.77
	Rural	9,198	0.90	2.90	0.91	2.75
	Average Needs	41,459	0.90	2.89	0.91	2.74
	Low Needs	17,614	0.89	2.75	0.90	2.63
	Charter School	12,619	0.91	2.80	0.92	2.66
	Religious and Independent	10,675	0.91	3.00	0.92	2.82
SWD	All Codes	18,576	0.91	2.87	0.92	2.73
SUA	All Codes	14,667	0.90	2.86	0.91	2.73
ELL	ELL=Y	18,221	0.90	2.92	0.91	2.79
SWD/SUA	SWD & SUA codes	12,742	0.89	2.84	0.90	2.71
ELL/SUA	SUA & ELL codes	1,836	0.89	2.84	0.90	2.73

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	156,788	0.93	3.02	0.93	2.88
Gender	Female	78,390	0.92	3.05	0.93	2.91
	Male	78,395	0.93	2.99	0.94	2.84
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	17,235	0.92	2.81	0.93	2.62
	African American	23,883	0.92	3.03	0.93	2.91
	Hispanic	39,502	0.92	3.05	0.92	2.93
	American Indian	1,171	0.92	3.05	0.93	2.90
	Multiracial	5,375	0.93	3.00	0.94	2.84
	Pacific Islander	331	0.91	3.03	0.92	2.88
	White	69,161	0.92	3.01	0.93	2.88
NRC	New York	45,775	0.93	3.01	0.94	2.83
	Big 4 Cities	6,124	0.92	2.90	0.93	2.81
	Urban/Suburban	11,230	0.91	3.01	0.92	2.92
	Rural	9,333	0.91	3.01	0.92	2.92
	Average Needs	40,690	0.91	3.02	0.92	2.90
	Low Needs	17,805	0.90	2.89	0.91	2.73
	Charter School	12,616	0.92	3.02	0.93	2.86
	Religious and Independent	13,215	0.92	3.06	0.93	2.95

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SWD	All Codes	18,929	0.92	2.95	0.92	2.86
SUA	All Codes	16,717	0.91	2.94	0.91	2.86
ELL	ELL=Y	16,985	0.91	3.00	0.91	2.92
SWD/SUA	SWD & SUA codes	14,102	0.90	2.92	0.91	2.84
ELL/SUA	SUA & ELL codes	2,004	0.88	2.88	0.89	2.82

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	151,633	0.93	3.03	0.94	2.86
Gender	Female	75,536	0.93	3.05	0.94	2.89
	Male	76,087	0.94	3.01	0.94	2.83
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	16,058	0.93	2.78	0.94	2.60
	African American	23,810	0.92	3.03	0.93	2.88
	Hispanic	38,464	0.92	3.05	0.93	2.91
	American Indian	1,072	0.93	3.02	0.94	2.86
	Multiracial	4,952	0.94	3.01	0.94	2.82
	Pacific Islander	347	0.93	3.03	0.93	2.88
	White	66,816	0.93	3.04	0.94	2.88
NRC	New York	45,810	0.94	3.00	0.95	2.81
	Big 4 Cities	5,972	0.92	2.86	0.93	2.75
	Urban/Suburban	10,799	0.91	2.98	0.92	2.87
	Rural	9,168	0.92	3.01	0.92	2.89
	Average Needs	39,588	0.92	3.05	0.93	2.90
	Low Needs	17,463	0.91	2.95	0.92	2.80
	Charter School	12,071	0.93	3.05	0.93	2.89
	Religious and Independent	10,762	0.92	3.05	0.93	2.92
SWD	All Codes	18,671	0.91	2.9	0.92	2.79
SUA	All Codes	16,878	0.91	2.89	0.91	2.79
ELL	ELL=Y	14,595	0.89	2.93	0.90	2.84
SWD/SUA	SWD & SUA codes	14,296	0.90	2.85	0.90	2.77
ELL/SUA	SUA & ELL codes	1,927	0.87	2.80	0.87	2.74

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	147,320	0.93	3.16	0.93	3.03
Gender	Female	72,766	0.92	3.18	0.93	3.05
	Male	74,543	0.93	3.13	0.94	3.00

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	15,238	0.93	2.94	0.94	2.82
	African American	23,885	0.91	3.15	0.92	3.03
	Hispanic	38,168	0.91	3.18	0.92	3.07
	American Indian	1,002	0.92	3.17	0.93	3.05
	Multiracial	4,660	0.94	3.13	0.94	2.98
	Pacific Islander	304	0.92	3.19	0.93	3.09
	White	63,920	0.92	3.16	0.93	3.04
NRC	New York	43,612	0.93	3.13	0.94	3.00
	Big 4 Cities	6,081	0.91	3.08	0.92	2.96
	Urban/Suburban	10,426	0.90	3.14	0.91	3.03
	Rural	8,705	0.91	3.17	0.91	3.06
	Average Needs	36,451	0.92	3.18	0.92	3.07
	Low Needs	16,115	0.92	3.06	0.92	2.96
	Charter School	13,045	0.93	3.14	0.93	3.03
	Religious and Independent	12,885	0.92	3.19	0.93	3.04
SWD	All Codes	17,931	0.89	3.02	0.89	2.93
SUA	All Codes	16,298	0.88	3.03	0.89	2.94
ELL	ELL=Y	12,313	0.85	3.03	0.85	2.96
SWD/SUA	SWD & SUA codes	13,549	0.86	2.98	0.87	2.90
ELL/SUA	SUA & ELL codes	1,750	0.80	2.92	0.81	2.87

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	141,730	0.92	3.30	0.93	3.14
Gender	Female	69,942	0.92	3.31	0.93	3.15
	Male	71,772	0.93	3.28	0.93	3.11
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	15,574	0.93	2.96	0.94	2.78
	African American	23,723	0.91	3.32	0.92	3.19
	Hispanic	36,585	0.91	3.35	0.91	3.22
	American Indian	976	0.91	3.35	0.92	3.20
	Multiracial	4,271	0.93	3.27	0.94	3.08
	Pacific Islander	319	0.93	3.26	0.94	3.07
	White	60,195	0.92	3.29	0.93	3.14
NRC	New York	45,398	0.93	3.25	0.94	3.06
	Big 4 Cities	5,926	0.90	3.23	0.91	3.10
	Urban/Suburban	9,601	0.89	3.32	0.90	3.21
	Rural	8,577	0.90	3.34	0.90	3.21

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Average Needs	33,879	0.91	3.33	0.92	3.19
	Low Needs	15,458	0.91	3.14	0.92	3.00
	Charter School	12,138	0.92	3.28	0.93	3.12
	Religious and Independent	10,753	0.92	3.35	0.93	3.20
SWD	All Codes	17,241	0.88	3.22	0.89	3.13
SUA	All Codes	15,508	0.88	3.23	0.89	3.14
ELL	ELL=Y	9,884	0.85	3.24	0.86	3.16
SWD/SUA	SWD & SUA codes	12,879	0.86	3.19	0.87	3.11
ELL/SUA	SUA & ELL codes	1,520	0.77	3.11	0.78	3.06

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	89,746	0.93	3.15	0.93	3.04
Gender	Female	43,300	0.92	3.15	0.93	3.04
	Male	46,428	0.93	3.13	0.93	3.03
	Non-Binary	-	-	-	-	-
Ethnicity	Asian	7,278	0.94	3.08	0.95	2.88
	African American	15,962	0.92	3.06	0.92	2.98
	Hispanic	24,336	0.92	3.12	0.92	3.03
	American Indian	624	0.92	3.11	0.92	3.01
	Multiracial	2,404	0.92	3.14	0.93	3.03
	Pacific Islander	189	0.93	3.16	0.94	3.01
	White	38,855	0.92	3.18	0.92	3.08
NRC	New York	26,541	0.93	3.14	0.94	3.00
	Big 4 Cities	4,929	0.92	2.90	0.93	2.83
	Urban/Suburban	6,536	0.88	3.02	0.89	2.97
	Rural	7,171	0.90	3.11	0.90	3.05
	Average Needs	22,475	0.90	3.18	0.90	3.10
	Low Needs	7,788	0.92	3.22	0.92	3.08
	Charter School	6,476	0.93	3.10	0.94	2.99
	Religious and Independent	7,830	0.93	3.17	0.94	3.04
SWD	All Codes	13,783	0.88	2.93	0.88	2.90
SUA	All Codes	12,468	0.88	2.94	0.88	2.90
ELL	ELL=Y	7,875	0.89	2.95	0.89	2.91
SWD/SUA	SWD & SUA codes	10,539	0.86	2.90	0.87	2.87
ELL/SUA	SUA & ELL codes	1,100	0.82	2.81	0.82	2.80

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 2.27 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 2.16 to 3.35, 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 P. Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000).

Appendix O 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, 65–72% 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 0.53 to 0.59. These are between "moderate" and "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 71–76% of students were estimated to be classified consistently to one of the four performance categories, and kappa ranged from 0.62 to 0.66. 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	152,259	72%	28%	0.59
4	153,437	66%	34%	0.54
5	159,499	70%	30%	0.59
6	157,750	65%	35%	0.53
7	144,053	70%	30%	0.59
8	138,380	68%	32%	0.57
Mathematics				
3	153,424	71%	29%	0.62
4	156,788	71%	29%	0.62
5	151,633	74%	26%	0.65
6	147,320	74%	26%	0.65
7	141,730	75%	25%	0.66
8	89,746	76%	24%	0.66

**Note.* Decision consistency was calculated based on the PBT conversion tables 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, 86–89% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.72 to 0.75. These are considered "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 89–92% of the classifications were estimated consistently, and kappa coefficients ranged from 0.78 to 0.82. 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	152,259	87%	13%	0.74
4	153,437	86%	14%	0.72
5	159,499	89%	11%	0.75
6	157,750	86%	14%	0.72
7	144,053	88%	12%	0.75
8	138,380	88%	12%	0.75
Mathematics				
3	153,424	89%	11%	0.78
4	156,788	90%	10%	0.79
5	151,633	91%	9%	0.82
6	147,320	91%	9%	0.82
7	141,730	92%	8%	0.82
8	89,746	92%	8%	0.82

**Note.* Decision consistency was calculated based on the PBT conversion tables 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 their underlying proficiency from 73% to 79% of the time across all performance levels and 90% to 92% 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 their true proficiency from 79% to 83% of the time across all performance levels and 93% to 95% of the time in regard to the Level III cut score.

Table 7.23. Decision Agreement (Accuracy) Estimates*

Grade	N-Count	Accuracy All Cuts	Accuracy Level III Cut
		ELA	
3	152,259	79%	90%
4	153,437	75%	90%
5	159,499	78%	92%
6	157,750	73%	90%
7	144,053	78%	92%
8	138,380	77%	91%
		Mathematics	
3	153,424	79%	93%
4	156,788	80%	93%
5	151,633	80%	93%
6	147,320	81%	94%
7	141,730	82%	94%
8	89,746	83%	95%

*Note. Decision agreement was calculated based on the PBT conversion tables 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 2022 Grades 3–8 ELA and Mathematics Tests. These include the scale score means, standard deviations, and performance level distributions for each grade's population and specific subgroups. Gender, ethnic identification, NRC, ELL, 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/SUA subgroup is defined as ELLs who use one or more ELL-related accommodation. The SWD/SUA subgroup is defined as examinees with disabilities who use one or more disability-related accommodation. 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 P.

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[6]. 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 ELLs, 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	
Grade	N-Count	Mean	SD
3	165,209	598.21	19.31
4	168,725	595.76	20.87
5	165,024	600.87	19.65
6	163,509	602.5	20.59
7	159,762	604.09	19.18
8	150,130	599.96	21.50

---

[6] In 2021–22, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

Table 8.2. ELA Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Reading	165,209	18	11.76	4.04
	Writing	165,209	16	7.39	3.77
4	Reading	168,725	18	11.47	4.07
	Writing	168,725	16	8.41	3.96
5	Reading	165,024	28	17.45	5.79
	Writing	165,024	16	9.52	4.00
6	Reading	163,509	28	18.28	5.40
	Writing	163,509	16	9.91	4.00
7	Reading	159,762	28	17.95	5.53
	Writing	159,762	18	12.11	4.50
8	Reading	150,130	28	18.19	5.91
	Writing	150,130	18	13.05	4.38

8.1.1.1. <u>ELA Grade 3</u>

Table 8.3 presents the scale score statistics and n-counts of demographic subgroups for Grade 3. The population scale score mean was 598.21 with a standard deviation of 19.31. 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 (608.49). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score of 586.68—12 scale score points below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and ELL subgroups scored 12–15 scale core points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 16 scale score points below the State mean.

Table 8.3. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	165,209	598.21	19.31
Gender	Female	82,152	600.33	18.94
	Male	83,055	596.13	19.45
	Non-Binary	-	-	-
Ethnicity	Asian	17,026	608.49	18.62
	African American	24,816	594.42	19.14
	Hispanic	44,664	594.07	18.49
	American Indian	1,188	597.75	18.72
	Multiracial	5,830	600.41	19.45
	Pacific Islander	370	600.07	19.46

Demographic Category		N-Count	Scale Score Mean	Scale Score SD
	White	71,199	599.50	18.91
NRC	New York	49,203	600.21	20.38
	Big 4 Cities	6,399	586.68	20.31
	Urban/Suburban	12,353	590.57	18.09
	Rural	9,252	592.64	16.32
	Average Needs	42,995	598.29	16.55
	Low Needs	19,829	607.34	16.03
	Charter	14,456	600.27	17.88
	Religious and Independent	10,722	589.59	23.58
SWD	All Codes	23,077	585.90	17.53
SUA	All Codes	12,081	584.95	16.24
ELL	ELL=Y	21,396	583.50	18.90
ELL/SUA	SUA & ELL codes	1,187	582.18	15.73
SWD/SUA	SWD & SUA codes	10,244	583.56	15.89

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 595.76 with a standard deviation of 20.87. 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, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (607.11). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 13 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored about 15–19 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 21 scale score points below the State mean.

Table 8.4. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD
State	All Students	168,725	595.76	20.87
Gender	Female	83,974	598.45	20.38
	Male	84,749	593.10	21.02
	Non-Binary	-	-	-
Ethnicity	Asian	18,478	607.11	20.20
	African American	25,991	590.89	20.16
	Hispanic	45,363	590.79	19.57
	American Indian	1,225	594.39	20.24
	Multiracial	5,581	598.09	20.93

	Demographic Category	N-Count	Scale Score Mean	SD
	Pacific Islander	354	598.15	19.79
	White	71,621	597.60	20.64
NRC	New York	51,191	597.45	22.18
	Big 4 Cities	6,271	582.44	21.32
	Urban/Suburban	12,618	587.40	19.23
	Rural	9,405	591.82	18.03
	Average Needs	42,363	596.53	18.06
	Low Needs	19,652	605.35	17.10
	Charter	13,884	597.71	18.58
	Religious and Independent	13,341	587.63	25.04
SWD	All Codes	23,972	580.76	18.71
SUA	All Codes	13,695	580.62	17.66
ELL	ELL=Y	19,974	576.85	19.56
ELL/SUA	SUA & ELL codes	1,227	574.30	16.01
SWD/SUA	SWD & SUA codes	11,086	578.37	17.00

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 600.87 with a standard deviation of 19.65. 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 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 (611.54). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 13 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored about 15–21 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 22 scale score points below the State mean.

Table 8.5. ELA Grade 5 Scale Score Distribution by Subgroup

	Demographic Category	N-Count	Scale Score Mean	SD
State	All Students	165,024	600.87	19.65
Gender	Female	81,758	603.46	18.65
	Male	83,256	598.33	20.26
	Non-Binary	-	-	-
Ethnicity	Asian	17,331	611.54	17.87
	African American	26,182	596.33	18.91
	Hispanic	44,743	597.06	18.42
	American Indian	1,160	598.68	18.66

Demographic Category		N-Count	Scale Score Mean	SD
	Multiracial	5,253	602.86	19.07
	Pacific Islander	371	604.48	18.20
	White	69,864	602.26	19.95
NRC	New York	51,772	602.42	19.80
	Big 4 Cities	6,166	587.80	20.96
	Urban/Suburban	12,359	593.92	18.19
	Rural	9,373	596.95	17.79
	Average Needs	42,022	601.76	17.11
	Low Needs	19,535	610.23	15.50
	Charter	13,284	603.05	16.92
	Religious and Independent	10,513	588.92	26.90
SWD	All Codes	24,153	586.16	18.47
SUA	All Codes	14,158	585.97	17.93
ELL	ELL=Y	17,399	579.96	20.07
ELL/SUA	SUA & ELL codes	1,225	578.75	16.14
SWD/SUA	SWD & SUA codes	11,392	583.69	17.37

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 602.50 with a standard deviation of 20.59. 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 ethnic groups, Asian students earned the highest mean score (613.54). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 10 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored about 16–24 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 24 scale score points below the State mean.

Table 8.6. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	SD
State	All Students	163,509	602.50	20.59
Gender	Female	80,555	605.58	19.86
	Male	82,942	599.52	20.84
	Non-Binary	-	-	-
Ethnicity	Asian	16,943	613.54	18.51
	African American	26,412	598.32	19.40
	Hispanic	45,881	598.64	19.22

Demographic Category		N-Count	Scale Score Mean	SD
	American Indian	1,103	601.38	19.32
	Multiracial	4,987	604.79	20.15
	Pacific Islander	330	605.98	18.31
	White	67,734	603.84	21.27
NRC	New York	50,630	604.00	20.68
	Big 4 Cities	6,280	592.34	21.04
	Urban/Suburban	12,251	594.90	19.24
	Rural	9,003	598.93	18.81
	Average Needs	39,528	603.55	18.56
	Low Needs	18,534	611.20	16.70
	Charter	14,216	604.53	17.95
	Religious and Independent	13,067	593.44	27.14
SWD	All Codes	23,457	586.48	18.04
SUA	All Codes	13,773	586.61	18.01
ELL	ELL=Y	16,419	578.85	19.33
ELL/SUA	SUA & ELL codes	1,221	578.95	14.25
SWD/SUA	SWD & SUA codes	10,759	583.93	17.21

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 604.09 with a standard deviation of 19.18. 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 from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (615.76). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 12 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored about 15–23 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 24 scale score points below the State mean.

Table 8.7. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	SD
State	All Students	159,762	604.09	19.18
Gender	Female	78,641	607.00	18.50
	Male	81,106	601.27	19.40
	Non-Binary	-	-	-
Ethnicity	Asian	17,349	615.76	17.64
	African American	26,934	600.25	18.09

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Hispanic	44,944	600.54	18.15
	American Indian	1,077	602.97	17.70
	Multiracial	4,648	605.58	19.33
	Pacific Islander	350	608.55	18.56
	White	64,349	604.93	19.36
NRC	New York	52,902	607.23	19.10
	Big 4 Cities	6,256	592.30	19.79
	Urban/Suburban	11,963	595.25	18.47
	Rural	9,002	598.96	17.66
	Average Needs	37,348	603.17	17.61
	Low Needs	18,167	611.12	16.00
	Charter	13,221	606.48	15.98
	Religious and Independent	10,903	598.07	24.50
SWD	All Codes	22,940	589.09	17.10
SUA	All Codes	13,673	588.45	17.27
ELL	ELL=Y	13,743	580.85	17.91
ELL/SUA	SUA & ELL codes	1,265	579.75	14.73
SWD/SUA	SWD & SUA codes	10,809	586.06	16.66

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 599.96 with a standard deviation of 21.50. Female students tended to outperform male students by 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 (611.77). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 12 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored about 17–27 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 28 scale score points below the State mean.

Table 8.8. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	150,130	599.96	21.50
Gender	Female	73,004	603.80	20.28
	Male	77,107	596.32	21.98
	Non-Binary	-	-	-
Ethnicity	Asian	17,134	611.77	19.53

Demographic Category		N-Count	Scale Score	
			Mean	SD
	African American	26,171	596.57	19.98
	Hispanic	42,312	596.26	20.67
	American Indian	1,075	598.47	19.02
	Multiracial	3,989	601.34	21.73
	Pacific Islander	368	603.72	19.29
	White	58,989	600.62	22.01
NRC	New York	52,287	602.84	20.52
	Big 4 Cities	6,249	587.85	23.09
	Urban/Suburban	11,375	590.98	21.21
	Rural	8,964	594.93	20.25
	Average Needs	34,215	598.95	20.23
	Low Needs	16,452	608.24	18.61
	Charter	12,373	603.86	17.58
	Religious and Independent	8,215	590.52	29.51
SWD	All Codes	21,673	583.42	19.47
SUA	All Codes	12,928	582.85	20.32
ELL	ELL=Y	12,581	573.42	21.22
ELL/SUA	SUA & ELL codes	1,029	572.24	18.22
SWD/SUA	SWD & SUA codes	10,284	580.07	19.47

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[7]. 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) outperformed High Needs (New York City, Big 4 Cities, Urban/Suburban, and Rural) and Average Needs districts. ELL, SWDs, and/or SUAs tended to under-perform the State population (All Students). Students taking the Chinese and Korean translations tended to outperform the other translation subgroups. This pattern of achievement was fairly consistent across all grades.

Table 8.9. Mathematics Scale Score Distribution Summary

Grade	N-Count	Scale Score	
		Mean	SD
3	166,446	596.21	20.54
4	169,535	595.48	22.21

[7] In 2021–22, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

Grade	N-Count	Scale Score	
		Mean	**SD**
5	163,950	595.20	21.65
6	160,087	596.99	20.18
7	154,425	597.51	20.07
8	97,284	595.72	20.66

Table 8.10. Mathematics Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	**Mean**	**SD**
3	Operations and Algebraic Thinking	166,446	19	11.32	4.99
	Number and Operations - Fractions	166,446	7	3.12	2.24
	Measurement and Data	166,446	10	5.97	2.52
4	Operations and Algebraic Thinking	169,535	10	6.81	2.53
	Number and Operations in Base Ten	169,535	11	6.03	3.18
	Number and Operations - Fractions	169,535	12	6.32	3.40
5	Number and Operations in Base Ten	163,950	13	7.07	3.48
	Number and Operations - Fractions	163,950	18	8.89	4.85
	Measurement and Data	163,950	13	6.60	3.89
6	Ratios and Proportional Relationships	160,087	12	5.84	2.97
	The Number System	160,087	9	4.58	2.63
	Expressions and Equations	160,087	19	9.80	5.20
7	Ratios and Proportional Relationships	154,425	14	7.52	3.53
	The Number System	154,425	11	6.45	3.10
	Expressions and Equations	154,425	14	7.27	3.69
8	Expressions and Equations	97,284	19	8.45	4.76
	Functions	97,284	14	6.11	3.49
	Geometry	97,284	13	5.58	3.43

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 596.21 with a standard deviation of 20.54. 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 (608.87). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—14 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored 13–17 scale score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 21 scale score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	166,446	596.21	20.54
Gender	Female	82,532	595.96	19.69
	Male	83,912	596.45	21.34
	Non-Binary	-	-	-
Ethnicity	Asian	17,324	608.87	19.45
	African American	24,465	591.03	20.57
	Hispanic	45,590	590.82	19.55
	American Indian	1182	595.72	20.28
	Multiracial	5,811	597.56	21.57
	Pacific Islander	378	598.51	20.18
	White	71,556	598.25	19.56
NRC	New York	50,278	597.38	21.23
	Big 4 Cities	6,490	582.38	21.64
	Urban/Suburban	12,680	587.59	19.53
	Rural	9,243	590.65	18.96
	Average Needs	43,130	596.36	18.65
	Low Needs	19,953	606.05	17.25
	Charter	13,719	600.71	19.25
	Religious and Independent	10953	589.51	21.56
SWD	All Codes	23,056	582.85	20.90
SUA	All Codes	11,289	579.06	20.04
ELL	ELL=Y	12,780	580.08	19.11
ELL/SUA	SUA & ELL codes	1089	575.60	19.61
SWD/SUA	SWD & SUA codes	9,407	576.98	19.74
ELL Test Language	Arabic	136	580.82	21.93
	Bengali	42	588.05	19.58
	Chinese (Traditional)	233	602.81	18.28
	Chinese (Simplified)	65	605.95	16.86
	Haitian-Creole	33	574.30	20.46
	Korean	29	605.48	26.06
	Russian	159	588.10	17.93
	Spanish	3,703	580.37	18.34
	All Translations	4,400	582.43	19.53

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 595.48 with a standard deviation of 22.21. 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 (610.28). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—17 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored 16–20 scale score points below the mean scale score for the population. Students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 25 scale score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup

			Scale Score	
Demographic Category		**N-Count**	**Mean**	**SD**
State	All Students	169,535	595.48	22.21
Gender	Female	83,956	595.00	21.63
	Male	85,576	595.95	22.75
	Non-Binary	-	-	-
Ethnicity	Asian	18,723	610.28	20.83
	African American	25,544	588.46	21.98
	Hispanic	46,289	588.97	21.14
	American Indian	1,222	593.51	21.50
	Multiracial	5,552	597.40	22.83
	Pacific Islander	358	598.94	19.63
	White	71,706	598.22	20.75
NRC	New York	52,096	596.02	23.32
	Big 4 Cities	6,341	578.48	22.57
	Urban/Suburban	12,863	585.41	21.09
	Rural	9,377	590.52	19.71
	Average Needs	42,238	596.61	19.42
	Low Needs	19,785	607.69	18.08
	Charter	13,253	599.75	21.39
	Religious and Independent	13582	588.86	22.72
SWD	All Codes	23,687	579.94	21.65
SUA	All Codes	12,838	577.95	20.86
ELL	ELL=Y	11,876	575.12	20.10
ELL/SUA	SUA & ELL codes	1,140	570.59	17.83
SWD/SUA	SWD & SUA codes	10,276	575.20	20.19
ELL Test Language	Arabic	141	573.79	20.08
	Bengali	35	586.97	18.93
	Chinese (Traditional)	260	604.07	18.14
	Chinese (Simplified)	54	605.70	18.39
	Haitian-Creole	37	579.30	19.30
	Korean	28	600.36	27.52

Demographic Category	N-Count	Scale Score Mean	SD
Russian	146	586.80	20.65
Spanish	3,537	573.90	19.17
All Translations	4,238	576.92	21.01

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 595.20 with a standard deviation of 21.65. Male students outperformed female students. 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 (610.68). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—16 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored 15–20 scale score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 24 scale score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-	Scale Score Mean	SD
State	All Students	163,950	595.20	21.65
Gender	Female	80,884	594.40	21.03
	Male	83,056	595.97	22.21
	Non-Binary	-	-	-
Ethnicity	Asian	17,456	610.68	20.16
	African American	25,478	587.35	21.06
	Hispanic	45,001	589.33	20.40
	American Indian	1,136	591.50	20.97
	Multiracial	5,121	596.75	21.81
	Pacific Islander	369	598.15	21.32
	White	69,267	597.94	20.36
NRC	New York	52,141	596.50	22.82
	Big 4 Cities	6,149	579.21	21.45
	Urban/Suburban	12,338	585.14	19.93
	Rural	9,213	589.90	19.37
	Average Needs	41,115	596.33	19.38
	Low Needs	19,331	607.12	17.44
	Charter	12,637	596.69	20.26
	Religious and Independent	11,026	586.81	21.77
SWD	All Codes	23,327	579.84	20.59

	Demographic Category	N-	Scale Score Mean	SD
SUA	All Codes	13,019	578.25	19.91
ELL	ELL=Y	10,328	574.83	18.01
ELL/SUA	SUA & ELL codes	1,107	570.89	16.93
SWD/SUA	SWD & SUA codes	10,464	575.52	19.06
ELL Test Language	Arabic	133	579.18	20.22
	Bengali	35	584.83	22.51
	Chinese (Traditional)	236	603.88	20.00
	Chinese (Simplified)	47	604.60	19.80
	Haitian-Creole	30	575.93	18.08
	Korean	16	604.31	24.74
	Russian	144	590.41	21.44
	Spanish	3,327	575.52	17.92
	All Translations	3,968	578.42	19.97

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 596.99 with a standard deviation of 20.18. 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 (610.77). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—12 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored 16–18 scale score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 23 scale score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup

	Demographic Category	N-	Scale Score Mean	SD
State	All Students	160,087	596.99	20.18
Gender	Female	78,322	596.62	19.81
	Male	81,755	597.34	20.52
	Non-Binary	-	-	-
Ethnicity	Asian	16,929	610.77	19.72
	African American	25,486	589.36	19.22
	Hispanic	44,995	590.86	18.74
	American Indian	1,077	593.22	19.48
	Multiracial	4,835	599.48	21.05
	Pacific Islander	334	598.84	19.47

Demographic Category		N-	Scale Score	
			Mean	**SD**
	White	66,280	600.46	18.80
NRC	New York	50,129	596.04	21.31
	Big 4 Cities	6,212	585.35	19.80
	Urban/Suburban	11,963	587.76	18.67
	Rural	8,750	593.04	18.16
	Average Needs	37,910	599.11	18.07
	Low Needs	18,153	608.85	16.97
	Charter	13,671	598.12	19.59
	Religious and Independent	13299	593.52	19.66
SWD	All Codes	22,071	580.81	18.28
SUA	All Codes	12,325	581.33	18.24
ELL	ELL=Y	9,805	578.76	16.34
ELL/SUA	SUA & ELL codes	1,018	573.76	14.88
SWD/SUA	SWD & SUA codes	9,710	578.27	17.29
ELL Test Language	Arabic	221	578.00	17.62
	Bengali	82	583.74	16.53
	Chinese (Traditional)	436	600.78	18.63
	Chinese (Simplified)	73	606.05	17.47
	Haitian-Creole	81	571.81	15.62
	Korean	27	609.04	20.32
	Russian	166	592.16	17.88
	Spanish	4,127	577.34	15.67
	All Translations	5,213	580.38	17.95

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 597.51 with a standard deviation of 20.07. 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 (612.08). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—15 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored 16–21 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 24 scale score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	SD
State	All Students	154,425	597.51	20.07
Gender	Female	75,398	597.92	19.45
	Male	79,009	597.12	20.65
	Non-Binary	-	-	-
Ethnicity	Asian	17,095	612.08	18.56
	African American	25,652	590.06	19.12
	Hispanic	43,592	592.02	18.71
	American Indian	1,037	594.43	18.87
	Multiracial	4,401	599.09	20.73
	Pacific Islander	355	599.70	20.31
	White	62,199	600.37	18.95
NRC	New York	51,756	598.90	20.82
	Big 4 Cities	6,146	582.43	20.37
	Urban/Suburban	11,380	587.09	18.64
	Rural	8,651	592.38	17.65
	Average Needs	35,181	598.20	17.75
	Low Needs	17,464	608.06	16.77
	Charter	12,701	599.06	18.38
	Religious and Independent	11,146	593.53	21.55
SWD	All Codes	21,288	581.49	18.64
SUA	All Codes	11,647	581.32	18.90
ELL	ELL=Y	7,982	576.67	18.50
ELL/SUA	SUA & ELL codes	940	573.43	16.15
SWD/SUA	SWD & SUA codes	9,113	578.29	18.14
ELL Test Language	Arabic	197	579.93	18.84
	Bengali	64	582.28	15.95
	Chinese (Traditional)	393	605.30	16.89
	Chinese (Simplified)	66	605.70	13.67
	Haitian-Creole	61	582.25	17.86
	Korean	19	601.63	21.53
	Russian	175	595.59	16.25
	Spanish	4,413	577.66	16.98
	All Translations	5,388	580.87	18.89

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 595.72 with a standard deviation of 20.66. Female students tended to outperform male students by around two scale score points. Asian, 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, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (610.34). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—13 scale score points below the population mean. The SWD, SUA, and ELL subgroups scored 13-16 scale score points below the mean scale score for the population. English Language Learners tested under accommodations were the lowest performing subgroup analyzed for English forms, scoring about 21 scale score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	97,284	595.72	20.66
Gender	Female	46,337	596.70	20.27
	Male	50,927	594.83	20.97
	Non-Binary	-	-	-
Ethnicity	Asian	8,117	610.34	21.54
	African American	16,955	589.47	20.38
	Hispanic	29,067	591.64	19.97
	American Indian	653	593.32	19.28
	Multiracial	2,442	595.79	20.57
	Pacific Islander	207	601.47	20.61
	White	39,739	598.42	19.13
NRC	New York	31,192	596.30	21.67
	Big 4 Cities	5,113	582.68	21.86
	Urban/Suburban	7,525	585.30	18.40
	Rural	7,216	593.26	18.01
	Average Needs	22,993	596.77	17.46
	Low Needs	8,480	606.52	18.27
	Charter	6,687	599.60	20.07
	Religious and Independent	8,078	596.11	22.75
SWD	All Codes	16,393	582.07	18.60
SUA	All Codes	9,528	582.66	18.78
ELL	ELL=Y	6,553	579.77	18.87
ELL/SUA	SUA & ELL codes	730	574.46	16.62
SWD/SUA	SWD & SUA codes	7,657	580.39	18.14
ELL Test Language	Arabic	186	583.44	18.91
	Bengali	50	588.88	19.30
	Chinese (Traditional)	356	608.61	20.92
	Chinese (Simplified)	33	610.00	19.17
	Haitian-Creole	67	575.75	17.39
	Korean	12	610.58	14.72

Demographic Category	N-Count	Scale Score	
		Mean	SD
Russian	114	591.61	20.37
Spanish	3,336	579.03	17.57
All Translations	4,154	582.51	20.13

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 Appendix T of the *2018 Standards Review Report* 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[8] 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 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 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, 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	165,209	19.33	34.50	40.22	5.95	46.17
4	168,725	24.77	33.76	24.84	16.62	41.46
5	165,024	29.77	32.60	21.92	15.70	37.62
6	163,509	24.24	18.90	23.96	32.91	56.87
7	159,762	20.67	30.76	29.38	19.19	48.57
8	150,130	18.76	30.89	29.16	21.19	50.35

[8] In 2021–22, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

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 46% of students achieved Level III and Level IV. About 50% 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 (68%) 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 23–38% of students in those same performance categories. Only about 16–19% of the SWD, SUA, and ELL subgroups 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%): Female (50%), Asian (68%), Multiracial (50%), Pacific Islander (47%), White (49%) students, and those enrolled in New York City (50%), Low Needs districts (67%), and Charter schools (52%).

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	165,209	19.33	34.50	40.22	5.95	46.17
Gender	Female	82,152	16.39	33.28	43.13	7.20	50.33
	Male	83,055	22.23	35.71	37.35	4.71	42.06
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,026	8.08	23.78	52.64	15.50	68.14
	African American	24,816	25.04	36.78	34.30	3.88	38.18
	Hispanic	44,664	24.62	38.58	33.44	3.36	36.80
	American Indian	1,188	19.70	36.36	38.30	5.64	43.94
	Multiracial	5,830	16.59	33.76	41.41	8.25	49.66
	Pacific Islander	370	16.49	36.49	38.65	8.38	47.03
	White	71,199	16.92	33.72	43.54	5.82	49.36
NRC	New York	49,203	18.01	31.62	40.94	9.43	50.37
	Big 4 Cities	6,399	42.80	33.93	20.97	2.30	23.27
	Urban/Suburban	12,353	30.69	41.00	26.33	1.98	28.31
	Rural	9,252	25.70	44.13	28.86	1.31	30.17
	Average Needs	42,995	16.05	40.08	40.54	3.33	43.87
	Low Needs	19,829	5.97	26.56	57.46	10.02	67.48
	Charter	14,456	15.10	33.28	45.90	5.71	51.61
	Religious and Independent	10,722	36.34	26.25	33.43	3.98	37.41
SWD	All Codes	23,077	40.08	40.73	18.13	1.05	19.18
SUA	All Codes	12,081	41.45	42.94	15.14	0.48	15.62
ELL	ELL=Y	21,396	45.53	35.82	17.79	0.86	18.65
ELL/SUA	SUA & ELL codes	1,187	47.94	40.69	11.20	0.17	11.37
SWD/SUA	SWD & SUA codes	10,244	44.50	42.53	12.65	0.31	12.96

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 41% of students achieved Level III and Level IV. About 46% of Female students were at Level III or above, as compared to 37% 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 (65%) students and students from Low Needs districts (62%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 19–32% of students in those same performance categories. Only about 11–13% of the SWD, SUA, and ELL 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 (41%): Female (46%), Asian (65%), Multiracial (45%), Pacific Islander (45%), and White (46%) students as well as those enrolled in New York City (45%), Low Needs districts (62%), and Charter schools (45%).

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	168,725	24.77	33.76	24.84	16.62	41.46
Gender	Female	83,974	20.81	32.76	26.72	19.71	46.43
	Male	84,749	28.70	34.75	22.99	13.56	36.55
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	18,478	10.92	24.28	29.42	35.38	64.80
	African American	25,991	32.54	35.52	21.18	10.76	31.94
	Hispanic	45,363	32.05	37.17	20.91	9.86	30.77
	American Indian	1,225	26.45	34.61	25.14	13.80	38.94
	Multiracial	5,581	22.83	32.65	24.17	20.35	44.52
	Pacific Islander	354	19.21	35.88	25.42	19.49	44.91
	White	71,621	21.03	33.47	27.54	17.96	45.50
NRC	New York	51,191	24.00	31.17	23.41	21.42	44.83
	Big 4 Cities	6,271	51.44	29.88	12.61	6.06	18.67
	Urban/Suburban	12,618	39.38	36.45	17.29	6.88	24.17
	Rural	9,405	30.77	40.05	20.75	8.42	29.17
	Average Needs	42,363	21.48	38.3	26.66	13.55	40.21
	Low Needs	19,652	8.68	29.14	34.97	27.21	62.18
	Charter	13,884	19.44	35.64	28.04	16.88	44.92
	Religious and Independent	13,341	36.87	28.97	22.10	12.05	34.15
SWD	All Codes	23,972	53.83	32.75	9.90	3.52	13.42
SUA	All Codes	13,695	54.30	34.02	9.38	2.31	11.69
ELL	ELL=Y	19,974	58.59	30.59	8.96	1.86	10.82
ELL/SUA	SUA & ELL codes	1,227	68.62	27.06	4.07	0.24	4.31
SWD/SUA	SWD & SUA codes	11,086	59.69	32.07	6.75	1.50	8.25

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 38% of students achieved Level III and Level IV. About 42% 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 (61%) students and students from Low Needs districts (58%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 17–28% of students in those same performance categories. Only about 6–11% of the SWD, SUA, and ELL 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 (38%): Female (42%), Asian (61%), Multiracial (41%), Pacific Islander (44%), and White (42%) students, as well as those enrolled in New York City (40%), Low Needs districts (58%), and Charter schools (40%).

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	165,024	29.77	32.60	21.92	15.70	37.62
Gender	Female	81,758	25.07	32.93	23.47	18.53	42.00
	Male	83,256	34.39	32.28	20.40	12.92	33.32
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,331	12.90	25.82	27.19	34.09	61.28
	African American	26,182	38.90	33.75	17.49	9.87	27.36
	Hispanic	44,743	37.00	35.29	17.98	9.74	27.72
	American Indian	1,160	35.17	32.16	20.86	11.81	32.67
	Multiracial	5,253	27.55	31.43	22.31	18.71	41.02
	Pacific Islander	371	22.37	33.15	25.88	18.60	44.48
	White	69,864	26.00	32.23	24.78	16.99	41.77
NRC	New York	51,772	28.45	31.36	20.71	19.48	40.19
	Big 4 Cities	6,166	57.95	25.14	11.60	5.32	16.92
	Urban/Suburban	12,359	44.47	33.86	15.04	6.63	21.67
	Rural	9,373	37.86	34.76	19.29	8.09	27.38
	Average Needs	42,022	27.15	36.12	23.86	12.86	36.72
	Low Needs	19,535	11.53	30.60	30.98	26.90	57.88
	Charter	13,284	24.28	35.43	24.22	16.06	40.28
	Religious and Independent	10,513	46.64	25.75	16.90	10.71	27.61
SWD	All Codes	24,153	62.34	26.96	7.91	2.79	10.70
SUA	All Codes	14,158	63.07	27.35	7.23	2.34	9.57
ELL	ELL=Y	17,399	70.41	23.99	4.65	0.95	5.60
ELL/SUA	SUA & ELL codes	1,225	79.67	18.61	1.47	0.24	1.71
SWD/SUA	SWD & SUA codes	11,392	68.43	24.92	5.33	1.32	6.65

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 57% of students achieved Level III and Level IV. About 62% of Female students were at Level III or above, as compared to 51% 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 (79%) students and students from Low Needs districts (76%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 36–48% of students in those same performance categories. Only about 12–22% of the SWD, SUA, and ELL 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 (57%): Female (62%), Asian (79%), Multiracial (61%), Pacific Islander (63%), and White (61%) students, as well as those from New York City (58%), Average (59%) and Low (76%) Needs districts, and Charter (61%) 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	163,509	24.24	18.90	23.96	32.91	56.87
Gender	Female	80,555	19.68	17.88	24.16	38.28	62.44
	Male	82,942	28.67	19.88	23.75	27.69	51.44
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	16,943	9.67	11.69	22.25	56.39	78.64
	African American	26,412	31.30	21.72	22.36	24.62	46.98
	Hispanic	45,881	30.06	22.15	23.33	24.46	47.79
	American Indian	1,103	26.29	21.40	22.57	29.74	52.31
	Multiracial	4,987	22.22	16.90	23.52	37.36	60.88
	Pacific Islander	330	17.58	19.70	24.24	38.48	62.72
	White	67,734	21.30	17.49	25.48	35.72	61.20
NRC	New York	50,630	23.40	18.47	21.61	36.52	58.13
	Big 4 Cities	6,280	44.12	19.89	19.33	16.66	35.99
	Urban/Suburban	12,251	37.48	22.82	21.44	18.25	39.69
	Rural	9,003	29.56	22.90	24.91	22.63	47.54
	Average Needs	39,528	21.52	19.95	27.24	31.29	58.53
	Low Needs	18,534	9.54	14.64	26.89	48.93	75.82
	Charter	14,216	19.03	19.78	25.82	35.36	61.18
	Religious and Independent	13,067	36.53	15.56	20.85	27.05	47.90
SWD	All Codes	23,457	56.25	22.27	13.69	7.79	21.48
SUA	All Codes	13,773	56.04	22.37	14.30	7.28	21.58
ELL	ELL=Y	16,419	68.26	19.84	9.20	2.71	11.91
ELL/SUA	SUA & ELL codes	1,221	77.15	17.12	4.83	0.90	5.73
SWD/SUA	SWD & SUA codes	10,759	62.30	21.30	11.64	4.76	16.40

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 49% of students achieved Level III and Level IV. About 54% of Female students were at Level III or above, as compared to 43% 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 (74%) students and students from Low Needs (65%) districts. The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 25–40% of students in those same performance categories. Only about 6–15% of the SWD, SUA, and ELL 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 (49%): Female (54%), Asian (74%), Multiracial (51%), Pacific Islander (57%), and White (51%) students, as well as those enrolled in New York City (54%), Low Needs (65%) districts, and Charter schools (54%).

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	159,762	20.67	30.76	29.38	19.19	48.57
Gender	Female	78,641	16.32	29.31	31.01	23.36	54.37
	Male	81,106	24.89	32.17	27.80	15.14	42.94
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,349	7.45	18.13	32.29	42.13	74.42
	African American	26,934	26.17	34.51	26.72	12.59	39.31
	Hispanic	44,944	25.32	34.89	26.83	12.96	39.79
	American Indian	1,077	22.19	33.61	27.76	16.43	44.19
	Multiracial	4,648	19.88	29.20	28.7	22.22	50.92
	Pacific Islander	350	14.00	28.86	30.86	26.29	57.15
	White	64,349	18.72	29.79	31.57	19.92	51.49
NRC	New York	52,902	17.01	28.58	28.72	25.68	54.40
	Big 4 Cities	6,256	44.01	31.12	18.41	6.46	24.87
	Urban/Suburban	11,963	36.15	34.96	21.17	7.72	28.89
	Rural	9,002	28.23	36.47	25.84	9.46	35.30
	Average Needs	37,348	20.70	34.11	30.56	14.63	45.19
	Low Needs	18,167	8.68	26.55	36.81	27.95	64.76
	Charter	13,221	14.39	31.34	34.80	19.47	54.27
	Religious and Independent	10,903	29.26	26.66	27.79	16.29	44.08
SWD	All Codes	22,940	50.17	34.44	12.37	3.02	15.39
SUA	All Codes	13,673	51.88	33.28	12.05	2.78	14.83
ELL	ELL=Y	13,743	66.47	27.61	5.23	0.69	5.92
ELL/SUA	SUA & ELL codes	1,265	74.23	24.11	1.50	0.16	1.66
SWD/SUA	SWD & SUA codes	10,809	57.86	31.26	9.24	1.64	10.88

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 50% of students achieved Level III and Level IV. About 58% of Female students were at Level III or above, as compared to 43% 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 (74%) students and students from Low Needs (68%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 28–42% of students in those same performance categories. Only about 6–16% of the SWD, SUA, and ELL 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%): Female (58%), Asian (74%), Multiracial (52%), Pacific Islander (57%), and White (53%) students, as well as those enrolled in New York City (55%) and Low Needs (68%) districts, and Charter schools (58%).

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	150,130	18.76	30.89	29.16	21.19	50.35
Gender	Female	73,004	13.49	28.67	31.84	26.00	57.84
	Male	77,107	23.75	32.99	26.63	16.64	43.27
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,134	7.43	18.26	30.82	43.49	74.31
	African American	26,171	22.17	35.49	27.71	14.63	42.34
	Hispanic	42,312	22.78	34.75	27.70	14.77	42.47
	American Indian	1,075	19.81	35.26	28.00	16.93	44.93
	Multiracial	3,989	18.95	28.58	28.08	24.39	52.47
	Pacific Islander	368	13.04	30.43	30.16	26.36	56.52
	White	58,989	17.60	29.84	30.47	22.09	52.56
NRC	New York	52,287	15.12	29.77	29.40	25.71	55.11
	Big 4 Cities	6,249	38.93	32.82	19.94	8.31	28.25
	Urban/Suburban	11,375	32.20	36.19	21.78	9.82	31.60
	Rural	8,964	25.35	36.40	26.39	11.86	38.25
	Average Needs	34,215	19.21	33.59	30.06	17.14	47.20
	Low Needs	16,452	8.30	23.88	34.45	33.38	67.83
	Charter	12,373	11.00	30.67	35.17	23.16	58.33
	Religious and Independent	8,215	31.47	26.31	24.53	17.70	42.23
SWD	All Codes	21,673	45.71	38.02	12.99	3.28	16.27
SUA	All Codes	12,928	47.18	36.36	13.01	3.46	16.47
ELL	ELL=Y	12,581	62.65	31.49	5.19	0.67	5.86
ELL/SUA	SUA & ELL codes	1,029	70.17	26.63	3.01	0.19	3.20
SWD/SUA	SWD & SUA codes	10,284	52.61	35.72	9.80	1.88	11.68

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[9] 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 students outperformed Female students in terms of Level III and above classifications except for grade eight. 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, Urban/Suburban, and Rural), and Religious and Independent schools. The Level III and above rates for SWD and SUA subgroups were low, compared to the total population of examinees. The subgroups that used the Korean or Chinese translations outperformed other test translation subgroups. The n-counts for some translation subgroups were 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

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
3	166,446	28.46	24.06	29.03	18.45	47.48
4	169,535	32.32	25.12	21.83	20.73	42.56
5	163,950	40.55	22.97	18.59	17.89	36.48
6	160,087	36.71	24.74	20.01	18.55	38.56
7	154,425	35.73	28.58	20.31	15.38	35.69
8	97,284	46.57	28.55	15.27	9.62	24.89

---

[9] In 2021–22, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–22 reporting year, school district access to the code was significantly limited and, thus, the 2021–22 technical report does not disaggregate data by nonbinary.

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 47% of students achieved Level III and Level IV. About 46% of Female and 49% of 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 (73%) students and students from Low Needs districts (69%). 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 16–23% of the SWD, SUA, and ELL subgroups 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%): Asian (73%), Multiracial (56%), Pacific Islander (58%), and White (60%) students, as well as those enrolled at Low Needs districts (69%) and Charter schools (57%). For ELLs who used translated test forms, the percentages of students earning at least a Level III ranged from 9% (Haitian-Creole) to 72% (Simplified Chinese).

Table 8.25. Mathematics 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	166,446	28.46	24.06	29.03	18.45	47.48
Gender	Female	82,532	28.77	25.13	29.08	17.02	46.10
	Male	83,912	28.16	23.00	28.99	19.85	48.84
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,324	11.19	15.44	32.88	40.49	73.37
	African American	24,465	38.16	24.78	24.39	12.66	37.05
	Hispanic	45,590	37.50	26.90	24.88	10.72	35.60
	American Indian	1,182	29.44	23.43	30.12	17.01	47.13
	Multiracial	5,811	27.16	22.77	28.05	22.03	50.08
	Pacific Islander	378	24.34	22.75	30.95	21.96	52.91
	White	71,556	23.63	24.22	32.41	19.75	52.16
NRC	New York	50,278	28.00	22.39	28.10	21.52	49.62
	Big 4 Cities	6,490	55.78	21.57	15.76	6.89	22.65
	Urban/Suburban	12,680	43.86	26.55	21.80	7.80	29.60
	Rural	9,243	37.08	28.46	25.40	9.06	34.46
	Average Needs	43,130	26.20	27.39	30.67	15.74	46.41
	Low Needs	19,953	10.80	19.83	38.16	31.21	69.37
	Charter	13,719	20.88	21.87	32.70	24.56	57.26
	Religious and Independent	10,953	39.91	23.91	24.93	11.25	36.18
SWD	All Codes	23,056	54.48	22.82	16.42	6.29	22.71
SUA	All Codes	11,289	61.97	21.92	12.51	3.60	16.11
ELL	ELL=Y	12,780	59.92	23.59	13.45	3.04	16.49
ELL/SU	SUA & ELL codes	1,089	68.87	19.74	8.91	2.48	11.39

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
SWD/S	SWD & SUA codes	9,407	66.26	20.45	10.52	2.76	13.28
ELL Test Language	Arabic	136	56.62	21.32	16.91	5.15	22.06
	Bengali	42	40.48	23.81	30.95	4.76	35.71
	Chinese (Traditional)	233	15.45	16.31	42.49	25.75	68.24
	Chinese (Simplified)	65	13.85	13.85	38.46	33.85	72.31
	Haitian-Creole	33	63.64	27.27	6.06	3.03	9.09
	Korean	29	17.24	13.79	34.48	34.48	68.96
	Russian	159	43.40	23.90	26.42	6.29	32.71
	Spanish	3,703	60.28	24.41	12.15	3.16	15.31
	All Translations	4,400	56.05	23.66	15.09	5.20	20.29

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 43% of students achieved Level III and Level IV. About 40% of Female students and 45% of 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 (71%) students and students from Low Needs (68%). 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 10–17% of the SWD, SUA, and ELL subgroups 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%): Asian (71%), Multiracial (47%), Pacific Islander (46%), and White (48%) students, as well as students enrolled in Average (44%) and Low (68%) Needs districts and Charter schools (49%). For ELLs who used translated test forms, the percentages of students earning at least a Level III ranged from 9% (Arabic) to 63% (Simplified Chinese).

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	169,535	32.32	25.12	21.83	20.73	42.56
Gender	Female	83,956	33.20	26.31	21.34	19.15	40.49
	Male	85,576	31.45	23.97	22.31	22.28	44.59
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	18,723	12.01	17.24	24.30	46.45	70.75
	African American	25,544	45.22	25.64	16.64	12.50	29.14
	Hispanic	46,289	43.42	27.25	17.60	11.73	29.33
	American Indian	1,222	36.91	25.45	20.95	16.69	37.64
	Multiracial	5,552	30.51	22.89	21.81	24.78	46.59
	Pacific Islander	358	25.98	28.49	22.91	22.63	45.54

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
	White	71,706	25.88	25.79	25.79	22.54	48.33
NRC	New York	52,096	33.20	23.72	19.63	23.44	43.07
	Big 4 Cities	6,341	64.82	18.45	10.12	6.61	16.73
	Urban/Suburban	12,863	50.31	26.22	15.21	8.26	23.47
	Rural	9,377	40.63	28.73	19.78	10.86	30.64
	Average Needs	42,238	28.04	28.37	25.33	18.27	43.60
	Low Needs	19,785	11.27	20.49	30.28	37.96	68.24
	Charter	13,253	25.51	25.41	23.18	25.90	49.08
	Religious and Independent	13,582	41.58	26.50	18.86	13.06	31.92
SWD	All Codes	23,687	61.96	20.71	11.08	6.25	17.33
SUA	All Codes	12,838	65.71	20.28	9.81	4.20	14.01
ELL	ELL=Y	11,876	70.50	19.53	7.16	2.82	9.98
ELL/SU	SUA & ELL codes	1,140	81.67	13.68	3.33	1.32	4.65
SWD/S	SWD & SUA codes	10,276	71.38	17.79	7.93	2.90	10.83
ELL Test Language	Arabic	141	70.21	21.28	6.38	2.13	8.51
	Bengali	35	45.71	25.71	20.00	8.57	28.57
	Chinese (Traditional)	260	15.00	23.85	30.77	30.38	61.15
	Chinese (Simplified)	54	14.81	22.22	27.78	35.19	62.97
	Haitian-Creole	37	75.68	8.11	13.51	2.70	16.21
	Korean	28	32.14	17.86	14.29	35.71	50.00
	Russian	146	45.21	30.82	16.44	7.53	23.97
	Spanish	3,537	73.65	18.38	6.05	1.92	7.97
	All Translations	4,238	67.72	19.25	8.45	4.58	13.03

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 36% of students achieved Level III and Level IV. About 34% of Female students and 39% of 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 (67%) students and students from Low Needs districts (61%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 14–24% of students in those same performance categories. Only about 6–13% of the SWD, SUA, and ELL subgroups earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (36%): Asian (67%), Pacific Islander (41%), and White (42%) students, as well as those enrolled in Average (37%) and Low (61%) Needs districts and Charter schools (37%). For ELLs who used translated test forms, the percentages of students earning at least a Level III ranged from 6% (Spanish) to 57% (Simplified 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	163,950	40.55	22.97	18.59	17.89	36.48
Gender	Female	80,884	41.90	24.14	18.12	15.84	33.96
	Male	83,056	39.25	21.82	19.04	19.89	38.93
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,456	15.39	17.56	22.44	44.61	67.05
	African American	25,478	56.05	21.60	13.42	8.93	22.35
	Hispanic	45,001	51.88	23.74	14.77	9.60	24.37
	American Indian	1,136	50.26	21.57	14.70	13.47	28.17
	Multiracial	5,121	39.56	21.54	18.53	20.37	38.90
	Pacific Islander	369	31.71	27.10	22.22	18.97	41.19
	White	69,267	33.76	24.45	22.05	19.74	41.79
NRC	New York	52,141	39.84	21.37	17.29	21.51	38.80
	Big 4 Cities	6,149	71.59	14.36	8.34	5.71	14.05
	Urban/Suburban	12,338	60.65	21.27	11.99	6.09	18.08
	Rural	9,213	50.05	25.12	16.21	8.63	24.84
	Average Needs	41,115	36.87	25.69	21.64	15.79	37.43
	Low Needs	19,331	16.12	22.99	27.80	33.08	60.88
	Charter	12,637	37.15	26.26	18.55	18.03	36.58
	Religious and Independent	11,026	56.68	21.45	12.34	9.53	21.87
SWD	All Codes	23,327	71.03	16.27	7.98	4.71	12.69
SUA	All Codes	13,019	73.68	15.49	7.40	3.43	10.83
ELL	ELL=Y	10,328	81.43	12.96	4.33	1.28	5.61
ELL/SUA	SUA & ELL codes	1,107	87.90	9.03	2.53	0.54	3.07
SWD/SUA	SWD & SUA codes	10,464	79.06	13.34	5.46	2.14	7.60
ELL Test Language	Arabic	133	71.43	18.80	6.02	3.76	9.78
	Bengali	35	57.14	20.00	20.00	2.86	22.86
	Chinese (Traditional)	236	22.88	23.73	24.58	28.81	53.39
	Chinese (Simplified)	47	25.53	17.02	23.4	34.04	57.44
	Haitian-Creole	30	76.67	16.67	6.67	-	6.67
	Korean	16	31.25	12.50	25.00	31.25	56.25
	Russian	144	47.22	28.47	11.11	13.19	24.30
	Spanish	3,327	80.40	13.89	4.57	1.14	5.71
	All Translations	3,968	74.40	15.27	6.50	3.83	10.33

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 39% of students achieved Level III and Level

IV. About 38% of Female students and 40% of 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 (67%) students and students from Low Needs districts (66%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 18–25% of students in those same performance categories. Only about 6–11% of the SWD, SUA, and ELL subgroups earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (39%): Asian (67%), Multiracial (44%), Pacific Islander (40%), and White (46%) students, as well as those enrolled in Average (43%) and Low (66%) Needs districts and Charter schools (40%). For ELLs who used translated test forms, the percentages of students earning at least a Level III ranged from 2% (Haitian-Creole) to 59% (Simplified Chinese).

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	160,087	36.71	24.74	20.01	18.55	38.56
Gender	Female	78,322	37.00	25.45	20.29	17.27	37.56
	Male	81,755	36.42	24.06	19.73	19.78	39.51
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	16,929	14.76	18.50	22.97	43.77	66.74
	African American	25,486	52.69	24.30	13.86	9.16	23.02
	Hispanic	44,995	49.26	25.99	15.20	9.55	24.75
	American Indian	1,077	44.38	25.44	17.92	12.26	30.18
	Multiracial	4,835	32.84	22.71	21.05	23.39	44.44
	Pacific Islander	334	30.84	29.34	18.56	21.26	39.82
	White	66,280	27.77	25.78	24.86	21.59	46.45
NRC	New York	50,129	40.91	23.53	16.40	19.16	35.56
	Big 4 Cities	6,212	61.93	19.77	11.12	7.18	18.30
	Urban/Suburban	11,963	55.41	24.39	13.62	6.58	20.20
	Rural	8,750	42.42	28.27	19.17	10.14	29.31
	Average Needs	37,910	29.99	27.22	24.72	18.07	42.79
	Low Needs	18,153	13.03	21.07	29.52	36.39	65.91
	Charter	13,671	34.01	25.99	20.28	19.73	40.01
	Religious and Independent	13,299	42.70	26.31	17.35	13.64	30.99
SWD	All Codes	22,071	72.48	16.72	7.17	3.64	10.81
SUA	All Codes	12,325	70.54	17.65	8.49	3.32	11.81
ELL	ELL=Y	9,805	77.22	17.01	4.35	1.42	5.77
ELL/SU	SUA & ELL codes	1,018	88.31	10.12	1.08	0.49	1.57
SWD/SU	SWD & SUA codes	9,710	77.72	14.62	5.65	2.00	7.65
ELL Test Language	Arabic	221	77.38	16.29	5.43	0.90	6.33
	Bengali	82	64.63	25.61	7.32	2.44	9.76

Demographic Category	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
Chinese (Traditional)	436	26.38	29.36	22.48	21.79	44.27
Chinese (Simplified)	73	20.55	20.55	23.29	35.62	58.91
Haitian-Creole	81	88.89	8.64	2.47	-	2.47
Korean	27	22.22	18.52	18.52	40.74	59.26
Russian	166	43.37	30.12	15.66	10.84	26.50
Spanish	4,127	81.61	14.32	3.46	0.61	4.07
All Translations	5,213	74.28	16.36	5.93	3.43	9.36

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 36% of students achieved Level III and Level IV. About 35% of Female students were at Level III or above, as compared to 36% 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 (67%) students and students from Low Needs districts (60%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 13–23% of students in those same performance categories. Only about 4–9% of the SWD, SUA, and ELL subgroups earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (36%): Asian (67%), Multiracial (40%), Pacific Islander (39%), and White (42%) students, as well as those enrolled in New York City (38%), Average (36%) and Low (60%) Needs districts and Charter schools (38%). For ELLs who used translated test forms, the percentages of students earning at least a Level III ranged from 3% (Spanish) to 56% (Simplified Chinese).

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	154,425	35.73	28.58	20.31	15.38	35.69
Gender	Female	75,398	35.13	29.74	19.93	15.21	35.14
	Male	79,009	36.30	27.48	20.67	15.55	36.22
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	17,095	12.78	20.22	25.54	41.47	67.01
	African American	25,652	51.33	28.06	13.84	6.78	20.62
	Hispanic	43,592	46.60	30.21	15.56	7.64	23.20
	American Indian	1,037	41.95	30.47	17.26	10.32	27.58
	Multiracial	4,401	33.45	26.68	21.56	18.31	39.87
	Pacific Islander	355	33.80	27.04	19.44	19.72	39.16
	White	62,199	28.01	30.08	24.85	17.06	41.91
NRC	New York	51,756	35.46	26.99	18.26	19.28	37.54
	Big 4 Cities	6,146	67.59	19.23	9.49	3.69	13.18

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
	Urban/Suburban	11,380	57.35	27.52	11.45	3.68	15.13
	Rural	8,651	44.64	32.76	16.52	6.08	22.60
	Average Needs	35,181	32.09	32.02	23.77	12.12	35.89
	Low Needs	17,464	14.19	25.88	30.74	29.19	59.93
	Charter	12,701	31.68	30.49	22.06	15.76	37.82
	Religious and Independent	11,146	40.21	30.15	18.51	11.13	29.64
SWD	All Codes	21,288	71.26	19.72	6.40	2.62	9.02
SUA	All Codes	11,647	70.65	20.13	6.77	2.45	9.22
ELL	ELL=Y	7,982	77.90	17.88	3.57	0.65	4.22
ELL/SU	SUA & ELL codes	940	87.34	11.06	1.60	-	1.60
SWD/SU	SWD & SUA codes	9,113	77.25	16.93	4.53	1.28	5.81
ELL Test Language	Arabic	197	69.54	22.84	6.09	1.52	7.61
	Bengali	64	68.75	25.00	4.69	1.56	6.25
	Chinese (Traditional)	393	18.83	27.48	31.30	22.39	53.69
	Chinese (Simplified)	66	12.12	31.82	39.39	16.67	56.06
	Haitian-Creole	61	70.49	19.67	8.20	1.64	9.84
	Korean	19	26.32	31.58	10.53	31.58	42.11
	Russian	175	40.00	36.00	14.29	9.71	24.00
	Spanish	4,413	79.47	17.09	2.97	0.48	3.45
	All Translations	5,388	72.16	19.02	6.07	2.75	8.82

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 25% of students achieved Level III and Level IV. About 26% of Female students were at Level III or above, as compared to 24% 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 (54%) students and students from Low Needs districts (45%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 8–18% of students in those same performance categories. Only about 6–7% of the SWD, SUA, and ELL subgroups earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (25%): Asian (54%), Pacific Islander (40%), and White (28%) students, as well as those enrolled in New York City (27%) and Low Needs districts (45%), and Charter (31%) and Religious and Independent (28%) schools. For ELLs who used translated test forms, the percentages of students earning at least a Level III ranged from 1% (Haitian-Creole) to 58% (Korean).

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	97,284	46.57	28.55	15.27	9.62	24.89
Gender	Female	46,337	44.47	29.33	16.37	9.83	26.20
	Male	50,927	48.49	27.84	14.26	9.41	23.67
	Non-Binary	-	-	-	-	-	-
Ethnicity	Asian	8,117	21.83	24.13	22.22	31.81	54.03
	African American	16,955	59.35	25.08	10.21	5.36	15.57
	Hispanic	29,067	55.75	26.29	11.92	6.04	17.96
	American Indian	653	53.29	27.41	11.79	7.50	19.29
	Multiracial	2,442	47.34	28.17	15.52	8.97	24.49
	Pacific Islander	207	34.30	26.09	24.15	15.46	39.61
	White	39,739	39.29	32.66	18.48	9.57	28.05
NRC	New York	31,192	47.33	26.08	14.20	12.39	26.59
	Big 4 Cities	5,113	72.89	14.98	7.28	4.85	12.13
	Urban/Suburban	7,525	69.13	22.68	6.68	1.50	8.18
	Rural	7,216	51.21	31.44	13.10	4.25	17.35
	Average Needs	22,993	41.84	35.41	17.34	5.41	22.75
	Low Needs	8,480	23.02	31.70	25.83	19.46	45.29
	Charter	6,687	39.52	29.21	17.63	13.64	31.27
	Religious and Independent	8,078	45.85	26.13	15.42	12.59	28.01
SWD	All Codes	16,393	75.78	18.04	4.50	1.68	6.18
SUA	All Codes	9,528	74.38	18.73	5.17	1.71	6.88
ELL	ELL=Y	6,553	78.86	15.44	4.49	1.21	5.70
ELL/SU	SUA & ELL codes	730	89.59	9.18	0.96	0.27	1.23
SWD/SU	SWD & SUA codes	7,657	79.12	16.22	3.58	1.08	4.66
ELL Test Language	Arabic	186	71.51	22.58	4.30	1.61	5.91
	Bengali	50	48.00	40.00	12.00	-	12.00
	Chinese (Traditional)	356	23.88	23.03	24.16	28.93	53.09
	Chinese (Simplified)	33	15.15	36.36	24.24	24.24	48.48
	Haitian-Creole	67	83.58	14.93	1.49	-	1.49
	Korean	12	8.33	33.33	33.33	25.00	58.33
	Russian	114	57.89	22.81	12.28	7.02	19.30
	Spanish	3,336	82.88	13.13	3.09	0.90	3.99
	All Translations	4,154	75.47	15.26	5.54	3.73	9.27

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.

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. (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.

Appendix A: ELA and Mathematics Test Configurations

Table A1. ELA Test Configuration

| Grade | Day | Session | Number of Items | | | | Total |
| | | | Multiple-Choice | | Constructed-Response | | |
			Operational	Embedded	Operational	Embedded	
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

| Grade | Day | Session | Number of Items | | | | Total |
| | | | Multiple-Choice | | Constructed-Response | | |
			Operational	Embedded	Operational	Embedded	
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: *2022 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	37.50
	2	2	51
		Total	**88.50**
4	1	1	45
	2	2	51
		Total	**96**
5	1	1	45
	2	2	51
		Total	**96**
6	1	1	46.50
	2	2	54.50
		Total	**101**
7	1	1	49.50
	2	2	54.50
		Total	**104**

			on Task (min.)
	1	1	49.50
8	2	2	54.50
		Total	**104**

Source: *2022 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 *2022 Teacher's Directions* and the *School Administrator's Manual* at http://www.nysed.gov/state-assessment/archive-2022-grades-3-8-test-manuals-and-teachers-directions.

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Strand	Point Range Target	Point Range Actual	% of Test Target	% of Test 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	21	45%–55%	48%
		Informational Text	20–24	23	45%–55%	52%
6	44	Literature	20–24	20	45%–55%	45%
		Informational Text	20–24	24	45%–55%	55%
7	46	Literature	20–26	26	43%–57%	57%
		Informational Text	20–26	20	43%–57%	43%
8	46	Literature	20–26	26	43%–57%	57%
		Informational Text	20–26	20	43%–57%	43%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Domain	Point Range Target	Point Range Actual	% of Test Target	% of Test 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	7	14%–24%	17%
		Measurement and Data	9–13	10	21%–31%	24%
		Geometry*	1–2	3	2%–5%	7%
4	46	Operations and Algebraic Thinking	7–11	10	15%–26%	22%
		Number and Operations in Base Ten	10–14	11	22%–30%	24%
		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%
		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	13	22%–30%	28%
		Geometry*	1–2	2	2%–4%	4%
5	48	Ratios and Proportional Relationships	10–14	12	21%–29%	25%
		The Number System	9–13	9	19%–27%	19%
6	48	Expressions and Equations	16–22	19	33%–46%	40%
		Geometry	5–9	8	10%–19%	17%

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

*There is a slight difference between the "Target % of Test" shown in these tables and the tables presented in the Guides to the 2022 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.

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 2022 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 2022 item selection included 2017–2019 embedded and stand-alone field-test items.[10]

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 flagged DIF items that do not exhibit bias.
- Provide 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 2019 test)
    - TCC, Test Information Curves, and Conditional SEM Curves for each subject and grade, again using the Spring 2019 operational test as a reference

---

[10] Due to the Covid-19 pandemic, there was no testing in 2020 and no field testing in 2021. Therefore, items eligible for the 2022 selection included 2017-2019 field-tested items.

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2022 NYSTP Grades 3–8 ELA and Mathematics Tests. Field test items that do not contribute to students' scores have been omitted. Additional details on the standards to which these items align may be found at: http://www.nysed.gov/next-generation-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.RL.3.6	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.4	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.3.5	Reading Standards for Literature	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.L.3.4	Language Standards	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.2	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.5	Reading Standards for Informational Text	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.3	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.3.2	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.3.6	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.3.6	Reading Standards for Literature	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.3.5	Reading Standards for Literature	Reading
			Session 2		
25	Constructed Response	2	CCSS.ELA-Literacy.RI.3.8	Reading Standards for Informational Text	Writing to Sources
26	Constructed Response	2	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Writing to Sources
27	Constructed Response	2	CCSS.ELA-Literacy.RI.3.7	Reading Standards for Informational Text	Writing to Sources

Question	Type	Points	Standard	Strand	Subscore
28	Constructed Response	2	CCSS.ELA-Literacy.RI.3.2	Reading Standards for Informational Text	Writing to Sources
29	Constructed Response	2	CCSS.ELA-Literacy.RI.3.3	Reading Standards for Informational Text	Writing to Sources
30	Constructed Response	2	CCSS.ELA-Literacy.RL.3.3	Reading Standards for Literature	Writing to Sources
31	Constructed Response	4	CCSS.ELA-Literacy.RL.3.2	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 2022 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.RI.4.8	Reading Standards for Informational Text	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.2	Reading Standards for Informational Text	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.L.4.4	Language Standards	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.3	Reading Standards for Informational Text	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.7	Reading Standards for Informational Text	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.2	Reading Standards for Informational Text	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.6	Reading Standards for Literature	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.4	Reading Standards for Literature	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.7	Reading Standards for Literature	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RL.4.2	Reading Standards for Literature	Reading
19	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.3	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.8	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.4	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.5	Reading Standards for Informational Text	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.3	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
24	Multiple Choice	1	CCSS.ELA-Literacy.RI.4.2	Reading Standards for Informational Text	Reading
Session 2					
25	Constructed Response	2	CCSS.ELA-Literacy.RI.4.3	Reading Standards for Informational Text	Writing to Sources
26	Constructed Response	2	CCSS.ELA-Literacy.RI.4.2	Reading Standards for Informational Text	Writing to Sources
27	Constructed Response	2	CCSS.ELA-Literacy.RI.4.4	Reading Standards for Informational Text	Writing to Sources
28	Constructed Response	2	CCSS.ELA-Literacy.RL.4.3	Reading Standards for Literature	Writing to Sources
29	Constructed Response	2	CCSS.ELA-Literacy.RL.4.6	Reading Standards for Literature	Writing to Sources
30	Constructed Response	2	CCSS.ELA-Literacy.RL.4.2	Reading Standards for Literature	Writing to Sources
31	Constructed Response	4	CCSS.ELA-Literacy.RL.4.3	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 2022 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.5.2	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	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.5.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.5	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.2	Reading Standards for Literature	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.6	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.3	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.2	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.4	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.2	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.5	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.2	Reading Standards for Literature	Reading
25	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
26	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.4	Reading Standards for Literature	Reading
27	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Reading
28	Multiple Choice	1	CCSS.ELA-Literacy.RL.5.6	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.RI.5.4	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.L.5.4	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.5.7	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.5.2	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RL.5.5	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RL.5.3	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RL.5.2	Reading Standards for Literature	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RI.5.6	Reading Standards for Informational Text	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RI.5.2	Reading Standards for Informational Text	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RI.5.5	Reading Standards for Informational Text	Writing to Sources
42	Constructed Response	4	CCSS.ELA-Literacy.RI.5.8	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 2022 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.6.5	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.4	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.L.6.4	Language Standards	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.2	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.6.4	Language Standards	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.2	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.6.4	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.5	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.2	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.L.6.4	Language Standards	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
25	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.4	Reading Standards for Literature	Reading
26	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.3	Reading Standards for Literature	Reading
27	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.6	Reading Standards for Literature	Reading
28	Multiple Choice	1	CCSS.ELA-Literacy.RL.6.2	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.4	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.4	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.7	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.8	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
35	Multiple Choice	1	CCSS.ELA-Literacy.RI.6.3	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RL.6.5	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RL.6.2	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RL.6.6	Reading Standards for Literature	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RI.6.6	Reading Standards for Informational Text	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RI.6.2	Reading Standards for Informational Text	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RI.6.5	Reading Standards for Informational Text	Writing to Sources
42	Constructed Response	4	CCSS.ELA-Literacy.RI.6.9	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 2022 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.7.3	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.7.2	Reading Standards for Literature	Reading
4	Multiple Choice	1	CCSS.ELA-Literacy.L.7.5	Language Standards	Reading
5	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.6	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	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.7.6	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.4	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.8	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.5	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.2	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.6	Reading Standards for Literature	Reading
23	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Reading
25	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.2	Reading Standards for Literature	Reading
26	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Reading
27	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Reading
28	Multiple Choice	1	CCSS.ELA-Literacy.RL.7.2	Reading Standards for Literature	Reading
29	Multiple Choice	1	CCSS.ELA-Literacy.L.7.4	Language Standards	Reading
30	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	CCSS.ELA-Literacy.L.7.4	Language Standards	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.5	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.2	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.6	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	CCSS.ELA-Literacy.RI.7.2	Reading Standards for Informational Text	Reading
Session 2					
36	Constructed Response	2	CCSS.ELA-Literacy.RI.7.3	Reading Standards for Informational Text	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RI.7.8	Reading Standards for Informational Text	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RI.7.4	Reading Standards for Informational Text	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RL.7.3	Reading Standards for Literature	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RL.7.4	Reading Standards for Literature	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RL.7.6	Reading Standards for Literature	Writing to Sources
42	Constructed Response	2	CCSS.ELA-Literacy.RL.7.5	Reading Standards for Literature	Writing to Sources
43	Constructed Response	4	CCSS.ELA-Literacy.RL.7.2	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 2022 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 G6. ELA Grade 8 Operational Item Map

Question	Type	Points	Standard	Strand	Subscore
			Session 1		
1	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.4	Reading Standards for Literature	Reading
2	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Reading
3	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.4	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.RL.8.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.2	Reading Standards for Literature	Reading
7	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.6	Reading Standards for Literature	Reading
8	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.6	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.3	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.2	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	CCSS.ELA-Literacy.L.8.4	Language Standards	Reading
13	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.5	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.2	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.L.8.4	Language Standards	Reading
24	Multiple Choice	1	CCSS.ELA-Literacy.RL.8.6	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.8.3	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.8.2	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
31	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.3	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.6	Reading Standards for Informational Text	Reading

Question	Type	Points	Standard	Strand	Subscore
34	Multiple Choice	1	CCSS.ELA-Literacy.RI.8.8	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.RI.8.3	Reading Standards for Informational Text	Writing to Sources
37	Constructed Response	2	CCSS.ELA-Literacy.RI.8.8	Reading Standards for Informational Text	Writing to Sources
38	Constructed Response	2	CCSS.ELA-Literacy.RI.8.4	Reading Standards for Informational Text	Writing to Sources
39	Constructed Response	2	CCSS.ELA-Literacy.RL.8.4	Reading Standards for Literature	Writing to Sources
40	Constructed Response	2	CCSS.ELA-Literacy.RL.8.5	Reading Standards for Literature	Writing to Sources
41	Constructed Response	2	CCSS.ELA-Literacy.RL.8.3	Reading Standards for Literature	Writing to Sources
42	Constructed Response	2	CCSS.ELA-Literacy.RL.8.6	Reading Standards for Literature	Writing to Sources
43	Constructed Response	4	CCSS.ELA-Literacy.RL.8.6	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 2022 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
4	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
7	Multiple Choice	1	CCSS.Math.Content.3.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
9	Multiple Choice	1	CCSS.Math.Content.3.NF.A.2b	Number and Operations - Fractions	Number and Operations - Fractions

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

Question	Type	Points	Standard	Cluster	Subscore
34	Constructed Response	2	CCSS.Math.Content.3.NBT.A.3	Number and Operations in Base Ten	
35	Constructed Response	2	CCSS.Math.Content.3.MD.C.5b	Measurement and Data	Measurement and Data
36	Constructed Response	2	CCSS.Math.Content.3.OA.D.8	Operations and Algebraic Thinking	Operations and Algebraic Thinking
37	Constructed Response	2	CCSS.Math.Content.3.OA.A.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
38	Constructed Response	2	CCSS.Math.Content.3.NF.A.3b	Number and Operations - Fractions	Number and Operations - Fractions
39	Constructed Response	2	CCSS.Math.Content.3.OA.D.9	Operations and Algebraic Thinking	Operations and Algebraic Thinking
40	Constructed Response	3	CCSS.Math.Content.3.G.A.2	Geometry	

*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.MD.C.5b	Measurement and Data	
2	Multiple Choice	1	CCSS.Math.Content.4.G.A.1	Geometry	
3	Multiple Choice	1	CCSS.Math.Content.3.MD.D.8	Measurement and Data	
6	Multiple Choice	1	CCSS.Math.Content.4.NBT.B.6	Number and Operations in Base Ten	Number and Operations in Base Ten
7	Multiple Choice	1	CCSS.Math.Content.4.NF.A.1	Number and Operations - Fractions	Number and Operations - Fractions
8	Multiple Choice	1	CCSS.Math.Content.4.MD.A.3	Measurement and Data	
9	Multiple Choice	1	CCSS.Math.Content.4.NBT.A.1	Number and Operations in Base Ten	Number and Operations in Base Ten

Question	Type	Points	Standard	Cluster	Subscore
10	Multiple Choice	1	CCSS.Math.Content.4.OA.A.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
12	Multiple Choice	1	CCSS.Math.Content.4.NBT.A.2	Number and Operations in Base Ten	Number and Operations in Base Ten
13	Multiple Choice	1	CCSS.Math.Content.4.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
14	Multiple Choice	1	CCSS.Math.Content.4.NF.B.3a	Number and Operations - Fractions	Number and Operations - Fractions
16	Multiple Choice	1	CCSS.Math.Content.4.MD.B.4	Measurement and Data	
17	Multiple Choice	1	CCSS.Math.Content.4.NBT.A.3	Number and Operations in Base Ten	Number and Operations in Base Ten
18	Multiple Choice	1	CCSS.Math.Content.4.NF.B.4b	Number and Operations - Fractions	Number and Operations - Fractions
20	Multiple Choice	1	CCSS.Math.Content.4.OA.B.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
21	Multiple Choice	1	CCSS.Math.Content.4.NF.A.2	Number and Operations - Fractions	Number and Operations - Fractions
22	Multiple Choice	1	CCSS.Math.Content.4.MD.C.6	Measurement and Data	
23	Multiple Choice	1	CCSS.Math.Content.4.OA.A.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
24	Multiple Choice	1	CCSS.Math.Content.4.MD.C.5a	Measurement and Data	
25	Multiple Choice	1	CCSS.Math.Content.4.NF.B.4a	Number and Operations - Fractions	Number and Operations - Fractions
27	Multiple Choice	1	CCSS.Math.Content.4.NBT.B.5	Number and Operations in Base Ten	Number and Operations in Base Ten
28	Multiple Choice	1	CCSS.Math.Content.4.G.A.2	Geometry	
30	Multiple Choice	1	CCSS.Math.Content.4.NBT.B.6	Number and Operations in Base Ten	Number and Operations in Base Ten
Session 2					
31	Multiple Choice	1	CCSS.Math.Content.4.NF.B.3d	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.3.MD.B.4	Measurement and Data	
34	Multiple Choice	1	CCSS.Math.Content.4.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking

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

*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	Measurement and Data	Measurement and Data	CCSS.Math.Content.4.MD.A.2
2	Multiple Choice	1	Number and Operations - Fractions	Number and Operations - Fractions	CCSS.Math.Content.5.NF.A.1
3	Multiple Choice	1	Geometry		CCSS.Math.Content.5.G.B.4

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

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

*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.EE.B.5	Expressions and Equations	Expressions and Equations
2	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
3	Multiple Choice	1	CCSS.Math.Content.6.G.A.4	Geometry	
4	Multiple Choice	1	CCSS.Math.Content.6.EE.C.9	Expressions and Equations	Expressions and Equations
5	Multiple Choice	1	CCSS.Math.Content.6.RP.A.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
7	Multiple Choice	1	CCSS.Math.Content.6.NS.A.1	The Number System	The Number System
8	Multiple Choice	1	CCSS.Math.Content.6.EE.B.6	Expressions and Equations	Expressions and Equations
9	Multiple Choice	1	CCSS.Math.Content.5.G.A.1	The Number System	The Number System
11	Multiple Choice	1	CCSS.Math.Content.6.EE.A.2c	Expressions and Equations	Expressions and Equations
12	Multiple Choice	1	CCSS.Math.Content.6.G.A.1	Geometry	
13	Multiple Choice	1	CCSS.Math.Content.6.EE.B.7	Expressions and Equations	Expressions and Equations
15	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
16	Multiple Choice	1	CCSS.Math.Content.6.EE.B.5	Expressions and Equations	Expressions and Equations
17	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
19	Multiple Choice	1	CCSS.Math.Content.6.EE.B.8	Expressions and Equations	Expressions and Equations
20	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
22	Multiple Choice	1	CCSS.Math.Content.6.NS.A.1	The Number System	The Number System
23	Multiple Choice	1	CCSS.Math.Content.6.G.A.1	Geometry	
24	Multiple Choice	1	CCSS.Math.Content.6.EE.A.1	Expressions and Equations	Expressions and Equations
26	Multiple Choice	1	CCSS.Math.Content.6.NS.C.6c	The Number System	The Number System

Question	Type	Points	Standard	Cluster	Subscore
27	Multiple Choice	1	CCSS.Math.Content.6.RP.A.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
28	Multiple Choice	1	CCSS.Math.Content.6.G.A.3	Geometry	
30	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
31	Multiple Choice	1	CCSS.Math.Content.6.EE.A.2a	Expressions and Equations	Expressions and Equations
Session 2					
32	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
33	Multiple Choice	1	CCSS.Math.Content.6.G.A.4	Geometry	
34	Multiple Choice	1	CCSS.Math.Content.6.RP.A.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
35	Multiple Choice	1	CCSS.Math.Content.6.NS.C.5	The Number System	The Number System
36	Multiple Choice	1	CCSS.Math.Content.6.G.A.1	Geometry	
37	Multiple Choice	1	CCSS.Math.Content.6.RP.A.3c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
38	Multiple Choice	1	CCSS.Math.Content.6.EE.B.6	Expressions and Equations	Expressions and Equations
39	Constructed Response	2	CCSS.Math.Content.6.EE.A.2a	Expressions and Equations	Expressions and Equations
40	Constructed Response	2	CCSS.Math.Content.6.NS.C.8	The Number System	The Number System
41	Constructed Response	2	CCSS.Math.Content.6.NS.B.4	The Number System	The Number System
42	Constructed Response	2	CCSS.Math.Content.6.G.A.2	Geometry	
43	Constructed Response	2	CCSS.Math.Content.6.EE.A.3	Expressions and Equations	Expressions and Equations
44	Constructed Response	2	CCSS.Math.Content.6.RP.A.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
45	Constructed Response	2	CCSS.Math.Content.6.EE.B.7	Expressions and Equations	Expressions and Equations

Question	Type	Points	Standard	Cluster	Subscore
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.SP.C.5	Statistics and Probability	
2	Multiple Choice	1	CCSS.Math.Content.7.RP.A.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
3	Multiple Choice	1	CCSS.Math.Content.7.NS.A.3	The Number System	The Number System
4	Multiple Choice	1	CCSS.Math.Content.7.SP.C.7b	Statistics and Probability	
5	Multiple Choice	1	CCSS.Math.Content.7.EE.A.1	Expressions and Equations	Expressions and Equations
6	Multiple Choice	1	CCSS.Math.Content.7.G.A.1	Geometry	
7	Multiple Choice	1	CCSS.Math.Content.7.EE.A.2	Expressions and Equations	Expressions and Equations
9	Multiple Choice	1	CCSS.Math.Content.6.SP.A.3	Statistics and Probability	
10	Multiple Choice	1	CCSS.Math.Content.7.SP.C.8b	Statistics and Probability	
11	Multiple Choice	1	CCSS.Math.Content.7.RP.A.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
12	Multiple Choice	1	CCSS.Math.Content.7.EE.B.4b	Expressions and Equations	Expressions and Equations
14	Multiple Choice	1	CCSS.Math.Content.7.EE.B.4a	Expressions and Equations	Expressions and Equations
15	Multiple Choice	1	CCSS.Math.Content.7.RP.A.2d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
18	Multiple Choice	1	CCSS.Math.Content.7.SP.C.6	Statistics and Probability	
19	Multiple Choice	1	CCSS.Math.Content.7.NS.A.1c	The Number System	The Number System

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

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

*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.EE.B.6	Expressions and Equations	Expressions and Equations
2	Multiple Choice	1	CCSS.Math.Content.7.G.B.5	Expressions and Equations	Expressions and Equations
3	Multiple Choice	1	CCSS.Math.Content.8.G.A.4	Geometry	Geometry
4	Multiple Choice	1	CCSS.Math.Content.8.SP.A.2	Statistics and Probability	
5	Multiple Choice	1	CCSS.Math.Content.8.EE.C.7a	Expressions and Equations	Expressions and Equations
6	Multiple Choice	1	CCSS.Math.Content.8.F.A.1	Functions	Functions
7	Multiple Choice	1	CCSS.Math.Content.8.EE.B.5	Expressions and Equations	Expressions and Equations
8	Multiple Choice	1	CCSS.Math.Content.8.EE.A.1	Expressions and Equations	Expressions and Equations
10	Multiple Choice	1	CCSS.Math.Content.8.EE.A.1	Expressions and Equations	Expressions and Equations

Question	Type	Points	Standard	Cluster	Subscore
11	Multiple Choice	1	CCSS.Math.Content.8.F.B.4	Functions	Functions
12	Multiple Choice	1	CCSS.Math.Content.8.EE.C.7b	Expressions and Equations	Expressions and Equations
14	Multiple Choice	1	CCSS.Math.Content.8.SP.A.1	Statistics and Probability	
15	Multiple Choice	1	CCSS.Math.Content.8.F.B.5	Functions	Functions
16	Multiple Choice	1	CCSS.Math.Content.8.G.A.1a	Geometry	Geometry
18	Multiple Choice	1	CCSS.Math.Content.7.G.B.5	Expressions and Equations	Expressions and Equations
19	Multiple Choice	1	CCSS.Math.Content.8.G.A.4	Geometry	Geometry
20	Multiple Choice	1	CCSS.Math.Content.8.SP.A.2	Statistics and Probability	
21	Multiple Choice	1	CCSS.Math.Content.8.SP.A.4	Statistics and Probability	
23	Multiple Choice	1	CCSS.Math.Content.8.G.C.9	Geometry	Geometry
24	Multiple Choice	1	CCSS.Math.Content.7.G.A.2	Expressions and Equations	Expressions and Equations
25	Multiple Choice	1	CCSS.Math.Content.8.EE.B.6	Expressions and Equations	Expressions and Equations
26	Multiple Choice	1	CCSS.Math.Content.8.EE.B.5	Expressions and Equations	Expressions and Equations
28	Multiple Choice	1	CCSS.Math.Content.8.EE.A.1	Expressions and Equations	Expressions and Equations
29	Multiple Choice	1	CCSS.Math.Content.8.F.B.4	Functions	Functions
32	Multiple Choice	1	CCSS.Math.Content.8.EE.C.7b	Expressions and Equations	Expressions and Equations
33	Multiple Choice	1	CCSS.Math.Content.8.F.A.3	Functions	Functions
Session 2					
34	Multiple Choice	1	CCSS.Math.Content.8.SP.A.3	Statistics and Probability	
35	Multiple Choice	1	CCSS.Math.Content.8.EE.A.1	Expressions and Equations	Expressions and Equations
36	Multiple Choice	1	CCSS.Math.Content.8.G.A.5	Geometry	Geometry

Question	Type	Points	Standard	Cluster	Subscore
37	Multiple Choice	1	CCSS.Math.Content.8.G.A.2	Geometry	Geometry
38	Multiple Choice	1	CCSS.Math.Content.8.F.A.1	Functions	Functions
39	Multiple Choice	1	CCSS.Math.Content.8.EE.C.7a	Expressions and Equations	Expressions and Equations
40	Multiple Choice	1	CCSS.Math.Content.8.G.A.3	Geometry	Geometry
41	Constructed Response	2	CCSS.Math.Content.8.F.B.4	Functions	Functions
42	Constructed Response	2	CCSS.Math.Content.8.G.A.2	Geometry	Geometry
43	Constructed Response	2	CCSS.Math.Content.8.G.A.5	Geometry	Geometry
44	Constructed Response	2	CCSS.Math.Content.8.EE.C.7b	Expressions and Equations	Expressions and Equations
45	Constructed Response	2	CCSS.Math.Content.8.F.A.2	Functions	Functions
46	Constructed Response	2	CCSS.Math.Content.8.G.C.9	Geometry	Geometry
47	Constructed Response	2	CCSS.Math.Content.8.EE.C.7b	Expressions and Equations	Expressions and Equations
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	NGLS	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	NGLS	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	NGLS	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 Selected Subgroups

As described in Section 3: Validity, a principal component 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, SWD, SUA, SWD/SUA students using disability accommodations, and ELLs using ELL -related accommodations (ELL & SUA). Tables L1 and L12 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		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**6.08**	**24.31**	**24.31**
	1.43	5.72	30.03
	1.10	4.40	34.43
SWD	**6.16**	**24.64**	**24.64**
	1.61	6.45	31.09
	1.09	4.35	35.44
	1.02	4.07	39.50
SUA	**5.67**	**22.69**	**22.69**
	1.63	6.50	29.19
	1.09	4.34	33.53
	1.05	4.19	37.72

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**5.48**	**21.92**	**21.92**
	1.33	5.32	27.24
	1.05	4.22	31.46
	1.03	4.12	35.58
	1.02	4.08	39.66
SWD	**5.95**	**23.82**	**23.82**
	1.42	5.68	29.49
	1.07	4.28	33.77
SUA	**5.60**	**22.42**	**22.42**
	1.42	5.68	28.10
	1.13	4.53	32.63

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

| Demographic Category | Extracted Factor | | |
| | Eigenvalue | Variance Accounted for | |
		%	Cumulative %
ELL	**6.11**	**17.47**	**17.47**
	1.51	4.32	21.79
	1.19	3.39	25.18
	1.13	3.22	28.40
	1.06	3.04	31.43
	1.03	2.95	34.38
	1.00	2.87	37.25
SWD	**6.89**	**19.68**	**19.68**
	1.58	4.53	24.20
	1.20	3.42	27.62
	1.05	3.00	30.61
	1.02	2.91	33.53
	1.00	2.86	36.39
SUA	**6.62**	**18.93**	**18.93**
	1.60	4.56	23.48
	1.19	3.41	26.90
	1.07	3.06	29.96
	1.03	2.96	32.92
	1.01	2.90	35.81

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

| Demographic Category | Extracted Factor | | |
| | Eigenvalue | Variance Accounted for | |
		%	Cumulative %
ELL	**5.66**	**16.18**	**16.18**
	1.63	4.66	20.84
	1.19	3.41	24.25
	1.11	3.17	27.41
	1.06	3.04	30.45
	1.04	2.96	33.41
	1.03	2.94	36.35
	1.01	2.90	39.25
	1.01	2.88	42.13
SWD	**6.53**	**18.66**	**18.66**
	1.76	5.03	23.69

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
	1.14	3.25	26.94
	1.08	3.08	30.01
	1.02	2.92	32.93
	1.00	2.86	35.79
SUA	**6.50**	**18.58**	**18.58**
	1.74	4.96	23.54
	1.10	3.14	26.68
	1.08	3.07	29.75
	1.02	2.92	32.67
	1.00	2.86	35.53

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**6.33**	**17.58**	**17.58**
	1.61	4.46	22.04
	1.19	3.30	25.33
	1.13	3.13	28.46
	1.10	3.06	31.52
	1.07	2.98	34.50
	1.04	2.90	37.40
	1.03	2.85	40.25
SWD	**7.30**	**20.29**	**20.29**
	1.88	5.21	25.50
	1.18	3.28	28.77
	1.07	2.96	31.74
	1.05	2.93	34.67
	1.02	2.83	37.49
	1.01	2.79	40.29
SUA	**7.34**	**20.38**	**20.38**
	1.75	4.87	25.26
	1.19	3.30	28.56
	1.09	3.03	31.59
	1.06	2.95	34.54
	1.02	2.82	37.37

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**6.23**	**17.29**	**17.29**
	1.83	5.09	22.38
	1.11	3.09	25.47
	1.09	3.02	28.49
	1.07	2.96	31.45
	1.05	2.91	34.36
	1.04	2.88	37.23
	1.02	2.83	40.06
	1.01	2.81	42.87
SWD	**7.22**	**20.06**	**20.06**
	2.10	5.84	25.90
	1.06	2.94	28.83
	1.05	2.91	31.75
	1.02	2.84	34.59
	1.00	2.79	37.38
SUA	**7.51**	**20.86**	**20.86**
	2.02	5.62	26.48
	1.06	2.93	29.42
	1.03	2.87	32.29
	1.00	2.78	35.07

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**7.33**	**21.57**	**21.57**
	1.69	4.98	26.54
	1.29	3.81	30.35
	1.12	3.30	33.65
	1.07	3.15	36.80
	1.01	2.98	39.77
SWD	**8.93**	**26.26**	**26.26**
	1.63	4.78	31.04
	1.15	3.37	34.41
	1.04	3.05	37.46

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
SUA	**7.97**	**23.43**	**23.43**
	1.57	4.63	28.06
	1.17	3.45	31.52
	1.09	3.22	34.73
	1.02	2.99	37.72

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**7.66**	**20.16**	**20.16**
	1.63	4.30	24.46
	1.27	3.35	27.81
	1.05	2.78	30.58
	1.04	2.74	33.32
SWD	**9.49**	**24.99**	**24.99**
	1.59	4.19	29.18
	1.12	2.96	32.14
SUA	**8.75**	**23.01**	**23.01**
	1.57	4.14	27.15
	1.12	2.95	30.11

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**6.54**	**17.20**	**17.20**
	1.49	3.93	21.13
	1.20	3.15	24.28
	1.17	3.07	27.35
	1.09	2.88	30.23
	1.06	2.80	33.03
	1.04	2.74	35.76
	1.03	2.70	38.47
	1.00	2.64	41.11

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
SWD	**9.22**	**24.28**	**24.28**
	1.37	3.61	27.88
	1.16	3.04	30.93
	1.03	2.72	33.64
	1.00	2.64	36.28
SUA	**8.61**	**22.67**	**22.67**
	1.43	3.75	26.42
	1.13	2.99	29.41
	1.04	2.75	32.15
	1.02	2.68	34.83

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**5.80**	**14.87**	**14.87**
	1.45	3.72	18.59
	1.33	3.42	22.01
	1.14	2.93	24.94
	1.09	2.79	27.72
	1.08	2.76	30.49
	1.07	2.75	33.23
	1.03	2.64	35.87
	1.02	2.62	38.49
SWD	**7.71**	**19.77**	**19.77**
	1.35	3.45	23.22
	1.13	2.91	26.12
	1.06	2.71	28.84
	1.02	2.62	31.45
SUA	**7.61**	**19.52**	**19.52**
	1.31	3.36	22.88
	1.14	2.92	25.79
	1.07	2.74	28.54
	1.05	2.69	31.23

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

Demographic Category	Extracted Factor		
	Eigenvalue	Variance Accounted for	
		%	Cumulative %
ELL	**6.07**	**14.81**	**14.81**
	1.53	3.73	18.54
	1.26	3.08	21.62
	1.16	2.84	24.46
	1.11	2.71	27.17
	1.09	2.66	29.82
	1.07	2.60	32.43
	1.06	2.57	35.00
	1.03	2.52	37.52
SWD	**7.59**	**18.52**	**18.52**
	1.35	3.29	21.81
	1.18	2.89	24.70
	1.05	2.56	27.26
	1.04	2.54	29.80
	1.02	2.49	32.29
	1.00	2.44	34.74
SUA	**7.62**	**18.59**	**18.59**
	1.32	3.22	21.81
	1.19	2.90	24.71
	1.08	2.63	27.34
	1.05	2.55	29.89
	1.03	2.50	32.39
	1.01	2.45	34.85

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

| Demographic Category | Extracted Factor | | |
| | Eigenvalue | Variance Accounted for | |
		%	Cumulative %
ELL	**7.49**	**18.26**	**18.26**
	1.39	3.40	21.66
	1.28	3.11	24.77
	1.16	2.83	27.60
	1.11	2.72	30.32
	1.03	2.51	32.83
	1.01	2.47	35.30
SWD	**7.23**	**17.63**	**17.63**
	1.34	3.27	20.91
	1.26	3.08	23.98
	1.13	2.77	26.75
	1.09	2.66	29.41
	1.01	2.47	31.88
SUA	**7.30**	**17.81**	**17.81**
	1.34	3.28	21.09
	1.25	3.04	24.13
	1.15	2.80	26.93
	1.10	2.69	29.62
	1.02	2.48	32.10

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

Table M1. ELA Grade 3 Classical Item Analysis

Item	Type	N-Count	*P*-Value	% Omit	PBis
1	MC	152,259	0.80	0	0.45
2	MC	152,259	0.90	0	0.40
3	MC	152,259	0.75	0	0.32
4	MC	152,259	0.49	0	0.37
5	MC	152,259	0.86	0	0.47
6	MC	152,259	0.51	0	0.21
7	MC	152,259	0.48	0	0.30
8	MC	152,259	0.76	0	0.48
9	MC	152,259	0.47	0	0.37
10	MC	152,259	0.70	0	0.41
11	MC	152,259	0.47	0	0.32
12	MC	152,259	0.61	0	0.34
13	MC	152,259	0.77	0	0.47
14	MC	152,259	0.81	0	0.45
15	MC	152,259	0.62	0	0.47
16	MC	152,259	0.61	0	0.48
17	MC	152,259	0.67	0	0.49
18	MC	152,259	0.61	0	0.36
19	CR	152,259	0.55	0	0.59
20	CR	152,259	0.53	0	0.56
21	CR	152,259	0.38	0	0.56
22	CR	152,259	0.56	0	0.61
23	CR	152,259	0.53	0	0.60
24	CR	152,259	0.57	0	0.68
25	CR	152,259	0.31	0	0.67

Table M2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	*P*-Value	% Omit	PBis
1	MC	153,437	0.75	0	0.35
2	MC	153,437	0.70	0	0.45
3	MC	153,437	0.67	0	0.36
4	MC	153,437	0.59	0	0.39

Item	Type	N-Count	*P*-Value	% Omit	PBis
5	MC	153,437	0.84	0	0.44
6	MC	153,437	0.81	0	0.50
7	MC	153,437	0.69	0	0.34
8	MC	153,437	0.51	0	0.43
9	MC	153,437	0.65	0	0.41
10	MC	153,437	0.64	0	0.47
11	MC	153,437	0.76	0	0.45
12	MC	153,437	0.65	0	0.39
13	MC	153,437	0.56	0	0.38
14	MC	153,437	0.42	0	0.21
15	MC	153,437	0.72	0	0.44
16	MC	153,437	0.54	0	0.30
17	MC	153,437	0.48	0	0.33
18	MC	153,437	0.55	0	0.39
19	CR	153,437	0.67	0	0.56
20	CR	153,437	0.52	0	0.49
21	CR	153,437	0.58	0	0.64
22	CR	153,437	0.56	0	0.60
23	CR	153,437	0.60	0	0.65
24	CR	153,437	0.49	0	0.60
25	CR	153,437	0.42	0	0.69

Table M3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis
1	MC	159,499	0.69	0	0.40
2	MC	159,499	0.74	0	0.43
3	MC	159,499	0.87	0	0.25
4	MC	159,499	0.51	0	0.26
5	MC	159,499	0.83	0	0.44
6	MC	159,499	0.73	0	0.42
7	MC	159,499	0.81	0	0.38
8	MC	159,499	0.69	0	0.39
9	MC	159,499	0.43	0	0.18
10	MC	159,499	0.71	0	0.44
11	MC	159,499	0.63	0	0.32
12	MC	159,499	0.77	0	0.42
13	MC	159,499	0.64	0	0.36
14	MC	159,499	0.68	0	0.42

Item	Type	N-Count	*p*-value	% Omit	PBis
15	MC	159,499	0.50	0	0.31
16	MC	159,499	0.59	0	0.45
17	MC	159,499	0.59	0	0.49
18	MC	159,499	0.47	0	0.27
19	MC	159,499	0.43	0	0.29
20	MC	159,499	0.68	0	0.38
21	MC	159,499	0.60	0	0.40
22	MC	159,499	0.80	0	0.47
23	MC	159,499	0.79	0	0.48
24	MC	159,499	0.59	0	0.40
25	MC	159,499	0.50	0	0.43
26	MC	159,499	0.30	0	0.17
27	MC	159,499	0.52	0	0.36
28	MC	159,499	0.51	0	0.43
29	CR	159,499	0.69	0	0.57
30	CR	159,499	0.73	0	0.60
31	CR	159,499	0.57	0	0.58
32	CR	159,499	0.66	0	0.58
33	CR	159,499	0.62	0	0.59
34	CR	159,499	0.62	0	0.61
35	CR	159,499	0.47	0	0.66

Table M4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis
1	MC	157,750	0.78	0	0.28
2	MC	157,750	0.73	0	0.17
3	MC	157,750	0.46	0	0.28
4	MC	157,750	0.59	0	0.37
5	MC	157,750	0.86	0	0.44
6	MC	157,750	0.75	0	0.34
7	MC	157,750	0.84	0	0.34
8	MC	157,750	0.81	0	0.41
9	MC	157,750	0.69	0	0.37
10	MC	157,750	0.45	0	0.25
11	MC	157,750	0.70	0	0.33
12	MC	157,750	0.74	0	0.41
13	MC	157,750	0.85	0	0.41
14	MC	157,750	0.55	0	0.16

Item	Type	N-Count	*p*-value	% Omit	PBis
15	MC	157,750	0.68	0	0.36
16	MC	157,750	0.67	0	0.35
17	MC	157,750	0.75	0	0.49
18	MC	157,750	0.69	0	0.27
19	MC	157,750	0.49	0	0.24
20	MC	157,750	0.53	0	0.31
21	MC	157,750	0.44	0	0.26
22	MC	157,750	0.44	0	0.27
23	MC	157,750	0.81	0	0.50
24	MC	157,750	0.59	0	0.36
25	MC	157,750	0.70	0	0.45
26	MC	157,750	0.67	0	0.40
27	MC	157,750	0.63	0	0.40
28	MC	157,750	0.57	0	0.38
29	CR	157,750	0.73	0	0.51
30	CR	157,750	0.62	0	0.54
31	CR	157,750	0.65	0	0.62
32	CR	157,750	0.74	0	0.61
33	CR	157,750	0.67	0	0.61
34	CR	157,750	0.68	0	0.60
35	CR	157,750	0.47	0	0.65

Table M5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis
1	MC	144,053	0.81	0	0.41
2	MC	144,053	0.36	0	0.25
3	MC	144,053	0.71	0	0.36
4	MC	144,053	0.76	0	0.29
5	MC	144,053	0.61	0	0.38
6	MC	144,053	0.69	0	0.30
7	MC	144,053	0.81	0	0.39
8	MC	144,053	0.56	0	0.33
9	MC	144,053	0.88	0	0.51
10	MC	144,053	0.72	0	0.47
11	MC	144,053	0.62	0	0.15
12	MC	144,053	0.85	0	0.49
13	MC	144,053	0.64	0	0.38
14	MC	144,053	0.49	0	0.23

Item	Type	N-Count	p-value	% Omit	PBis
15	MC	144,053	0.73	0	0.41
16	MC	144,053	0.64	0	0.35
17	MC	144,053	0.70	0	0.46
18	MC	144,053	0.42	0	0.35
19	MC	144,053	0.75	0	0.44
20	MC	144,053	0.34	0	0.11
21	MC	144,053	0.54	0	0.35
22	MC	144,053	0.70	0	0.37
23	MC	144,053	0.40	0	0.29
24	MC	144,053	0.50	0	0.36
25	MC	144,053	0.61	0	0.42
26	MC	144,053	0.71	0	0.49
27	MC	144,053	0.59	0	0.40
28	MC	144,053	0.82	0	0.48
29	CR	144,053	0.77	0	0.58
30	CR	144,053	0.71	0	0.64
31	CR	144,053	0.73	0	0.68
32	CR	144,053	0.76	0	0.61
33	CR	144,053	0.65	0	0.63
34	CR	144,053	0.75	0	0.66
35	CR	144,053	0.69	0	0.59
36	CR	144,053	0.52	0	0.67

Table M6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	p-value	% Omit	PBis
1	MC	138,380	0.84	0	0.43
2	MC	138,380	0.79	0	0.42
3	MC	138,380	0.81	0	0.45
4	MC	138,380	0.65	0	0.39
5	MC	138,380	0.43	0	0.28
6	MC	138,380	0.73	0	0.44
7	MC	138,380	0.54	0	0.22
8	MC	138,380	0.61	0	0.34
9	MC	138,380	0.45	0	0.21
10	MC	138,380	0.58	0	0.23
11	MC	138,380	0.69	0	0.30
12	MC	138,380	0.52	0	0.36
13	MC	138,380	0.64	0	0.33

Item	Type	N-Count	p-value	% Omit	PBis
14	MC	138,380	0.70	0	0.43
15	MC	138,380	0.71	0	0.48
16	MC	138,380	0.70	0	0.44
17	MC	138,380	0.68	0	0.49
18	MC	138,380	0.70	0	0.45
19	MC	138,380	0.56	0	0.43
20	MC	138,380	0.64	0	0.38
21	MC	138,380	0.78	0	0.41
22	MC	138,380	0.62	0	0.36
23	MC	138,380	0.57	0	0.37
24	MC	138,380	0.55	0	0.35
25	MC	138,380	0.69	0	0.44
26	MC	138,380	0.78	0	0.56
27	MC	138,380	0.58	0	0.42
28	MC	138,380	0.65	0	0.38
29	CR	138,380	0.81	0	0.62
30	CR	138,380	0.82	0	0.64
31	CR	138,380	0.76	0	0.58
32	CR	138,380	0.76	0	0.62
33	CR	138,380	0.77	0	0.59
34	CR	138,380	0.73	0	0.64
35	CR	138,380	0.74	0	0.64
36	CR	138,380	0.58	0	0.68

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	p-value	% Omit	PBis
1	MC	153,424	0.78	0	0.44
2	MC	153,424	0.62	0	0.48
3	MC	153,424	0.35	0	0.40
4	MC	153,424	0.79	0	0.51
5	MC	153,424	0.83	0	0.37
6	MC	153,424	0.54	0	0.51
7	MC	153,424	0.63	0	0.46
8	MC	153,424	0.63	0	0.56
9	MC	153,424	0.68	0	0.53
10	MC	153,424	0.40	0	0.16
11	MC	153,424	0.64	0	0.55
12	MC	153,424	0.75	0	0.50

Item	Type	N-Count	p-value	% Omit	PBis
13	MC	153,424	0.41	0	0.53
14	MC	153,424	0.65	0	0.52
15	MC	153,424	0.75	0	0.41
16	MC	153,424	0.63	0	0.50
17	MC	153,424	0.51	0	0.38
18	MC	153,424	0.78	0	0.48
19	MC	153,424	0.29	0	0.28
20	MC	153,424	0.96	0	0.27
21	MC	153,424	0.82	0	0.52
22	MC	153,424	0.93	0	0.35
23	MC	153,424	0.57	0	0.51
24	MC	153,424	0.45	0	0.55
25	MC	153,424	0.75	0	0.57
26	MC	153,424	0.50	0	0.26
27	MC	153,424	0.36	0	0.48
28	CR	153,424	0.55	0	0.69
29	CR	153,424	0.60	0	0.60
30	CR	153,424	0.22	0	0.48
31	CR	153,424	0.44	0	0.44
32	CR	153,424	0.59	0	0.66
33	CR	153,424	0.40	0	0.60
34	CR	153,424	0.55	0	0.67

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	p-value	% Omit	PBis
1	MC	156,788	0.77	0	0.42
2	MC	156,788	0.73	0	0.41
3	MC	156,788	0.67	0	0.48
4	MC	156,788	0.69	0	0.49
5	MC	156,788	0.73	0	0.45
6	MC	156,788	0.54	0	0.56
7	MC	156,788	0.41	0	0.53
8	MC	156,788	0.63	0	0.61
9	MC	156,788	0.53	0	0.50
10	MC	156,788	0.90	0	0.40
11	MC	156,788	0.76	0	0.51
12	MC	156,788	0.47	0	0.41
13	MC	156,788	0.62	0	0.50

Item	Type	N-Count	*p*-value	% Omit	PBis
14	MC	156,788	0.65	0	0.43
15	MC	156,788	0.46	0	0.41
16	MC	156,788	0.74	0	0.39
17	MC	156,788	0.69	0	0.40
18	MC	156,788	0.44	0	0.48
19	MC	156,788	0.55	0	0.52
20	MC	156,788	0.55	0	0.49
21	MC	156,788	0.61	0	0.46
22	MC	156,788	0.71	0	0.46
23	MC	156,788	0.47	0	0.48
24	MC	156,788	0.74	0	0.51
25	MC	156,788	0.82	0	0.48
26	MC	156,788	0.67	0	0.56
27	MC	156,788	0.88	0	0.41
28	MC	156,788	0.44	0	0.49
29	MC	156,788	0.54	0	0.36
30	MC	156,788	0.68	0	0.55
31	MC	156,788	0.81	0	0.42
32	CR	156,788	0.69	0	0.58
33	CR	156,788	0.25	0	0.53
34	CR	156,788	0.36	0	0.49
35	CR	156,788	0.50	0	0.60
36	CR	156,788	0.40	0	0.68
37	CR	156,788	0.55	0	0.68
38	CR	156,788	0.33	0	0.66

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis
1	MC	151,633	0.73	0	0.49
2	MC	151,633	0.59	0	0.64
3	MC	151,633	0.62	0	0.34
4	MC	151,633	0.72	0	0.46
5	MC	151,633	0.47	0	0.51
6	MC	151,633	0.50	0	0.49
7	MC	151,633	0.46	0	0.27
8	MC	151,633	0.52	0	0.50
9	MC	151,633	0.62	0	0.54
10	MC	151,633	0.82	0	0.48

Item	Type	N-Count	*p*-value	% Omit	PBis
11	MC	151,633	0.68	0	0.50
12	MC	151,633	0.46	0	0.36
13	MC	151,633	0.60	0	0.54
14	MC	151,633	0.38	0	0.53
15	MC	151,633	0.43	0	0.48
16	MC	151,633	0.43	0	0.52
17	MC	151,633	0.56	0	0.38
18	MC	151,633	0.57	0	0.58
19	MC	151,633	0.45	0	0.38
20	MC	151,633	0.58	0	0.49
21	MC	151,633	0.56	0	0.49
22	MC	151,633	0.51	0	0.32
23	MC	151,633	0.60	0	0.53
24	MC	151,633	0.36	0	0.62
25	MC	151,633	0.53	0	0.41
26	MC	151,633	0.79	0	0.33
27	MC	151,633	0.65	0	0.53
28	MC	151,633	0.69	0	0.52
29	MC	151,633	0.89	0	0.36
30	MC	151,633	0.61	0	0.61
31	MC	151,633	0.78	0	0.41
32	CR	151,633	0.33	0	0.65
33	CR	151,633	0.24	0	0.58
34	CR	151,633	0.30	0	0.57
35	CR	151,633	0.51	0	0.65
36	CR	151,633	0.55	0	0.62
37	CR	151,633	0.43	0	0.76
38	CR	151,633	0.44	0	0.73

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis
1	MC	147,320	0.68	0	0.42
2	MC	147,320	0.53	0	0.43
3	MC	147,320	0.54	0	0.48
4	MC	147,320	0.71	0	0.46
5	MC	147,320	0.55	0	0.40
6	MC	147,320	0.38	0	0.47
7	MC	147,320	0.55	0	0.51

Item	Type	N-Count	p-value	% Omit	PBis
8	MC	147,320	0.48	0	0.36
9	MC	147,320	0.49	0	0.43
10	MC	147,320	0.50	0	0.60
11	MC	147,320	0.66	0	0.48
12	MC	147,320	0.63	0	0.58
13	MC	147,320	0.62	0	0.57
14	MC	147,320	0.54	0	0.39
15	MC	147,320	0.44	0	0.46
16	MC	147,320	0.71	0	0.49
17	MC	147,320	0.55	0	0.43
18	MC	147,320	0.33	0	0.37
19	MC	147,320	0.61	0	0.44
20	MC	147,320	0.51	0	0.47
21	MC	147,320	0.59	0	0.51
22	MC	147,320	0.56	0	0.40
23	MC	147,320	0.44	0	0.47
24	MC	147,320	0.40	0	0.42
25	MC	147,320	0.53	0	0.32
26	MC	147,320	0.43	0	0.53
27	MC	147,320	0.79	0	0.46
28	MC	147,320	0.72	0	0.45
29	MC	147,320	0.53	0	0.52
30	MC	147,320	0.38	0	0.33
31	MC	147,320	0.61	0	0.49
32	CR	147,320	0.40	0	0.62
33	CR	147,320	0.46	0	0.59
34	CR	147,320	0.54	0	0.61
35	CR	147,320	0.22	0	0.61
36	CR	147,320	0.36	0	0.59
37	CR	147,320	0.12	0	0.51
38	CR	147,320	0.52	0	0.63
39	CR	147,320	0.55	0	0.64

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	p-value	% Omit	PBis
1	MC	141,730	0.75	0	0.42
2	MC	141,730	0.74	0	0.52
3	MC	141,730	0.67	0	0.44

Item	Type	N-Count	*p*-value	% Omit	PBis
4	MC	141,730	0.49	0	0.53
5	MC	141,730	0.65	0	0.39
6	MC	141,730	0.62	0	0.52
7	MC	141,730	0.53	0	0.29
8	MC	141,730	0.48	0	0.22
9	MC	141,730	0.48	0	0.53
10	MC	141,730	0.54	0	0.43
11	MC	141,730	0.55	0	0.42
12	MC	141,730	0.69	0	0.53
13	MC	141,730	0.63	0	0.39
14	MC	141,730	0.38	0	0.29
15	MC	141,730	0.64	0	0.47
16	MC	141,730	0.42	0	0.31
17	MC	141,730	0.72	0	0.44
18	MC	141,730	0.71	0	0.45
19	MC	141,730	0.57	0	0.42
20	MC	141,730	0.63	0	0.48
21	MC	141,730	0.54	0	0.47
22	MC	141,730	0.59	0	0.63
23	MC	141,730	0.59	0	0.40
24	MC	141,730	0.63	0	0.42
25	MC	141,730	0.50	0	0.54
26	MC	141,730	0.40	0	0.31
27	MC	141,730	0.59	0	0.45
28	MC	141,730	0.50	0	0.32
29	MC	141,730	0.81	0	0.34
30	MC	141,730	0.52	0	0.32
31	MC	141,730	0.50	0	0.48
32	MC	141,730	0.34	0	0.31
33	MC	141,730	0.51	0	0.37
34	CR	141,730	0.39	0	0.71
35	CR	141,730	0.66	0	0.62
36	CR	141,730	0.81	0	0.48
37	CR	141,730	0.55	0	0.49
38	CR	141,730	0.45	0	0.74
39	CR	141,730	0.45	0	0.57
40	CR	141,730	0.38	0	0.66
41	CR	141,730	0.30	0	0.71

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis
1	MC	89,746	0.83	0	0.38
2	MC	89,746	0.63	0	0.44
3	MC	89,746	0.73	0	0.44
4	MC	89,746	0.60	0	0.50
5	MC	89,746	0.60	0	0.34
6	MC	89,746	0.46	0	0.45
7	MC	89,746	0.67	0	0.27
8	MC	89,746	0.65	0	0.41
9	MC	89,746	0.53	0	0.44
10	MC	89,746	0.15	0	0.29
11	MC	89,746	0.49	0	0.45
12	MC	89,746	0.43	0	0.44
13	MC	89,746	0.48	0	0.45
14	MC	89,746	0.47	0	0.47
15	MC	89,746	0.71	0	0.39
16	MC	89,746	0.66	0	0.41
17	MC	89,746	0.51	0	0.48
18	MC	89,746	0.42	0	0.37
19	MC	89,746	0.51	0	0.46
20	MC	89,746	0.52	0	0.38
21	MC	89,746	0.57	0	0.50
22	MC	89,746	0.54	0	0.45
23	MC	89,746	0.47	0	0.49
24	MC	89,746	0.32	0	0.41
25	MC	89,746	0.47	0	0.43
26	MC	89,746	0.60	0	0.43
27	MC	89,746	0.63	0	0.42
28	MC	89,746	0.49	0	0.39
29	MC	89,746	0.55	0	0.38
30	MC	89,746	0.73	0	0.42
31	MC	89,746	0.66	0	0.44
32	MC	89,746	0.47	0	0.45
33	MC	89,746	0.61	0	0.38
34	CR	89,746	0.31	0	0.59
35	CR	89,746	0.37	0	0.64
36	CR	89,746	0.17	0	0.60
37	CR	89,746	0.29	0	0.68

Item	Type	N-Count	p-value	% Omit	PBis
38	CR	89,746	0.42	0	0.65
39	CR	89,746	0.26	0	0.68
40	CR	89,746	0.24	0	0.64
41	CR	89,746	0.23	0	0.52

Appendix N: IRT Statistics

Tables N.1 through N.12 show item calibration results for the operational items.

Table N1. ELA Grade 3 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3	d4
1	1	0.528	0.503	0.103				
2	1	0.830	0.605	0.177				
3	1	1.011	-0.752	0.172				
4	1	0.515	-0.348	0.095				
5	1	0.727	-0.673	0.120				
6	1	0.480	0.426	0.099				
7	1	1.127	0.002	0.207				
8	1	1.127	-0.260	0.206				
9	1	0.953	-0.911	0.129				
10	1	0.589	-0.217	0.133				
11	1	0.954	-1.123	0.138				
12	1	1.048	-0.060	0.199				
13	1	0.272	0.074	0.034				
14	1	0.495	-1.464	0.017				
15	1	0.873	-1.219	0.031				
16	1	1.322	-1.141	0.209				
17	1	0.737	0.458	0.162				
18	1	1.053	-1.778	0.062				
19	2	0.970	-0.230		1.056	-1.056		
20	2	0.932	-0.125		1.172	-1.172		
21	2	0.835	0.544		0.837	-0.837		
22	2	1.131	-0.157		1.148	-1.148		
23	2	1.121	-0.283		1.088	-1.088		
24	2	1.406	-0.288		0.922	-0.922		
25	4	1.028	0.894		1.614	0.396	-0.526	-1.484

Table N2. ELA Grade 4 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3	d4
1	1	1.036	-1.123	0.217				
2	1	0.692	-0.100	0.142				
3	1	0.550	-1.352	0.033				
4	1	0.889	-0.537	0.155				
5	1	1.272	-0.926	0.195				
6	1	0.606	-0.513	0.157				
7	1	1.001	-0.709	0.229				
8	1	0.916	0.287	0.151				
9	1	0.636	-0.524	0.097				
10	1	0.514	-1.028	0.029				
11	1	0.800	-0.239	0.208				
12	1	1.081	-0.142	0.206				
13	1	0.638	-0.022	0.093				

Item	Max Pts	a	b	c	d1	d2	d3	d4
14	1	0.803	0.696	0.221				
15	1	0.412	-0.182	0.029				
16	1	0.784	-0.758	0.119				
17	1	0.755	0.165	0.187				
18	1	0.472	1.429	0.203				
19	2	0.696	-0.097		1.231	-1.231		
20	2	1.074	-0.312		0.821	-0.821		
21	2	0.933	-0.911		1.209	-1.209		
22	2	0.902	0.038		0.798	-0.798		
23	2	1.147	-0.386		0.822	-0.822		
24	2	0.926	-0.273		0.882	-0.882		
25	4	0.903	0.410		1.389	0.688	-0.536	-1.541

Table N3. ELA Grade 5 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3	d4
1	1	1.254	0.434	0.214				
2	1	0.999	0.380	0.167				
3	1	1.141	-0.774	0.279				
4	1	1.134	0.586	0.283				
5	1	1.114	0.293	0.288				
6	1	1.230	-0.685	0.283				
7	1	1.251	1.608	0.223				
8	1	0.784	-0.322	0.249				
9	1	0.707	-1.244	0.021				
10	1	0.729	-0.057	0.261				
11	1	0.684	-0.677	0.071				
12	1	0.452	-0.705	0.021				
13	1	0.238	0.875	0.031				
14	1	0.911	-0.480	0.215				
15	1	1.467	0.151	0.240				
16	1	0.613	-0.671	0.111				
17	1	0.898	0.973	0.290				
18	1	1.216	0.165	0.241				
19	1	0.732	0.958	0.201				
20	1	0.929	0.742	0.273				
21	1	1.054	0.231	0.289				
22	1	0.679	-1.473	0.046				
23	1	0.465	-2.600	0.030				
24	1	1.038	-1.016	0.249				
25	1	0.890	-0.577	0.245				
26	1	0.400	0.299	0.108				
27	1	0.736	-0.491	0.173				
28	1	0.790	-0.673	0.178				
29	2	0.958	-1.005		0.816	-0.816		
30	2	0.820	-0.331		0.910	-0.910		

Item	Max Pts	a	b	c	d1	d2	d3	d4
31	2	0.801	-0.878		0.846	-0.846		
32	2	0.839	-0.730		0.903	-0.903		
33	2	0.890	-0.554		0.982	-0.982		
34	2	0.917	-0.493		0.833	-0.833		
35	4	0.759	0.177		1.548	0.715	-0.617	-1.646

Table N4. ELA Grade 6 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3	d4
1	1	0.605	-1.835	0.085				
2	1	0.490	0.682	0.135				
3	1	0.982	-1.346	0.175				
4	1	0.595	-0.854	0.228				
5	1	0.226	-2.489	0.031				
6	1	0.462	-1.465	0.150				
7	1	0.700	0.054	0.193				
8	1	0.755	-1.703	0.020				
9	1	0.383	0.724	0.095				
10	1	0.723	-1.384	0.090				
11	1	0.195	-0.442	0.020				
12	1	0.704	-0.895	0.128				
13	1	0.634	-0.606	0.167				
14	1	0.466	-1.134	0.032				
15	1	0.842	-0.509	0.172				
16	1	0.466	0.836	0.124				
17	1	1.197	-0.866	0.235				
18	1	0.597	-0.793	0.036				
19	1	0.639	-0.451	0.086				
20	1	0.685	0.053	0.158				
21	1	0.505	-0.447	0.018				
22	1	0.545	-0.785	0.089				
23	1	0.336	0.242	0.036				
24	1	0.706	1.067	0.227				
25	1	0.517	-0.841	0.052				
26	1	1.087	-0.595	0.224				
27	1	0.372	-1.354	0.017				
28	1	0.484	0.152	0.099				
29	2	0.935	-0.605		0.615	-0.615		
30	2	0.809	-0.565		1.035	-1.035		
31	2	0.720	-1.180		0.944	-0.944		
32	2	1.031	-0.999		0.757	-0.757		
33	2	0.973	-0.717		0.759	-0.759		
34	2	0.974	-0.743		0.790	-0.790		
35	4	0.801	0.160		1.577	0.606	-0.643	-1.540

Table N5. ELA Grade 7 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3	d4
1	1	0.666	-0.138	0.122				
2	1	1.056	-0.427	0.230				
3	1	0.822	1.025	0.192				
4	1	0.836	0.485	0.199				
5	1	0.704	-0.400	0.266				
6	1	0.906	0.007	0.223				
7	1	1.067	-0.951	0.240				
8	1	0.630	-0.124	0.227				
9	1	1.033	-0.312	0.270				
10	1	0.746	-0.706	0.187				
11	1	0.864	0.527	0.264				
12	1	0.465	2.780	0.250				
13	1	0.739	-0.935	0.113				
14	1	0.915	0.811	0.171				
15	1	0.359	0.592	0.132				
16	1	1.042	-0.392	0.255				
17	1	1.217	-1.084	0.256				
18	1	1.526	-1.128	0.288				
19	1	0.474	-0.216	0.054				
20	1	0.188	-1.355	0.037				
21	1	0.534	-0.667	0.039				
22	1	0.400	-1.238	0.027				
23	1	0.847	-0.899	0.298				
24	1	0.620	-0.617	0.219				
25	1	0.632	1.356	0.170				
26	1	0.678	-1.407	0.130				
27	1	0.576	-0.338	0.081				
28	1	0.451	-1.415	0.126				
29	2	1.000	-1.295		0.968	-0.968		
30	2	1.107	-0.854		0.757	-0.757		
31	2	1.335	-0.868		0.695	-0.695		
32	2	0.920	-0.850		0.847	-0.847		
33	2	1.214	-0.922		0.582	-0.582		
34	2	1.087	-1.139		0.855	-0.855		
35	2	1.048	-0.615		0.775	-0.775		
36	4	0.794	-0.044		1.345	0.640	-0.533	-1.452

Table N6. ELA Grade 8 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3	d4
1	1	0.930	-0.270	0.285				
2	1	0.719	1.271	0.289				
3	1	0.294	-0.534	0.037				
4	1	0.468	-0.500	0.123				
5	1	0.591	-0.049	0.197				

6	1	0.748	0.125	0.415			
7	1	0.777	0.441	0.200			
8	1	1.051	-0.855	0.405			
9	1	1.132	-0.668	0.379			
10	1	0.782	-0.081	0.249			
11	1	1.002	-0.323	0.321			
12	1	1.302	-0.335	0.459			
13	1	0.648	0.983	0.186			
14	1	0.383	0.553	0.213			
15	1	0.892	0.046	0.302			
16	1	0.927	0.455	0.269			
17	1	0.804	0.132	0.286			
18	1	1.743	-0.482	0.319			
19	1	1.065	0.393	0.296			
20	1	1.196	0.282	0.274			
21	1	0.997	-0.182	0.291			
22	1	1.034	-0.400	0.228			
23	1	1.260	-0.023	0.369			
24	1	0.802	-0.812	0.274			
25	1	0.629	-0.288	0.161			
26	1	0.994	-0.289	0.261			
27	1	1.163	0.284	0.232			
28	1	1.095	-0.273	0.228			
29	2	0.991	-1.194		0.935	-0.935	
30	2	1.180	-1.333		0.737	-0.737	
31	2	1.305	-1.322		0.738	-0.738	
32	2	1.080	-1.085		0.760	-0.760	
33	2	0.959	-1.141		0.690	-0.690	
34	2	1.090	-0.931		0.688	-0.688	
35	2	1.070	-0.932		0.593	-0.593	
36	4	0.799	-0.323		1.197	0.785	-0.488	-1.494

Table N7. Mathematics Grade 3 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3
1	1	1.125	1.350	0.155			
2	1	1.044	-0.219	0.145			
3	1	1.673	0.532	0.116			
4	1	0.712	-1.597	0.050			
5	1	1.204	0.053	0.155			
6	1	1.810	0.894	0.161			
7	1	1.116	-2.430	0.039			
8	1	1.251	-0.261	0.097			
9	1	1.376	-0.764	0.195			
10	1	1.170	-0.814	0.190			
11	1	1.899	0.719	0.129			
12	1	1.638	-0.903	0.196			

13	1	0.980	-0.134	0.190			
14	1	1.222	-0.598	0.218			
15	1	0.918	-1.032	0.108			
16	1	1.233	-0.363	0.164			
17	1	1.356	-0.219	0.158			
18	1	1.167	0.009	0.227			
19	1	0.898	1.016	0.327			
20	1	0.912	-0.586	0.273			
21	1	1.314	-0.157	0.204			
22	1	1.267	0.166	0.154			
23	1	1.120	-2.013	0.020			
24	1	1.461	0.380	0.108			
25	1	1.776	1.461	0.331			
26	1	1.124	0.601	0.267			
27	1	1.715	-0.581	0.182			
28	2	0.650	1.205		0.062	-0.062	
29	2	0.701	0.382		-0.336	0.336	
30	2	0.898	-0.315		0.427	-0.427	
31	2	0.685	-0.367		0.089	-0.089	
32	2	0.441	0.343		0.494	-0.494	
33	2	0.937	-0.122		-0.094	0.094	
34	3	0.643	-0.156		0.236	-0.105	-0.131

Table N8. Mathematics Grade 4 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3
1	1	1.066	0.033	0.222			
2	1	0.813	0.407	0.232			
3	1	1.603	0.560	0.163			
4	1	1.277	-0.283	0.419			
5	1	1.487	0.149	0.145			
6	1	1.010	0.583	0.186			
7	1	1.399	0.035	0.250			
8	1	0.978	-0.461	0.208			
9	1	1.474	-0.261	0.209			
10	1	1.161	-1.103	0.105			
11	1	1.159	-0.659	0.163			
12	1	1.418	0.531	0.099			
13	1	0.999	-0.510	0.248			
14	1	1.645	-0.181	0.128			
15	1	1.222	0.497	0.135			
16	1	1.273	0.651	0.221			
17	1	1.047	0.012	0.286			
18	1	1.164	-0.722	0.155			
19	1	1.405	-0.137	0.293			
20	1	1.173	-0.479	0.071			
21	1	1.094	-1.630	0.015			
22	1	1.049	-0.258	0.209			

Item	Max Pts	a	b	c	d1	d2	d3
23	1	0.867	-1.351	0.022			
24	1	0.906	0.296	0.078			
25	1	1.100	-0.206	0.376			
26	1	1.562	0.240	0.213			
27	1	1.225	0.227	0.170			
28	1	1.245	-1.647	0.098			
29	1	0.700	-0.616	0.129			
30	1	1.340	0.265	0.213			
31	1	1.330	-0.011	0.475			
32	2	0.672	0.941		-0.752	0.752	
33	2	1.018	0.364		0.127	-0.127	
34	2	0.644	-0.674		-0.899	0.899	
35	2	0.967	-0.137		0.381	-0.381	
36	2	0.738	0.005		0.493	-0.493	
37	2	0.522	0.685		0.266	-0.266	
38	3	0.719	0.586		-0.591	0.046	0.545

Table N9. Mathematics Grade 5 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3
1	1	1.403	0.270	0.246			
2	1	1.647	0.097	0.237			
3	1	1.815	-0.105	0.146			
4	1	1.690	-0.146	0.074			
5	1	1.266	0.161	0.224			
6	1	1.462	-0.950	0.173			
7	1	1.142	-0.366	0.186			
8	1	0.773	0.715	0.268			
9	1	1.766	0.564	0.157			
10	1	1.282	-0.343	0.203			
11	1	1.333	0.416	0.199			
12	1	1.238	0.589	0.147			
13	1	0.900	0.372	0.259			
14	1	1.280	1.123	0.331			
15	1	1.959	0.544	0.061			
16	1	1.565	0.643	0.115			
17	1	0.877	-0.447	0.452			
18	1	0.975	0.764	0.231			
19	1	0.997	-1.721	0.048			
20	1	1.056	-0.489	0.223			
21	1	0.991	0.416	0.237			
22	1	0.864	-1.014	0.141			
23	1	1.158	-0.547	0.191			
24	1	1.368	0.433	0.165			
25	1	0.511	-0.571	0.044			
26	1	1.540	0.374	0.225			
27	1	1.496	-0.021	0.216			

Item	Max Pts	a	b	c	d1	d2	d3
28	1	1.616	-0.052	0.263			
29	1	0.985	0.743	0.206			
30	1	1.178	-0.081	0.142			
31	1	1.496	0.040	0.138			
32	2	0.777	0.868		0.559	-0.559	
33	2	1.358	0.237		0.179	-0.179	
34	2	0.753	0.897		-1.078	1.078	
35	2	0.870	-0.017		0.311	-0.311	
36	2	0.838	0.574		-0.360	0.360	
37	2	0.749	-0.176		0.150	-0.150	
38	3	0.806	0.217		0.114	0.142	-0.256

Table N10. Mathematics Grade 6 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3
1	1	0.972	-0.560	0.179			
2	1	1.596	0.311	0.212			
3	1	0.936	0.300	0.181			
4	1	1.323	-0.104	0.273			
5	1	1.269	0.736	0.139			
6	1	1.705	0.550	0.151			
7	1	1.122	0.346	0.177			
8	1	0.963	0.765	0.328			
9	1	1.099	0.557	0.146			
10	1	1.053	-0.132	0.302			
11	1	1.400	-0.210	0.109			
12	1	1.025	-0.162	0.093			
13	1	1.614	0.651	0.206			
14	1	1.119	-1.053	0.043			
15	1	1.672	0.665	0.354			
16	1	1.223	0.191	0.305			
17	1	1.224	1.110	0.161			
18	1	1.561	0.358	0.252			
19	1	1.874	0.244	0.144			
20	1	1.325	0.820	0.183			
21	1	1.334	-0.343	0.245			
22	1	0.901	0.709	0.231			
23	1	1.494	0.266	0.224			
24	1	1.245	-0.268	0.292			
25	1	1.127	0.444	0.306			
26	1	0.680	-0.020	0.097			
27	1	0.940	1.109	0.199			
28	1	1.530	0.478	0.307			
29	1	1.255	0.025	0.222			
30	1	1.078	0.530	0.206			
31	1	1.506	-0.187	0.136			
32	2	0.771	0.578		0.572	-0.572	

Item		Max Pts	a	b	c	d1	d2	d3
33		2	0.850	-0.066		0.421	-0.421	
34		2	0.665	0.144		-0.111	0.111	
35		2	0.704	0.375		-0.218	0.218	
36		2	0.983	1.036		0.191	-0.191	
37		2	0.928	1.383		-0.869	0.869	
38		2	0.725	-0.145		-0.228	0.228	
39		3	0.581	-0.094		-0.956	1.287	-0.331

Table N11. Mathematics Grade 7 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3
1	1	0.820	0.068	0.166			
2	1	0.912	0.579	0.252			
3	1	0.950	-0.609	0.254			
4	1	1.395	0.465	0.302			
5	1	1.896	-0.121	0.307			
6	1	0.993	1.332	0.365			
7	1	0.824	0.711	0.284			
8	1	0.835	1.215	0.211			
9	1	0.946	-0.564	0.176			
10	1	1.450	0.371	0.160			
11	1	1.234	0.303	0.218			
12	1	1.551	0.128	0.314			
13	1	1.387	1.197	0.204			
14	1	0.793	-0.771	0.385			
15	1	1.026	0.295	0.284			
16	1	1.237	0.228	0.279			
17	1	1.085	0.248	0.083			
18	1	1.006	0.333	0.222			
19	1	0.971	-0.239	0.173			
20	1	0.929	-0.259	0.223			
21	1	1.306	0.884	0.342			
22	1	1.033	0.157	0.355			
23	1	0.881	0.067	0.276			
24	1	1.050	-0.364	0.279			
25	1	1.261	0.218	0.340			
26	1	2.209	-0.040	0.147			
27	1	1.448	-0.004	0.230			
28	1	1.368	1.042	0.272			
29	1	1.689	-0.407	0.280			
30	1	1.109	0.330	0.151			
31	1	1.381	0.303	0.150			
32	1	1.233	0.916	0.383			
33	1	1.717	1.082	0.271			
34	2	1.234	0.181		-0.097	0.097	
35	2	0.623	0.221		0.118	-0.118	
36	2	0.920	0.459		0.263	-0.263	

Item	Max Pts	a	b	c	d1	d2	d3
37	2	0.646	-1.183		-1.864	1.864	
38	2	0.868	-0.541		-0.049	0.049	
39	2	1.014	0.345		-1.052	1.052	
40	2	0.431	-0.184		-2.177	2.177	
41	3	1.016	0.681		0.500	-0.360	-0.140

Table N12. Mathematics Grade 8 OP Item Parameter Estimates

Item	Max Pts	a	b	c	d1	d2	d3
1	1	1.148	0.697	0.271			
2	1	1.176	0.521	0.210			
3	1	0.414	-0.939	0.064			
4	1	1.610	0.506	0.297			
5	1	0.612	-0.067	0.164			
6	1	1.010	-1.200	0.131			
7	1	1.343	0.239	0.349			
8	1	1.363	0.321	0.331			
9	1	1.314	1.742	0.077			
10	1	1.645	1.064	0.164			
11	1	1.158	-0.497	0.243			
12	1	1.379	0.481	0.236			
13	1	1.302	0.023	0.204			
14	1	1.301	0.571	0.218			
15	1	0.909	0.246	0.157			
16	1	0.680	0.710	0.112			
17	1	0.890	0.507	0.234			
18	1	1.802	0.614	0.248			
19	1	1.292	0.146	0.20			
20	1	1.444	0.747	0.214			
21	1	1.408	0.191	0.335			
22	1	0.833	-0.344	0.175			
23	1	0.955	0.264	0.306			
24	1	1.020	0.019	0.296			
25	1	0.912	-0.725	0.126			
26	1	1.471	0.441	0.226			
27	1	0.876	-0.437	0.246			
28	1	1.273	0.476	0.166			
29	1	1.172	0.618	0.216			
30	1	1.007	-0.191	0.226			
31	1	0.604	-0.127	0.048			
32	1	1.727	0.666	0.257			
33	1	1.059	0.538	0.160			
34	2	0.870	0.468		0.246	-0.246	
35	2	0.982	1.136		-0.453	0.453	
36	2	0.660	0.682		-0.758	0.758	
37	2	1.158	0.826		0.165	-0.165	
38	2	1.000	0.937		0.026	-0.026	

Item	Max Pts	a	b	c	d1	d2	d3
39	2	1.048	0.686		-0.061	0.061	
40	2	0.860	0.295		0.060	-0.060	
41	3	0.514	1.269		0.548	-0.274	-0.274

Appendix O: 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

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

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 $C_{h-1} \leq X < C_h$ 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:

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

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

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

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

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, $x \in L_h$ ($h = 1, 2, \ldots, H$), 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

$\quad w$ 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 P: Raw-to-Scale Score and Scale Score Frequency Tables

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

Table P1. PBT ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	535	25
1	540	20
2	544	16
3	549	13
4	553	11
5	558	9
6	563	7
7	567	7
8	570	6
9	573	6
10	575	6
11	578	5
12	580	5
13	583	5
14	585	5
15	587	5
16	590	5
17	592	5
18	594	5
19	597	5
20	599	5
21	602	5
22	604	5
23	607	5
24	609	5
25	612	5
26	615	6
27	618	6
28	621	6
29	624	7
30	629	7
31	634	8
32	640	10
33	652	14
34	656	17

Table P2. PBT ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	528	23
1	533	20
2	537	17
3	542	14
4	546	13
5	551	11
6	556	9
7	560	8
8	564	7
9	568	7
10	571	6
11	573	6
12	576	6
13	578	6
14	581	6
15	584	6
16	585	6
17	588	6
18	590	6
19	592	6
20	594	6
21	597	6
22	599	6
23	603	6
24	604	6
25	607	6
26	610	6
27	613	6
28	619	7
29	620	7
30	624	8
31	630	9
32	637	11
33	649	16
34	654	18

Table P3. PBT ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	518	35
1	523	30
2	527	26
3	532	22
4	536	19

Raw Score	Scale Score	Standard Error
5	541	16
6	545	14
7	550	12
8	555	10
9	559	9
10	563	8
11	567	7
12	570	7
13	573	6
14	575	6
15	578	6
16	580	6
17	582	6
18	584	5
19	586	5
20	588	5
21	590	5
22	592	5
23	594	5
24	595	5
25	597	5
26	599	5
27	600	5
28	602	5
29	604	5
30	605	5
31	607	5
32	609	5
33	611	5
34	613	5
35	615	5
36	618	5
37	622	6
38	623	6
39	626	7
40	630	7
41	635	8
42	642	10
43	654	16
44	658	19

Table P4. PBT ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	517	31
1	521	27
2	526	23
3	530	20
4	535	17
5	539	15
6	544	13
7	549	11
8	553	9
9	558	8
10	561	8
11	564	7
12	567	7
13	570	6
14	572	6
15	574	6
16	576	6
17	579	6
18	581	5
19	582	5
20	584	5
21	586	5
22	588	5
23	590	5
24	592	5
25	594	5
26	596	5
27	598	5
28	600	5
29	602	6
30	604	6
31	606	6
32	608	6
33	611	6
34	614	6
35	616	7
36	619	7
37	622	7
38	626	8
39	631	9
40	636	10
41	643	12
42	652	15
43	657	17

Raw Score	Scale Score	Standard Error
44	662	19

Table P5. PBT ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	518	43
1	523	36
2	528	30
3	532	24
4	537	20
5	541	17
6	546	14
7	550	11
8	555	9
9	560	8
10	564	7
11	567	6
12	570	6
13	572	6
14	574	5
15	576	5
16	578	5
17	580	5
18	582	5
19	584	5
20	585	5
21	587	4
22	589	4
23	591	4
24	592	4
25	593	4
26	595	4
27	597	4
28	598	4
29	600	4
30	602	4
31	603	5
32	605	5
33	607	5
34	609	5
35	611	5
36	613	5
37	615	6
38	618	6
39	623	6

Raw Score	Scale Score	Standard Error
40	624	7
41	628	8
42	632	9
43	639	10
44	647	13
45	652	15
46	656	17

Table P6. PBT ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	502	58
1	507	48
2	512	40
3	516	33
4	521	27
5	525	22
6	530	18
7	535	15
8	539	13
9	544	11
10	548	9
11	553	8
12	557	7
13	561	7
14	564	6
15	567	6
16	569	6
17	572	6
18	574	6
19	576	6
20	578	5
21	580	5
22	582	5
23	584	5
24	586	5
25	587	5
26	589	5
27	591	5
28	592	5
29	594	5
30	596	5
31	597	5
32	599	5
33	601	5

Raw Score	Scale Score	Standard Error
34	603	5
35	605	5
36	607	5
37	609	6
38	612	6
39	614	6
40	617	7
41	621	7
42	625	8
43	631	9
44	638	12
45	652	18
46	657	21

Table P7. PBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	522	33
1	526	27
2	531	21
3	536	18
4	540	15
5	545	13
6	550	12
7	554	11
8	561	10
9	565	8
10	569	8
11	572	7
12	575	6
13	577	6
14	579	5
15	581	5
16	583	5
17	585	5
18	587	5
19	588	4
20	590	4
21	591	4
22	593	4
23	594	4
24	596	4
25	597	4
26	600	4
27	601	4

Raw Score	Scale Score	Standard Error
28	602	4
29	604	4
30	605	4
31	607	4
32	609	5
33	611	5
34	613	5
35	615	5
36	618	5
37	621	6
38	624	6
39	628	7
40	634	9
41	645	14
42	649	17

Table P8. PBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	517	51
1	522	42
2	526	34
3	531	28
4	536	23
5	540	19
6	545	15
7	549	13
8	554	11
9	559	10
10	563	9
11	567	8
12	570	7
13	573	7
14	576	6
15	578	6
16	580	6
17	582	5
18	584	5
19	585	5
20	588	5
21	589	5
22	590	5
23	592	5
24	593	4
25	595	4

Raw Score	Scale Score	Standard Error
26	596	4
27	597	4
28	599	4
29	600	4
30	602	4
31	603	4
32	604	4
33	606	4
34	607	4
35	609	4
36	611	5
37	612	5
38	614	5
39	616	5
40	619	6
41	621	6
42	624	7
43	629	8
44	634	10
45	645	15
46	649	18

Table P9. PBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	523	50
1	528	41
2	532	34
3	537	28
4	542	24
5	546	20
6	551	17
7	555	15
8	560	13
9	566	10
10	571	8
11	574	7
12	577	7
13	580	6
14	582	6
15	584	5
16	586	5
17	588	5
18	589	4
19	591	4

Raw Score	Scale Score	Standard Error
20	592	4
21	593	4
22	595	4
23	596	4
24	597	4
25	598	4
26	599	4
27	601	4
28	602	4
29	603	4
30	604	4
31	605	4
32	607	4
33	608	4
34	609	4
35	611	4
36	612	4
37	614	4
38	616	4
39	617	5
40	620	5
41	622	6
42	625	6
43	629	7
44	634	9
45	644	13
46	648	16

Table P10. PBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	529	60
1	534	50
2	538	42
3	543	34
4	547	28
5	552	23
6	556	19
7	561	15
8	565	12
9	571	10
10	575	8
11	579	7
12	581	6
13	584	6

Raw Score	Scale Score	Standard Error
14	586	5
15	587	5
16	589	5
17	590	5
18	592	4
19	593	4
20	595	4
21	596	4
22	597	4
23	598	4
24	599	4
25	601	4
26	602	4
27	603	4
28	604	4
29	605	4
30	606	4
31	607	4
32	609	4
33	610	4
34	611	4
35	612	4
36	614	4
37	616	4
38	617	4
39	618	4
40	620	4
41	622	5
42	624	5
43	626	5
44	629	6
45	632	7
46	637	8
47	645	12
48	650	15

Table P11. PBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	518	95
1	522	80
2	527	67
3	532	56
4	536	46
5	541	37

Raw Score	Scale Score	Standard Error
6	546	30
7	550	25
8	555	20
9	559	16
10	564	13
11	570	10
12	574	9
13	577	8
14	580	7
15	582	6
16	585	6
17	586	5
18	588	5
19	590	5
20	591	4
21	593	4
22	594	4
23	595	4
24	596	4
25	597	4
26	598	4
27	600	3
28	601	3
29	602	3
30	603	3
31	604	3
32	605	3
33	606	3
34	607	3
35	608	4
36	609	4
37	611	4
38	612	4
39	613	4
40	615	4
41	616	4
42	618	4
43	619	4
44	621	5
45	624	5
46	626	6
47	630	6
48	634	8
49	642	11
50	647	14

Table P12. PBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	524	68
1	529	58
2	533	50
3	538	42
4	542	35
5	547	29
6	551	24
7	556	20
8	560	16
9	565	14
10	571	11
11	576	9
12	580	8
13	583	7
14	585	6
15	588	6
16	590	6
17	591	5
18	593	5
19	595	5
20	596	4
21	598	4
22	599	4
23	600	4
24	601	4
25	602	4
26	604	4
27	605	4
28	606	4
29	607	3
30	608	3
31	609	3
32	610	3
33	611	3
34	612	3
35	613	4
36	615	4
37	616	4
38	617	4
39	618	4
40	620	4
41	622	4
42	623	4
43	625	5

Raw Score	Scale Score	Standard Error
44	627	5
45	629	5
46	632	6
47	636	7
48	642	9
49	651	14
50	656	17

Table P13. CBT ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	537	23
1	542	18
2	546	15
3	551	12
4	555	10
5	560	8
6	565	7
7	569	6
8	572	6
9	575	6
10	577	5
11	580	5
12	582	5
13	585	5
14	587	5
15	589	5
16	592	5
17	594	5
18	596	5
19	599	5
20	601	5
21	604	5
22	606	5
23	609	5
24	611	5
25	614	6
26	617	6
27	620	6
28	623	6
29	626	7
30	631	8
31	636	9

Raw Score	Scale Score	Standard Error
32	642	10
33	654	16
34	656	17

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

Table P14. CBT ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	530	22
1	535	18
2	539	16
3	544	14
4	548	12
5	553	10
6	558	9
7	562	8
8	566	7
9	570	7
10	573	6
11	575	6
12	578	6
13	580	6
14	583	6
15	586	6
16	587	6
17	590	6
18	592	6
19	594	6
20	596	6
21	599	6
22	601	6
23	605	6
24	606	6
25	609	6
26	612	6
27	615	7
28	621	8
29	622	8
30	626	8
31	632	10
32	639	12
33	651	17
34	654	18

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

Table P15. CBT ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	520	33
1	525	28
2	529	24
3	534	20
4	538	18
5	543	15
6	547	13
7	552	11
8	557	9
9	561	8
10	565	8
11	569	7
12	572	7
13	575	6
14	577	6
15	580	6
16	582	6
17	584	5
18	586	5
19	588	5
20	590	5
21	592	5
22	594	5
23	596	5
24	597	5
25	599	5
26	601	5
27	602	5
28	604	5
29	606	5
30	607	5
31	609	5
32	611	5
33	613	5
34	615	5
35	617	5
36	620	6
37	624	6
38	625	6
39	628	7
40	632	8
41	637	9

Raw Score	Scale Score	Standard Error
42	644	11
43	656	17
44	658	19

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

Table P16. CBT ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	519	29
1	523	25
2	528	21
3	532	19
4	537	16
5	541	14
6	546	12
7	551	10
8	555	9
9	560	8
10	563	7
11	566	7
12	569	6
13	572	6
14	574	6
15	576	6
16	578	6
17	581	5
18	583	5
19	584	5
20	586	5
21	588	5
22	590	5
23	592	5
24	594	5
25	596	5
26	598	5
27	600	5
28	602	6
29	604	6
30	606	6
31	608	6
32	610	6
33	613	6
34	616	7

Raw Score	Scale Score	Standard Error
35	618	7
36	621	7
37	624	8
38	628	8
39	633	9
40	638	10
41	645	12
42	654	16
43	659	18
44	662	19

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

Table P17. CBT ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	520	40
1	525	33
2	530	27
3	534	23
4	539	18
5	543	15
6	548	12
7	552	11
8	557	9
9	562	7
10	566	6
11	569	6
12	572	6
13	574	5
14	576	5
15	578	5
16	580	5
17	582	5
18	584	5
19	586	4
20	587	4
21	589	4
22	591	4
23	593	4
24	594	4
25	595	4
26	597	4
27	599	4

Raw Score	Scale Score	Standard Error
28	600	4
29	602	5
30	604	5
31	605	5
32	607	5
33	609	5
34	611	5
35	613	5
36	615	6
37	617	6
38	620	6
39	625	7
40	626	7
41	630	8
42	634	9
43	641	11
44	649	14
45	654	16
46	656	17

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

Table P18. CBT ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	504	55
1	509	45
2	514	36
3	518	31
4	523	25
5	527	21
6	532	17
7	537	14
8	541	12
9	546	10
10	550	9
11	555	7
12	559	7
13	563	6
14	566	6
15	569	6
16	571	6
17	574	6
18	576	6

Raw Score	Scale Score	Standard Error
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	594	5
29	596	5
30	598	5
31	599	5
32	601	5
33	603	5
34	605	5
35	607	5
36	609	6
37	611	6
38	614	6
39	616	6
40	619	7
41	623	8
42	627	9
43	633	10
44	640	13
45	654	19
46	657	21

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

Table P19. CBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	524	30
1	528	25
2	533	20
3	538	16
4	542	14
5	547	13
6	552	11
7	556	10
8	563	9
9	567	8

Raw Score	Scale Score	Standard Error
10	571	7
11	574	6
12	577	6
13	579	5
14	581	5
15	583	5
16	585	5
17	587	5
18	589	4
19	590	4
20	592	4
21	593	4
22	595	4
23	596	4
24	598	4
25	599	4
26	602	4
27	603	4
28	604	4
29	606	4
30	607	4
31	609	5
32	611	5
33	613	5
34	615	5
35	617	5
36	620	6
37	623	6
38	626	7
39	630	8
40	636	10
41	647	16
42	649	17

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

Table P20. CBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	519	47
1	524	38
2	528	32
3	533	25
4	538	20
5	542	17
6	547	14
7	551	12
8	556	10
9	561	9
10	565	8
11	569	7
12	572	7
13	575	6
14	578	6
15	580	6
16	582	5
17	584	5
18	586	5
19	587	5
20	590	5
21	591	5
22	592	4
23	594	4
24	595	4
25	597	4
26	598	4
27	599	4
28	601	4
29	602	4
30	604	4
31	605	4
32	606	4
33	608	4
34	609	4
35	611	5
36	613	5
37	614	5
38	616	5
39	618	5
40	621	6
41	623	6

Raw Score	Scale Score	Standard Error
42	626	7
43	631	9
44	636	11
45	647	16
46	649	18

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

Table P21. CBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	525	46
1	530	38
2	534	32
3	539	26
4	544	22
5	548	19
6	553	16
7	557	14
8	562	12
9	568	9
10	573	8
11	576	7
12	579	6
13	582	6
14	584	5
15	586	5
16	588	5
17	590	4
18	591	4
19	593	4
20	594	4
21	595	4
22	597	4
23	598	4
24	599	4
25	600	4
26	601	4
27	603	4
28	604	4
29	605	4
30	606	4
31	607	4
32	609	4

Raw Score	Scale Score	Standard Error
33	610	4
34	611	4
35	613	4
36	614	4
37	616	5
38	618	5
39	619	5
40	622	6
41	624	6
42	627	7
43	631	8
44	636	10
45	646	15
46	648	16

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

Table P22. CBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	531	55
1	536	45
2	540	38
3	545	31
4	549	26
5	554	21
6	558	17
7	563	14
8	567	12
9	573	9
10	577	8
11	581	6
12	583	6
13	586	5
14	588	5
15	589	5
16	591	5
17	592	4
18	594	4
19	595	4
20	597	4
21	598	4
22	599	4
23	600	4

Raw Score	Scale Score	Standard Error
24	601	4
25	603	4
26	604	4
27	605	4
28	606	4
29	607	4
30	608	4
31	609	4
32	611	4
33	612	4
34	613	4
35	614	4
36	616	4
37	618	4
38	619	4
39	620	4
40	622	5
41	624	5
42	626	5
43	628	6
44	631	6
45	634	7
46	639	9
47	647	13
48	650	15

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

Table P23. CBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	520	88
1	524	75
2	529	62
3	534	50
4	538	42
5	543	34
6	548	27
7	552	22
8	557	18
9	561	15
10	566	12
11	572	9
12	576	8

Raw Score	Scale Score	Standard Error
13	579	7
14	582	6
15	584	6
16	587	5
17	588	5
18	590	5
19	592	4
20	593	4
21	595	4
22	596	4
23	597	4
24	598	4
25	599	3
26	600	3
27	602	3
28	603	3
29	604	3
30	605	3
31	606	3
32	607	3
33	608	4
34	609	4
35	610	4
36	611	4
37	613	4
38	614	4
39	615	4
40	617	4
41	618	4
42	620	4
43	621	5
44	623	5
45	626	5
46	628	6
47	632	7
48	636	8
49	644	12
50	647	14

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

Table P24. CBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	526	64
1	531	54
2	535	47
3	540	39
4	544	33
5	549	27
6	553	23
7	558	18
8	562	15
9	567	13
10	573	10
11	578	8
12	582	7
13	585	7
14	587	6
15	590	5
16	592	5
17	593	5
18	595	5
19	597	4
20	598	4
21	600	4
22	601	4
23	602	4
24	603	4
25	604	4
26	606	4
27	607	3
28	608	3
29	609	3
30	610	3
31	611	3
32	612	3
33	613	4
34	614	4
35	615	4
36	617	4
37	618	4
38	619	4
39	620	4
40	622	4
41	624	5

Raw Score	Scale Score	Standard Error
42	625	5
43	627	5
44	629	5
45	631	6
46	634	7
47	638	8
48	644	10
49	653	15
50	656	17

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

Table P25. ELA Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
535	54	0.03	54	0.0
537	16	0.01	70	0.0
540	186	0.11	256	0.2
542	31	0.02	287	0.2
544	534	0.32	821	0.5
546	100	0.06	921	0.6
549	948	0.57	1,869	1.1
551	223	0.13	2,092	1.3
553	1,492	0.90	3,584	2.2
555	363	0.22	3,947	2.4
558	1,782	1.08	5,729	3.5
560	493	0.30	6,222	3.8
563	2,034	1.23	8,256	5.0
565	563	0.34	8,819	5.3
567	2,277	1.38	11,096	6.7
569	624	0.38	11,720	7.1
570	2,402	1.45	14,122	8.5
572	735	0.44	14,857	9.0
573	2,710	1.64	17,567	10.6
575	3,898	2.36	21,465	13.0
577	933	0.56	22,398	13.6
578	3,434	2.08	25,832	15.6
580	4,805	2.91	30,637	18.5
582	1,293	0.78	31,930	19.3
583	4,194	2.54	36,124	21.9

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
585	6,007	3.64	42,131	25.5
587	6,407	3.88	48,538	29.4
589	1,688	1.02	50,226	30.4
590	5,207	3.15	55,433	33.6
592	7,291	4.41	62,724	38.0
594	7,690	4.65	70,414	42.6
596	1,897	1.15	72,311	43.8
597	6,290	3.81	78,601	47.6
599	8,448	5.11	87,049	52.7
601	1,881	1.14	88,930	53.8
602	6,647	4.02	95,577	57.9
604	8,811	5.33	104,388	63.2
606	1,810	1.10	106,198	64.3
607	6,652	4.03	112,850	68.3
609	8,318	5.03	121,168	73.3
611	1,531	0.93	122,699	74.3
612	6,344	3.84	129,043	78.1
614	1,364	0.83	130,407	78.9
615	6,013	3.64	136,420	82.6
617	1,154	0.70	137,574	83.3
618	5,623	3.40	143,197	86.7
620	964	0.58	144,161	87.3
621	5,259	3.18	149,420	90.4
623	773	0.47	150,193	90.9
624	4,622	2.80	154,815	93.7
626	569	0.34	155,384	94.1
629	3,726	2.26	159,110	96.3
631	378	0.23	159,488	96.5
634	2,816	1.70	162,304	98.2
636	231	0.14	162,535	98.4
640	1,679	1.02	164,214	99.4
642	75	0.05	164,289	99.4
652	697	0.42	164,986	99.9
654	28	0.02	165,014	99.9
656	195	0.12	165,209	100.0

Table P26. ELA Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
528	39	0.02	39	0.0
530	13	0.01	52	0.0
533	140	0.08	192	0.1
535	27	0.02	219	0.1
537	463	0.27	682	0.4
539	72	0.04	754	0.4
542	946	0.56	1,700	1.0
544	154	0.09	1,854	1.1
546	1,475	0.87	3,329	2.0
548	258	0.15	3,587	2.1
551	1,775	1.05	5,362	3.2
553	403	0.24	5,765	3.4
556	2,008	1.19	7,773	4.6
558	536	0.32	8,309	4.9
560	2,306	1.37	10,615	6.3
562	660	0.39	11,275	6.7
564	2,577	1.53	13,852	8.2
566	788	0.47	14,640	8.7
568	2,754	1.63	17,394	10.3
570	855	0.51	18,249	10.8
571	3,041	1.80	21,290	12.6
573	4,320	2.56	25,610	15.2
575	1,128	0.67	26,738	15.8
576	3,598	2.13	30,336	18.0
578	4,937	2.93	35,273	20.9
580	1,241	0.74	36,514	21.6
581	4,035	2.39	40,549	24.0
583	1,252	0.74	41,801	24.8
584	4,422	2.63	46,234	27.4
585	4,690	2.78	50,924	30.2
586	1,374	0.81	52,298	31.0
587	1,506	0.89	53,804	31.9
588	5,060	3.00	58,864	34.9
590	6,835	4.05	65,699	38.9
592	7,111	4.21	72,810	43.2
594	7,605	4.51	80,415	47.7

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
596	1,793	1.06	82,208	48.7
597	6,199	3.67	88,407	52.4
599	8,445	5.01	96,852	57.4
601	1,908	1.13	98,760	58.5
603	6,684	3.96	105,444	62.5
604	6,737	3.99	112,181	66.5
605	1,909	1.13	114,090	67.6
606	1,835	1.09	115,925	68.7
607	6,726	3.99	122,651	72.7
609	1,679	1.00	124,330	73.7
610	6,751	4.00	131,081	77.7
612	1,730	1.03	132,811	78.7
613	6,364	3.77	139,175	82.5
615	1,504	0.89	140,679	83.4
619	6,139	3.64	146,818	87.0
620	5,336	3.16	152,154	90.2
621	1,340	0.79	153,494	91.0
622	1,125	0.67	154,619	91.6
624	4,597	2.72	159,216	94.4
626	827	0.49	160,043	94.9
630	3,547	2.10	163,590	97.0
632	469	0.28	164,059	97.2
637	2,454	1.45	166,513	98.7
639	262	0.16	166,775	98.8
649	1,371	0.81	168,146	99.7
651	80	0.05	168,226	99.7
654	499	0.30	168,725	100.0

Table P27. ELA Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
518	15	0.01	15	0.0
520	2	0.00	17	0.0
523	27	0.02	44	0.0
525	1	0.00	45	0.0
527	47	0.03	92	0.1
529	5	0.00	97	0.1

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
532	153	0.09	250	0.2
534	23	0.01	273	0.2
536	284	0.17	557	0.3
538	47	0.03	604	0.4
541	533	0.32	1,137	0.7
543	95	0.06	1,232	0.7
545	817	0.50	2,049	1.2
547	173	0.10	2,222	1.3
550	1,061	0.64	3,283	2.0
552	252	0.15	3,535	2.1
555	1,243	0.75	4,778	2.9
557	278	0.17	5,056	3.1
559	1,364	0.83	6,420	3.9
561	397	0.24	6,817	4.1
563	1,474	0.89	8,291	5.0
565	482	0.29	8,773	5.3
567	1,599	0.97	10,372	6.3
569	505	0.31	10,877	6.6
570	1,697	1.03	12,574	7.6
572	625	0.38	13,199	8.0
573	1,913	1.16	15,112	9.2
575	2,744	1.66	17,856	10.8
577	773	0.47	18,629	11.3
578	2,332	1.41	20,961	12.7
580	3,192	1.93	24,153	14.6
582	3,488	2.11	27,641	16.7
584	3,690	2.24	31,331	19.0
586	3,991	2.42	35,322	21.4
588	4,445	2.69	39,767	24.1
590	4,586	2.78	44,353	26.9
592	4,781	2.90	49,134	29.8
594	5,192	3.15	54,326	32.9
595	3,938	2.39	58,264	35.3
596	1,295	0.78	59,559	36.1
597	5,462	3.31	65,021	39.4
599	5,714	3.46	70,735	42.9
600	4,541	2.75	75,276	45.6

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
601	1,373	0.83	76,649	46.4
602	6,199	3.76	82,848	50.2
604	6,518	3.95	89,366	54.2
605	5,175	3.14	94,541	57.3
606	1,511	0.92	96,052	58.2
607	6,883	4.17	102,935	62.4
609	6,998	4.24	109,933	66.6
611	7,150	4.33	117,083	70.9
613	6,994	4.24	124,077	75.2
615	7,088	4.30	131,165	79.5
617	1,331	0.81	132,496	80.3
618	5,332	3.23	137,828	83.5
620	1,284	0.78	139,112	84.3
622	5,265	3.19	144,377	87.5
623	4,700	2.85	149,077	90.3
624	1,190	0.72	150,267	91.1
625	984	0.60	151,251	91.7
626	4,021	2.44	155,272	94.1
628	779	0.47	156,051	94.6
630	3,263	1.98	159,314	96.5
632	567	0.34	159,881	96.9
635	2,364	1.43	162,245	98.3
637	358	0.22	162,603	98.5
642	1,406	0.85	164,009	99.4
644	154	0.09	164,163	99.5
654	615	0.37	164,778	99.9
656	64	0.04	164,842	99.9
658	182	0.11	165,024	100.0

Table P28. ELA Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
517	18	0.01	18	0.0
519	1	0.00	19	0.0
521	21	0.01	40	0.0
523	1	0.00	41	0.0
526	48	0.03	89	0.1
528	8	0.00	97	0.1

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
530	113	0.70	210	0.1
532	8	0.00	218	0.1
535	256	0.16	474	0.3
537	28	0.02	502	0.3
539	405	0.25	907	0.6
541	53	0.03	960	0.6
544	567	0.35	1,527	0.9
546	94	0.06	1,621	1.0
549	720	0.44	2,341	1.4
551	140	0.09	2,481	1.5
553	872	0.53	3,353	2.1
555	201	0.12	3,554	2.2
558	1,004	0.61	4,558	2.8
560	277	0.17	4,835	3.0
561	1,161	0.71	5,996	3.7
563	387	0.24	6,383	3.9
564	1,281	0.78	7,664	4.7
566	436	0.27	8,100	5.0
567	1,375	0.84	9,475	5.8
569	533	0.33	10,008	6.1
570	1,544	0.94	11,552	7.1
572	2,268	1.39	13,820	8.5
574	2,499	1.53	16,319	10.0
576	2,743	1.68	19,062	11.7
578	784	0.48	19,846	12.1
579	2,114	1.29	21,960	13.4
581	3,089	1.89	25,049	15.3
582	2,357	1.44	27,406	16.8
583	893	0.55	28,299	17.3
584	3,539	2.16	31,838	19.5
586	3,750	2.29	35,588	21.8
588	4,041	2.47	39,629	24.2
590	4,298	2.63	43,927	26.9
592	4,616	2.82	48,543	29.7
594	4,905	3.00	53,448	32.7
596	5,249	3.21	58,697	35.9
598	5,684	3.48	64,381	39.4

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
600	6,148	3.76	70,529	43.1
602	6,575	4.02	77,104	47.2
604	6,918	4.23	84,022	51.4
606	7,418	4.54	91,440	55.9
608	7,925	4.85	99,365	60.8
610	1,912	1.17	101,277	61.9
611	6,372	3.90	107,649	65.8
613	2,050	1.25	109,699	67.1
614	6,575	4.02	116,274	71.1
616	8,667	5.30	124,941	76.4
618	1,926	1.18	126,867	77.6
619	6,766	4.14	133,633	81.7
621	1,868	1.14	135,501	82.9
622	6,107	3.73	141,608	86.6
624	1,720	1.05	143,328	87.7
626	5,366	3.28	148,694	90.9
628	1,400	0.86	150,094	91.8
631	4,536	2.77	154,630	94.6
633	1,013	0.62	155,643	95.2
636	3,326	2.03	158,969	97.2
638	680	0.42	159,649	97.6
643	1,947	1.19	161,596	98.8
645	369	0.23	161,965	99.1
652	965	0.59	162,930	99.6
654	135	0.08	163,065	99.7
657	346	0.21	163,411	99.9
659	36	0.02	163,447	100.0
662	62	0.04	163,509	100.0

Table P29. ELA Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
518	15	0.01	15	0.0
520	4	0.00	19	0.0
523	16	0.01	35	0.0
525	1	0.00	36	0.0
528	30	0.02	66	0.0
530	5	0.00	71	0.0

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
532	72	0.05	143	0.1
534	20	0.01	163	0.1
537	158	0.10	321	0.2
539	28	0.02	349	0.2
541	288	0.18	637	0.4
543	63	0.04	700	0.4
546	459	0.29	1,159	0.7
548	117	0.07	1,276	0.8
550	614	0.38	1,890	1.2
552	170	0.11	2,060	1.3
555	735	0.46	2,795	1.7
557	224	0.14	3,019	1.9
560	808	0.51	3,827	2.4
562	290	0.18	4,117	2.6
564	955	0.60	5,072	3.2
566	346	0.22	5,418	3.4
567	1,032	0.65	6,450	4.0
569	414	0.26	6,864	4.3
570	1,107	0.69	7,971	5.0
572	1,764	1.10	9,735	6.1
574	1,904	1.19	11,639	7.3
576	2,092	1.31	13,731	8.6
578	2,210	1.38	15,941	10.0
580	2,351	1.47	18,292	11.4
582	2,664	1.67	20,956	13.1
584	2,715	1.70	23,671	14.8
585	2,211	1.38	25,882	16.2
586	849	0.53	26,731	16.7
587	3,067	1.92	29,798	18.7
589	3,223	2.02	33,021	20.7
591	3,557	2.23	36,578	22.9
592	2,832	1.77	39,410	24.7
593	4,019	2.52	43,429	27.2
594	1,101	0.69	44,530	27.9
595	4,179	2.62	48,709	30.5
597	4,680	2.93	53,389	33.4
598	3,650	2.28	57,039	35.7

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
599	1,279	0.80	58,318	36.5
600	5,299	3.32	63,617	39.8
602	5,772	3.61	69,389	43.4
603	4,692	2.94	74,081	46.4
604	1,485	0.93	75,566	47.3
605	6,601	4.13	82,167	51.4
607	6,742	4.22	88,909	55.7
609	7,256	4.54	96,165	60.2
611	7,529	4.71	103,694	64.9
613	7,784	4.87	111,478	69.8
615	7,977	4.99	119,455	74.8
617	1,747	1.09	121,202	75.9
618	6,342	3.97	127,544	79.8
620	1,562	0.98	129,106	80.8
623	6,167	3.86	135,273	84.7
624	5,593	3.50	140,866	88.2
625	1,371	0.86	142,237	89.0
626	1,146	0.72	143,383	89.7
628	4,961	3.11	148,344	92.9
630	878	0.55	149,222	93.4
632	4,034	2.53	153,256	95.9
634	653	0.41	153,909	96.3
639	2,760	1.73	156,669	98.1
641	383	0.24	157,052	98.3
647	1,657	1.04	158,709	99.3
649	179	0.11	158,888	99.5
652	674	0.42	159,562	99.9
654	45	0.03	159,607	99.9
656	155	0.10	159,762	100.0

Table P30. ELA Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
502	11	0.01	11	0.0
504	1	0.00	12	0.0
507	10	0.01	22	0.0
509	2	0.00	24	0.0
512	33	0.02	57	0.0

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
514	6	0.00	63	0.0
516	50	0.03	113	0.1
518	9	0.01	122	0.1
521	131	0.09	253	0.2
523	24	0.02	277	0.2
525	244	0.16	521	0.3
527	54	0.04	575	0.4
530	344	0.23	919	0.6
532	87	0.06	1,006	0.7
535	459	0.31	1,465	1.0
537	120	0.08	1,585	1.1
539	561	0.37	2,146	1.4
541	171	0.11	2,317	1.5
544	725	0.48	3,042	2.0
546	224	0.15	3,266	2.2
548	794	0.53	4,060	2.7
550	295	0.20	4,355	2.9
553	846	0.56	5,201	3.5
555	357	0.24	5,558	3.7
557	933	0.62	6,491	4.3
559	421	0.28	6,912	4.6
561	1,050	0.70	7,962	5.3
563	481	0.32	8,443	5.6
564	1,127	0.75	9,570	6.4
566	536	0.36	10,106	6.7
567	1,291	0.86	11,397	7.6
569	1,953	1.30	13,350	8.9
571	593	0.39	13,943	9.3
572	1,538	1.02	15,481	10.3
574	2,249	1.50	17,730	11.8
576	2,375	1.58	20,105	13.4
578	2,476	1.65	22,581	15.0
580	2,695	1.80	25,276	16.8
582	2,882	1.92	28,158	18.8
584	3,091	2.06	31,249	20.8
586	3,116	2.08	34,365	22.9
587	2,454	1.63	36,819	24.5

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
588	893	0.59	37,712	25.1
589	3,559	2.37	41,271	27.5
591	3,821	2.55	45,092	30.0
592	3,080	2.05	48,172	32.1
593	1,035	0.69	49,207	32.8
594	4,430	2.95	53,637	35.7
596	4,691	3.12	58,328	38.9
597	3,753	2.50	62,081	41.4
598	1,235	0.82	63,316	42.2
599	5,534	3.69	68,850	45.9
601	5,684	3.79	74,534	49.6
603	6,134	4.09	80,668	53.7
605	6,516	4.34	87,184	58.1
607	6,887	4.59	94,071	62.7
609	7,303	4.86	101,374	67.5
611	1,627	1.08	103,001	68.6
612	5,985	3.99	108,986	72.6
614	7,753	5.16	116,739	77.8
616	1,575	1.05	118,314	78.8
617	6,258	4.17	124,572	83.0
619	1,585	1.06	126,157	84.0
621	5,940	3.96	132,097	88.0
623	1,296	0.86	133,393	88.9
625	5,413	3.61	138,806	92.5
627	1,038	0.69	139,844	93.1
631	4,309	2.87	144,153	96.0
633	702	0.47	144,855	96.5
638	2,957	1.97	147,812	98.5
640	413	0.28	148,225	98.7
652	1,373	0.91	149,598	99.6
654	164	0.11	149,762	99.8
657	368	0.25	150,130	100.0

Table P31. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
522	11	0.01	11	0.0
524	2	0.00	13	0.0
526	27	0.02	40	0.0
528	8	0.00	48	0.0
531	93	0.06	141	0.1
533	37	0.02	178	0.1
536	300	0.18	478	0.3
538	124	0.07	602	0.4
540	588	0.35	1,190	0.7
542	199	0.12	1,389	0.8
545	1,091	0.66	2,480	1.5
547	339	0.20	2,819	1.7
550	1,655	0.99	4,474	2.7
552	505	0.30	4,979	3.0
554	2,153	1.29	7,132	4.3
556	633	0.38	7,765	4.7
561	2,636	1.58	10,401	6.2
563	709	0.43	11,110	6.7
565	2,884	1.73	13,994	8.4
567	777	0.47	14,771	8.9
569	3,023	1.82	17,794	10.7
571	839	0.50	18,633	11.2
572	3,043	1.83	21,676	13.0
574	863	0.52	22,539	13.5
575	3,121	1.88	25,660	15.4
577	4,089	2.46	29,749	17.9
579	4,317	2.59	34,066	20.5
581	4,396	2.64	38,462	23.1
583	4,362	2.62	42,824	25.7
585	4,548	2.73	47,372	28.5
587	4,556	2.74	51,928	31.2
588	3,715	2.23	55,643	33.4
589	998	0.60	56,641	34.0
590	4,701	2.82	61,342	36.9
591	3,874	2.33	65,216	39.2
592	1,109	0.67	66,325	39.8

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
593	4,862	2.92	71,187	42.8
594	3,895	2.34	75,082	45.1
595	1,040	0.62	76,122	45.7
596	5,164	3.10	81,286	48.8
597	4,056	2.44	85,342	51.3
598	1,062	0.64	86,404	51.9
599	1,012	0.61	87,416	52.5
600	4,155	2.50	91,571	55.0
601	4,243	2.55	95,814	57.6
602	5,444	3.27	101,258	60.8
603	1,057	0.64	102,315	61.5
604	5,339	3.21	107,654	64.7
605	4,569	2.75	112,223	67.4
606	1,088	0.65	113,311	68.1
607	5,545	3.33	118,856	71.4
609	5,618	3.38	124,474	74.8
611	5,595	3.36	130,069	78.1
613	5,670	3.41	135,739	81.6
615	5,453	3.28	141,192	84.8
617	765	0.46	141,957	85.3
618	4,675	2.81	146,632	88.1
620	709	0.43	147,341	88.5
621	4,358	2.62	151,699	91.1
623	581	0.35	152,280	91.5
624	4,088	2.46	156,368	93.9
626	423	0.25	156,791	94.2
628	3,557	2.14	160,348	96.3
630	359	0.22	160,707	96.6
634	2,747	1.65	163,454	98.2
636	227	0.14	163,681	98.3
645	1,812	1.09	165,493	99.4
647	121	0.07	165,614	99.5
649	832	0.50	166,446	100.0

Table P32. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
517	13	0.01	13	0.0
519	4	0.00	17	0.0
522	19	0.01	36	0.0
524	5	0.00	41	0.0
526	67	0.04	108	0.1
528	20	0.01	128	0.1
531	169	0.10	297	0.2
533	43	0.03	340	0.2
536	473	0.28	813	0.5
538	137	0.08	950	0.6
540	930	0.55	1,880	1.1
542	239	0.14	2,119	1.2
545	1,467	0.87	3,586	2.1
547	425	0.25	4,011	2.4
549	1,993	1.18	6,004	3.5
551	553	0.33	6,557	3.9
554	2,437	1.44	8,994	5.3
556	674	0.40	9,668	5.7
559	2,859	1.69	12,527	7.4
561	777	0.46	13,304	7.8
563	3,166	1.87	16,470	9.7
565	834	0.49	17,304	10.2
567	3,302	1.95	20,606	12.2
569	824	0.49	21,430	12.6
570	3,337	1.97	24,767	14.6
572	889	0.52	25,656	15.1
573	3,203	1.89	28,859	17.0
575	840	0.50	29,699	17.5
576	3,299	1.95	32,998	19.5
578	4,069	2.40	37,067	21.9
580	4,194	2.47	41,261	24.3
582	4,233	2.50	45,494	26.8
584	4,271	2.52	49,765	29.4
585	3,265	1.93	53,030	31.3
586	910	0.54	53,940	31.8
587	849	0.50	54,789	32.3

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
588	3,330	1.96	58,119	34.3
589	3,356	1.98	61,475	36.3
590	4,212	2.48	65,687	38.7
591	898	0.53	66,585	39.3
592	4,406	2.60	70,991	41.9
593	3,493	2.06	74,484	43.9
594	983	0.58	75,467	44.5
595	4,396	2.59	79,863	47.1
596	3,405	2.01	83,268	49.1
597	4,225	2.49	87,493	51.6
598	1,010	0.60	88,503	52.2
599	4,363	2.57	92,866	54.8
600	3,590	2.12	96,456	56.9
601	928	0.55	97,384	57.4
602	4,551	2.68	101,935	60.1
603	3,552	2.10	105,487	62.2
604	4,481	2.64	109,968	64.9
605	933	0.55	110,901	65.4
606	4,538	2.68	115,439	68.1
607	3,792	2.24	119,231	70.3
608	935	0.55	120,166	70.9
609	4,696	2.77	124,862	73.6
611	4,834	2.85	129,696	76.5
612	3,831	2.26	133,527	78.8
613	863	0.51	134,390	79.3
614	4,765	2.81	139,155	82.1
616	4,803	2.83	143,958	84.9
618	774	0.46	144,732	85.4
619	4,012	2.37	148,744	87.7
621	4,680	2.76	153,424	90.5
623	584	0.34	154,008	90.8
624	3,821	2.25	157,829	93.1
626	500	0.29	158,329	93.4
629	3,553	2.10	161,882	95.5
631	418	0.25	162,300	95.7
634	3,049	1.80	165,349	97.5
636	327	0.19	165,676	97.7

Scale			Cumulative	
Score	Freq.	Pct.	Freq.	Pct.
645	2,308	1.36	167,984	99.1
647	184	0.11	168,168	99.2
649	1,367	0.81	169,535	100.0

Table P33. Mathematics Grade 5 Scale Score Frequency Distribution

Scale			Cumulative	
Score	Freq.	Pct.	Freq.	Pct.
523	11	0.01	11	0.0
525	3	0.00	14	0.0
528	35	0.02	49	0.0
530	3	0.00	52	0.0
532	122	0.07	174	0.1
534	38	0.02	212	0.1
537	339	0.21	551	0.3
539	95	0.06	646	0.4
542	747	0.46	1,393	0.8
544	223	0.14	1,616	1.0
546	1,426	0.87	3,042	1.9
548	364	0.22	3,406	2.1
551	2,210	1.35	5,616	3.4
553	605	0.37	6,221	3.8
555	3,099	1.89	9,320	5.7
557	760	0.46	10,080	6.1
560	3,761	2.29	13,841	8.4
562	1,005	0.61	14,846	9.1
566	4,101	2.50	18,947	11.6
568	1,102	0.67	20,049	12.2
571	4,257	2.60	24,306	14.8
573	1,188	0.72	25,494	15.5
574	4,196	2.56	29,690	18.1
576	1,104	0.67	30,794	18.8
577	3,939	2.40	34,733	21.2
579	1,139	0.69	35,872	21.9
580	3,876	2.36	39,748	24.2
582	4,790	2.92	44,538	27.2
584	4,674	2.85	49,212	30.0
586	4,493	2.74	53,705	32.8
588	4,386	2.68	58,091	35.4

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
589	3,335	2.03	61,426	37.5
590	967	0.59	62,393	38.1
591	4,096	2.50	66,489	40.6
592	3,150	1.92	69,639	42.5
593	4,150	2.53	73,789	45.0
594	910	0.56	74,699	45.6
595	3,908	2.38	78,607	47.9
596	3,034	1.85	81,641	49.8
597	3,804	2.32	85,445	52.1
598	3,853	2.35	89,298	54.5
599	3,824	2.33	93,122	56.8
600	848	0.52	93,970	57.3
601	3,644	2.22	97,614	59.5
602	2,867	1.75	100,481	61.3
603	3,663	2.23	104,144	63.5
604	3,720	2.27	107,864	65.8
605	3,640	2.22	111,504	68.0
606	692	0.42	112,196	68.4
607	3,656	2.23	115,852	70.7
608	2,953	1.80	118,805	72.5
609	3,751	2.29	122,556	74.8
610	782	0.48	123,338	75.2
611	3,758	2.29	127,096	77.5
612	3,037	1.85	130,133	79.4
613	715	0.44	130,848	79.8
614	3,771	2.30	134,619	82.1
616	3,822	2.33	138,441	84.4
617	3,217	1.96	141,658	86.4
618	611	0.37	142,269	86.8
619	640	0.39	142,909	87.2
620	3,202	1.95	146,111	89.1
622	3,762	2.29	149,873	91.4
624	567	0.35	150,440	91.8
625	3,280	2.00	153,720	93.8
627	484	0.30	154,204	94.1
629	3,104	1.89	157,308	95.9
631	390	0.24	157,698	96.2

| Scale | | | Cumulative | |
Score	Freq.	Pct.	Freq.	Pct.
634	2,710	1.65	160,408	97.8
636	272	0.17	160,680	98.0
644	2,039	1.24	162,719	99.2
646	163	0.10	162,882	99.3
648	1,068	0.65	163,950	100.0

Table P34. Mathematics Grade 6 Scale Score Frequency Distribution

| Scale | | | Cumulative | |
Score	Freq.	Pct.	Freq.	Pct.
529	15	0.01	15	0.0
531	1	0.00	16	0.0
534	30	0.02	46	0.0
536	9	0.01	55	0.0
538	128	0.08	183	0.1
540	40	0.02	223	0.1
543	360	0.22	583	0.4
545	99	0.06	682	0.4
547	806	0.50	1,488	0.9
549	228	0.14	1,716	1.1
552	1,532	0.96	3,248	2.0
554	390	0.24	3,638	2.3
556	2,315	1.45	5,953	3.7
558	662	0.41	6,615	4.1
561	3,163	1.98	9,778	6.1
563	907	0.57	10,685	6.7
565	3,624	2.26	14,309	8.9
567	1,056	0.66	15,365	9.6
571	4,133	2.58	19,498	12.2
573	1,149	0.72	20,647	12.9
575	4,165	2.60	24,812	15.5
577	1,168	0.73	25,980	16.2
579	4,126	2.58	30,106	18.8
581	5,123	3.20	35,229	22.0
583	1,105	0.69	36,334	22.7
584	3,902	2.44	40,236	25.1
586	4,868	3.04	45,104	28.2
587	3,575	2.23	48,679	30.4

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
588	1,094	0.68	49,773	31.1
589	4,530	2.83	54,303	33.9
590	3,462	2.16	57,765	36.1
591	995	0.62	58,760	36.7
592	4,381	2.74	63,141	39.4
593	3,361	2.10	66,502	41.5
594	1,052	0.66	67,554	42.2
595	4,153	2.59	71,707	44.8
596	3,142	1.96	74,849	46.8
597	4,118	2.57	78,967	49.3
598	4,041	2.52	83,008	51.9
599	3,892	2.43	86,900	54.3
600	951	0.59	87,851	54.9
601	3,886	2.43	91,737	57.3
602	2,889	1.80	94,626	59.1
603	3,741	2.34	98,367	61.4
604	3,720	2.32	102,087	63.8
605	3,629	2.27	105,716	66.0
606	3,724	2.33	109,440	68.4
607	3,519	2.20	112,959	70.6
608	827	0.52	113,786	71.1
609	3,409	2.13	117,195	73.2
610	2,519	1.57	119,714	74.8
611	3,361	2.10	123,075	76.9
612	3,339	2.09	126,414	79.0
613	769	0.48	127,183	79.4
614	3,211	2.01	130,394	81.5
616	3,195	2.00	133,589	83.4
617	2,379	1.49	135,968	84.9
618	3,150	1.97	139,118	86.9
619	659	0.41	139,777	87.3
620	3,006	1.88	142,783	89.2
622	2,752	1.72	145,535	90.9
624	2,797	1.75	148,332	92.7
626	2,648	1.65	150,980	94.3
628	430	0.27	151,410	94.6
629	2,004	1.25	153,414	95.8

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
631	385	0.24	153,799	96.1
632	1,776	1.11	155,575	97.2
634	319	0.20	155,894	97.4
637	1,505	0.94	157,399	98.3
639	244	0.15	157,643	98.5
645	1,302	0.81	158,945	99.3
647	159	0.10	159,104	99.4
650	983	0.61	160,087	100.0

Table P35. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
518	10	0.01	10	0.0
520	2	0.00	12	0.0
522	23	0.01	35	0.0
524	5	0.00	40	0.0
527	42	0.03	82	0.1
529	7	0.00	89	0.1
532	102	0.07	191	0.1
534	23	0.01	214	0.1
536	241	0.16	455	0.3
538	66	0.04	521	0.3
541	487	0.32	1,008	0.7
543	143	0.09	1,151	0.7
546	813	0.53	1,964	1.3
548	272	0.18	2,236	1.4
550	1,278	0.83	3,514	2.3
552	425	0.28	3,939	2.6
555	1,807	1.17	5,746	3.7
557	570	0.37	6,316	4.1
559	2,241	1.45	8,557	5.5
561	767	0.50	9,324	6.0
564	2,657	1.72	11,981	7.8
566	861	0.56	12,842	8.3
570	2,895	1.87	15,737	10.2
572	1,012	0.66	16,749	10.8
574	3,234	2.09	19,983	12.9

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
576	1,016	0.66	20,999	13.6
577	3,277	2.12	24,276	15.7
579	1,020	0.66	25,296	16.4
580	3,426	2.22	28,722	18.6
582	4,406	2.85	33,128	21.5
584	1,055	0.68	34,183	22.1
585	3,381	2.19	37,564	24.3
586	3,500	2.27	41,064	26.6
587	1,091	0.71	42,155	27.3
588	4,364	2.83	46,519	30.1
590	4,426	2.87	50,945	33.0
591	3,207	2.08	54,152	35.1
592	1,017	0.66	55,169	35.7
593	4,212	2.73	59,381	38.5
594	3,143	2.04	62,524	40.5
595	3,997	2.59	66,521	43.1
596	3,916	2.54	70,437	45.6
597	3,895	2.52	74,332	48.1
598	3,741	2.42	78,073	50.6
599	906	0.59	78,979	51.1
600	3,621	2.34	82,600	53.5
601	2,632	1.70	85,232	55.2
602	3,584	2.32	88,816	57.5
603	3,523	2.28	92,339	59.8
604	3,496	2.26	95,835	62.1
605	3,471	2.25	99,306	64.3
606	3,376	2.19	102,682	66.5
607	3,382	2.19	106,064	68.7
608	3,376	2.19	109,440	70.9
609	3,339	2.16	112,779	73.0
610	749	0.49	113,528	73.5
611	3,441	2.23	116,969	75.7
612	2,681	1.74	119,650	77.5
613	3,421	2.22	123,071	79.7
614	715	0.46	123,786	80.2
615	3,514	2.28	127,300	82.4
616	2,694	1.74	129,994	84.2

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
617	676	0.44	130,670	84.6
618	3,245	2.10	133,915	86.7
619	2,635	1.71	136,550	88.4
620	589	0.38	137,139	88.8
621	3,271	2.12	140,410	90.9
623	485	0.31	140,895	91.2
624	2,603	1.69	143,498	92.9
626	2,907	1.88	146,405	94.8
628	394	0.26	146,799	95.1
630	2,368	1.53	149,167	96.6
632	295	0.19	149,462	96.8
634	1,934	1.25	151,396	98.0
636	237	0.15	151,633	98.2
642	1,592	1.03	153,225	99.2
644	126	0.08	153,351	99.3
647	1,074	0.70	154,425	100.0

Table P36. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
524	16	0.02	16	0.0
526	4	0.00	20	0.0
529	27	0.03	47	0.0
531	4	0.00	51	0.1
533	57	0.06	108	0.1
535	9	0.01	117	0.1
538	115	0.12	232	0.2
540	32	0.03	264	0.3
542	259	0.27	523	0.5
544	86	0.09	609	0.6
547	580	0.60	1,189	1.2
549	174	0.18	1,363	1.4
551	1,032	1.06	2,395	2.5
553	353	0.36	2,748	2.8
556	1,580	1.62	4,328	4.4
558	505	0.52	4,833	5.0
560	2,129	2.19	6,962	7.2

Scale Score	Freq.	Pct.	Cumulative Freq.	Pct.
562	697	0.72	7,659	7.9
565	2,542	2.61	10,201	10.5
567	866	0.89	11,067	11.4
571	2,910	2.99	13,977	14.4
573	954	0.98	14,931	15.3
576	2,947	3.03	17,878	18.4
578	1,053	1.08	18,931	19.5
580	2,973	3.06	21,904	22.5
582	1,025	1.05	22,929	23.6
583	2,862	2.94	25,791	26.5
585	3,627	3.73	29,418	30.2
587	959	0.99	30,377	31.2
588	2,612	2.68	32,989	33.9
590	3,261	3.35	36,250	37.3
591	2,256	2.32	38,506	39.6
592	875	0.90	39,381	40.5
593	3,057	3.14	42,438	43.6
595	2,868	2.95	45,306	46.6
596	2,003	2.06	47,309	48.6
597	812	0.83	48,121	49.5
598	2,697	2.77	50,818	52.2
599	1,951	2.01	52,769	54.2
600	2,597	2.67	55,366	56.9
601	2,490	2.56	57,856	59.5
602	2,390	2.46	60,246	61.9
603	608	0.62	60,854	62.6
604	2,329	2.39	63,183	64.9
605	1,599	1.64	64,782	66.6
606	2,160	2.22	66,942	68.8
607	2,185	2.25	69,127	71.1
608	2,073	2.13	71,200	73.2
609	1,877	1.93	73,077	75.1
610	1,850	1.90	74,927	77.0
611	1,765	1.81	76,692	78.8
612	1,731	1.78	78,423	80.6
613	1,671	1.72	80,094	82.3
614	358	0.37	80,452	82.7

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
615	1,628	1.67	82,080	84.4
616	1,257	1.29	83,337	85.7
617	1,456	1.50	84,793	87.2
618	1,487	1.53	86,280	88.7
619	262	0.27	86,542	89.0
620	1,388	1.43	87,930	90.4
622	1,281	1.32	89,211	91.7
623	1,087	1.12	90,298	92.8
624	192	0.20	90,490	93.0
625	1,187	1.22	91,677	94.2
627	1,104	1.13	92,781	95.4
629	1,022	1.05	93,803	96.4
631	87	0.09	93,890	96.5
632	864	0.89	94,754	97.4
634	84	0.09	94,838	97.5
636	804	0.83	95,642	98.3
638	56	0.06	95,698	98.4
642	685	0.70	96,383	99.1
644	54	0.06	96,437	99.1
651	476	0.49	96,913	99.6
653	32	0.03	96,945	99.7
656	339	0.35	97,284	100.0