New York State Testing Program 2024: English Language Arts and Mathematics Grades 3–8

Technical Report
Prepared for the New York State Education Department by NWEA

2024

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 2024 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 2024 Grades 3–8 ELA and Mathematics NYSTP measures grade-level-appropriate student knowledge and skills as defined by the New York State Next Generation Learning Standards for ELA and mathematics. The tests are designed to allow the classification of student proficiency into four performance levels (Level 1, Level 2, Level 3, and Level 4). Likewise, the test provides opportunities for students at each performance level to demonstrate their knowledge and skills in the Next Generation 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 or independent schools may participate in the testing program, but their participation is not mandatory. In 2024, some religious or independent schools participated in the testing program across all grades. These schools were included in the data analyses. Public school and charter school students are required to take all State assessments administered at their grade, except for students who took a Regents-level examination in mathematics or 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 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/sam-g3-8-2024.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 meet expectations set by the New York State Next Generation Learning Standards. Results of the tests are used to determine whether schools, districts, and the State meet the required progress objectives specified in the New York State Every Student Succeeds Act (ESSA, 2015) accountability system. Several types of scores are available from the Grades 3–8 ELA and Mathematics Tests, which are discussed in this section.

1.4.1. Scale Scores

The scale scores quantify the proficiency measured by the Grades 3–8 ELA and Mathematics Tests at each grade. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades nor 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 can be used to determine student progress within schools and districts, support registration of schools and districts, determine the 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 1, Level 2, Level 3, or Level 4 for the Grades 3–8 ELA and Mathematics Tests. The definition of each performance level is as follows:

- **NYS Level 1:** Students performing at this level are below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered insufficient for the expectations at this grade.

- **NYS Level 2:** Students performing at this level are partially proficient in standards for their grade. They demonstrate knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered partial but insufficient for the expectations at this grade. Students performing at Level 2 are considered on track to meet current New York State high school graduation requirements but are not yet proficient in the Next Generation Learning Standards at this grade.

- **NYS Level 3:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered sufficient for the expectations at this grade.

- **NYS Level 4:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels 1, 2, 3, and 4 were established during the standard-setting process in summer 2023. The standard setting process is described in detail in Appendix Q in the 2023 operational technical report (NYSED, 2023).

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 Next Generation 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 the major clusters. Table 1.1 and Table 1.2 present the reporting subscore categories and the point values that correspond to each of the 2024 tests.

Table 1.1. ELA Subscore Categories and Total Possible Score Points

Grade	Total Subscore Points	
	Reading	Writing to Sources
3	23	10
4	23	14
5	26	14
6	26	14
7	33	14
8	33	14

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 13	Number and Operations—Fractions 10	Measurement and Data 9
4	Operations and Algebraic Thinking 9	Number and Operations in Base Ten 11	Number and Operations—Fractions 11
5	Number and Operations in Base Ten 14	Number and Operations—Fractions 17	Measurement and Data 11
6	Ratios and Proportional Relationships 10	The Number System 11	Expressions, Equations, and Inequalities 17
7	Ratios and Proportional Relationships 15	The Number System 11	Expressions, Equations, and Inequalities 14
8	Expressions, Equations, and Inequalities 16	Functions 10	Geometry 16

1.5. Testing Accommodations

In accordance with federal law under the Individuals with Disabilities Education Act (IDEA, 2004) and the "Fairness in Testing" section of the *Standards for Educational and Psychological Testing* (AERA et al., 2014, pp. 49–72), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. This 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 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/sam-g3-8-2024.pdf.

1.6. Test Transcriptions

For visually impaired students, large type and braille editions of the test books are provided. In most cases, students dictate and/or record their responses, and teachers transcribe student responses to multiple-choice items onto scannable answer sheets and transcribe responses to 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 and printed by NWEA. 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 in order to provide students with 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 locations and bilingual glossaries.

Section 2: Test Design and Development

2.1. Test Descriptions

The 2024 Grades 3–8 ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) items based on the New York State Next Generation Learning Standards. The tests were administered in New York State classrooms during a three-day period for paper-based tests and a thirty-day period for computer-based tests from April to May 2024. 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 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/sam-g3-8-2024.pdf.

2.1.1. ELA Tests

The 2024 Grade 3–8 ELA Tests are designed to measure student literacy, as defined by the Next Generation Learning Standards. The tests assess reading, writing, and language standards by using multiple-choice, short-response, and extended-response items. All items are based on close readings of informational, literary, or paired texts. All texts are drawn from authentic, grade-appropriate works.

Multiple-choice items are designed to assess reading and language standards. Multiple-choice items require students to analyze different aspects of a given text, including central idea, style elements, character and plot development, and vocabulary.

Short-response items are designed to assess reading and language standards. These are single items in which students use textual evidence to support their answers to inferential questions. These items ask 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 are expected to write in complete sentences. Appendix H provides the rubric for the short-response items.

Extended-response items are designed to assess reading, writing, and language standards, with a primary focus on the writing standard. Extended-response items require comprehension and analysis of paired texts in Grades 4–8. Paired texts require students to read and analyze two related texts. Paired texts are related by theme, genre, tone, time period, or other characteristics. Many extended-response items ask students to express a position and support it with text-based evidence. For paired texts, students are expected to synthesize ideas between and draw evidence from both texts. Extended-response items require 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 2024 Grade 3–8 Mathematics Tests are designed to measure student understanding of mathematics, as defined by the Next Generation Learning Standards. The tests require 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 contain multiple-choice, short-

response (1-point and 2-point), and extended-response (3-point) items. For multiple-choice items, students select the correct response from four answer choices. For short- and extended-response items, students write an answer to an open-ended question. Some items require students to show their work or to explain, in words, how they arrived at their answers.

Mathematics multiple-choice items are used mainly to assess standard algorithms and conceptual standards. Multiple-choice items incorporate the Next Generation Learning Standards, some in real-world applications. Many multiple-choice items require students to complete multiple steps. Likewise, many of these items are linked to more than one standard, drawing on the simultaneous application of multiple skills and concepts.

Short-response items are used mainly to assess conceptual and application standards. These items require students to complete a task and show their work. Like multiple-choice items, short-response items often require 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 are 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 require students to complete two or more tasks (or a more extensive problem) and show their work. Some items also assess 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 Grades 3–8 ELA Tests are composed of two sessions per grade and administered over two days during the testing administration window. Each day consists of one session. Session 1 contains literary and informational reading passages, multiple-choice (MC) items, and two 2-point constructed-response (CR) items based on the passages. For Grade 3, Session 2 contains reading passages, MC items, and three 2-point CR items based on those passages. For Grades 4 through 8, Session 2 contains reading passages, MC items, three 2-point CR items, and one 4-point CR item based on those passages.

The Grades 3–8 Mathematics Tests are composed of two sessions per grade and administered over two days during the testing administration window. Each day consists of one session. Session 1 contains MC items, and Session 2 contains MC items as well as 1-point CR items, 2-point CR items, and one 3-point CR item.

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 2024 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 are not disclosed and they look the same as operational test items, students are 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 2024, although it did not eliminate the need for them. See Section 2.9: Field Testing for more information about field testing.

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 passage selection, item writing, educator item review, operational forms construction, 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 4 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.

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 is determined.

The New York State Next Generation Learning Standards for ELA are organized into four strands: reading, writing, language, and speaking/listening. Due to administration constraints, speaking/listening has been 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 review

the reading, writing, and language standards and recommend content coverage by standard and item type, based on the depth and breadth of each standard.

The New York State Next Generation 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 review the mathematics standards and recommend 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 blueprints and actual numbers 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 numbers of points on the 2024 operational tests. Tables A3 and A4 in Appendix A display the anticipated testing times by grade for ELA and mathematics, respectively.

2.5. Passage Selection and Item Review Criteria Documents

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

Passage review criteria documents were created based on the passage-selection guidelines and are used to evaluate each potential passage and determine whether it can be used to measure the New York State Next Generation Learning Standards for ELA. The criteria documents are used to determine whether each passage suggested for testing use is grade appropriate, fair, and possesses the necessary characteristics to assess each standard. Specifically, passages are 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 Grades 3–8 ELA Tests are used to help ensure that each item is clear and fair, measures a specific standard or standards with fidelity, and conforms to the specifications for each item type. Each section of the criteria includes pertinent questions used to determine whether an item is of sufficient quality to move forward in the development process. The first two 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 being 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 grade appropriate 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 2-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 Grades 3–8 Mathematics Tests are used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to the New York State Next Generation 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 of text 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 being 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 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 (e.g., making sure that each 3-point constructed-response item involves a multi-step process and requires students to show work).

2.5.1. Principles of Universal Design

To create tests as equitable as possible for students, principles of Universal Design are 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 et al., 2002, p. 5). 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 Next Generation Learning Standards–aligned test items. To do so, independent passage finders are recruited and trained, using passage-selection resources such as the passage-selection criteria. Passage finders are given assignments based on the test-blueprint requirements. Passage finders submit passages, along with completed criteria documents and source information, to ELA content specialists, who review the passages against the agreed-upon criteria. Passages that do not meet the criteria are rejected, and passages that do meet the criteria are moved forward in the process, where the text from scanned copies of the original sources is entered into templates. Once in the templates, readability metrics are determined for each text. Passages are 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 go through these review steps, ELA content specialists post the passages and complete criteria documents for NYSED's review and approval for moving forward in the process.

NYSED staff retrieve and review the passages and criteria documents. If NYSED staff determine that a passage does not meet the criteria, the passage is rejected, and NYSED staff provide an explanation for the rejection.

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

2.7. Item Development

Item development for the 2024 test forms was conducted during the recent annual development cycles. The goal of item development is to develop a sufficient number of high-quality, Next Generation 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 train item writers. The item writers have teaching or assessment experience in the content area for which they are writing items; experience in writing for large-scale, high-stakes assessments; and, at a minimum, a bachelor's degree in either education and/or the content area for which they are assigned. The item writers are given specific assignments, based on the test blueprints. For ELA, the item writers are also provided with the completed passage-criteria documents.

At least two content specialists review each item provided by item writers. After the content specialists agree all the items meet the criteria, copyeditors review the items. The mathematics items are also reviewed by content specialists in science and social studies and by research librarians, when necessary. The ELA and mathematics content specialists evaluate the feedback from the different internal groups and edit the items accordingly. The items are then posted for NYSED's review and approval to proceed with the process.

NYSED content experts retrieve and review the items. If NYSED staff determine that an item does not meet the criteria, NYSED staff provide an explanation for rejection or revision. If NYSED staff determine that an item meets the criteria but can be improved with editing, the staff member records notes for the edits. Those notes are reviewed during meetings at which NWEA content specialists and NYSED staff review and edit all the items to ensure that they meet the criteria. All passages and items accepted at that meeting are moved forward for educator item review.

2.8. Educator Item Review

After being reviewed by NYSED, the items are presented to panels of New York State educators. Based on their expertise, educators are assigned to grade- and content-specific groups where they review the items. The reviews are facilitated by NWEA content specialists and are attended by NYSED staff. For ELA, reviewers first read and then discuss the passages before reviewing items. For ELA and mathematics, the educators use 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 item 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 review the items, they discuss their judgments about them. If the educators feel that an item does not align to the standards, does not meet quality standards, or is not fair, they make recommendations for editing the item. NYSED staff and NWEA content specialists later review the recommendations and make 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 are field tested than are needed for the operational forms because that enables tests to be constructed with items that include the best possible characteristics from both a content and psychometric perspective.

There were two types of approaches used for field testing in Spring 2024: embedded field testing and stand-alone field testing. With embedded field testing, field test items were administered within the 2024 operational test forms for most multiple-choice items. With stand-alone field testing, field test items were administered separately from the 2024 operational forms later in Spring 2024, which included CR items and a small number of MC items.

A variety of analyses were conducted to better understand how the items field tested in 2024 may perform on future operational forms. All 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 New York State population in terms of student achievement on prior years' tests, student gender, student ethnicity, and school Needs Resource Capacity (NRC) category. 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 *NYSTP 2024: ELA and Mathematics Grades 3–8 Field Test Technical Report*.

2.10. Rangefinding and Pre-Scoring

The 2024 operational items were field tested in 2019, 2022, and 2023. NWEA conducted rangefinding for items field tested in 2019 and 2022 and pre-scoring for items field tested in 2023 and beyond.

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 NWEA Scoring Directors oversee this process.

The first step in the rangefinding process is 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. A grounding guide set contains student responses that illustrate the full range of scores on the rubric. This set is 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 reviews the pre-approved grounding guide set, groups of committee members familiarize themselves with each item type, scoring a small number of responses representative of each of the different score points. After a group-scoring exercise, committee members independently score other student responses. The committee then reviews and discusses their results and determines consensus scores for the responses. The rangefinding results are used to build training materials for NWEA scorers, who score the field test responses to constructed-response items.

The purpose of pre-scoring, which began in 2023, mirrors that of rangefinding. During the pre-score process, NYSED content experts and NWEA scoring experts come to consensus regarding the scoring of the field test items. After reviewing and scoring a variety of student responses that represent all possible score points for each item, NYSED and NWEA staffs compare and discuss results and determine consensus scores for the responses. These pre-score results are used to

build training materials that NWEA scorers use to score the field test responses to constructed-response items.

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

Test items were selected from the pools of available ELA and mathematics items. These items were field tested either by embedded field testing or stand-alone field testing in 2019, 2022, or 2023.

The test-construction process involves several iterative steps. Three criteria govern 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

NWEA content specialists are provided the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines are 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 is 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 are due to content considerations that require the inclusion of particular items. Decisions to use such items are made very carefully.

Using the pool of field tested items, NWEA content specialists make preliminary selections for each grade and content area. The selections are then reviewed by the content leads for each content area to make sure that the items conform to the different criteria. If the content criteria are not met, new items are selected. After the content leads' review, the item selections are reviewed by NWEA psychometricians. If items with undesirable statistics are selected, the psychometricians propose items with more desirable statistics. The content specialists and their leads then review those items. Once the NWEA content teams and the psychometric teams are satisfied that the content and statistics of the selected items and the proposed whole forms meet the requirements, the items are given to NYSED staff (including content and assessment experts) to review. NWEA content specialists and psychometricians traveled to Albany, New York, in September 2023 to finalize item selection and test creation for the 2024 tests 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 August 24–25, 2023, in Albany, New York, educators from around the State worked with NYSED and NWEA to review the content of the proposed operational ELA passages and individual ELA and mathematics 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 resulted in test forms that meet psychometric parameters and contain items that meet content criteria.

On August 24, 2023, educators reviewed Grades 4, 6, and 8. Different educators reviewed Grades 3, 5, and 7 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 a general session and then 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 their own checklist and had access to NWEA'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 became familiar with 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 became familiar with the process, the educators reviewed groups of items (e.g., 4 to 6 items, followed by a 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 NWEA 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 NWEA and NYSED proposed for any necessary replacements. Following each session with educators, NYSED and NWEA 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 is completed, NWEA creates test forms. The test forms are reviewed by NWEA content specialists and are

posted for NYSED to review. NYSED and NWEA review the forms to look for any errors in spelling, capitalization, punctuation, grammar, and formatting. They also confirm that each multiple-choice item has a single correct answer.

2.14. Final Eyes Committees

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

Following the Final Eyes review and after NYSED approves edits made as a result of the review, the tests are then considered final and produced for administration.

2.15. Proficiency and Performance Standards

In August 2023, after the operational administration of the 2023 tests, a standard setting meeting occurred in Albany, where approximately 65 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 for the Next Generation Learning Standards. These recommendations were presented to the Commissioner, who, in turn, adopted the recommended standards set forth by the committees. For additional details, see Section 8 and Appendix Q of the 2023 operational technical report (NYSED, 2023).

Each grade has four performance levels. Three cut points demarcate the performance levels needed to demonstrate each ascending level of performance. Section 6.5.1 contains information about the raw-score-to-scale-score conversion tables, CSEMs, and the performance levels.

Section 3: Validity

The *Standards for Educational and Psychological Testing* refers to validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014, p. 11). Test validation is an ongoing process of gathering evidence from many sources to evaluate the soundness of the desired score interpretations or uses. This evidence is acquired from studies of the content of the test as well as studies involving scores produced by the test. Additionally, reliability has to be taken into account 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* addresses 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 & Lehmann, 1991). Tests are now also used for the purposes of accountability. Specific to student-level outcomes, the NYSTP documents student performance in ELA as defined by the New York State ELA Next Generation Learning Standards and in mathematics as defined by the New York State Mathematics Next Generation Learning Standards.

For test-score interpretations to be appropriate for this purpose, the content of the test must be carefully matched to the specified standards. The *Standards for Educational and Psychological Testing* states that content-related evidence of validity is a central concern during test development (AERA et al., 2014). 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 a 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 standards (see Tables B1 and B2 in Appendix B). The NYSTP test-development process requires specific attention to content representation and balance within each test form. New York State educators were involved in test construction at various development stages. For example, during the item-review process, they reviewed field test items for alignment with the Next Generation Learning Standards. Section 2: Test Design and Development contains more information specific to the item-review process.

As a means of collecting further content validity evidence, a third-party alignment study was conducted by ACS Ventures, LLC in November 2023 to evaluate the degree to which the tests measure the content standards they are supposed to measure. See the *Evaluation of Alignment of*

New York State Assessment Program to the New York Next Generation Learning Standards for the full details of this alignment study.

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 because high coefficients imply that the test items measure the same domain of skill and are reliable and consistent. Reliability coefficients of the tests for total populations and subgroups of students are presented in Section 7.1: Test Reliability. For the total population, the ELA reliability coefficients (Cronbach's alpha) ranged from 0.87 to 0.92. For all subgroups, the reliability coefficients were greater than or equal to 0.74, except for the non-binary gender group. 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, except for the non-binary gender group. Overall, the high internal consistency of the NYSTP Grades 3–8 ELA and Mathematics Tests provide 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 item response theory (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, except for one item in Grade 4 mathematics. 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 one another within their respective content areas. This relationship of the items within the ELA or mathematics tests shows 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. A large main factor found using this analysis 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 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 New York State public, charter, and religious or 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 with 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.

Factor analyses related to the Grades 3–8 ELA and Mathematics Tests indicate 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 show that the first eigenvalues were at least 5 times as large as the second eigenvalues for all grades.

All the Grades 3–8 ELA and Mathematics Tests exhibited the first principal component, accounting for more than 20.50% and 25.01% of the test variance, respectively. Table 3.1 and Table 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			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	**7.24**	**25.86**	**25.86**
	2	1.27	4.54	30.40
	3	1.03	3.67	34.06
4	1	**7.34**	**25.32**	**25.32**
	2	1.26	4.35	29.67
	3	1.16	4.01	33.68
	4	1.01	3.48	37.16
5	1	**6.56**	**20.50**	**20.50**
	2	1.30	4.06	24.56
	3	1.03	3.22	27.78
	4	1.00	3.13	30.91
6	1	**7.17**	**22.41**	**22.41**
	2	1.41	4.40	26.81
	3	1.08	3.36	30.18
	4	1.01	3.14	33.32
7	1	**8.12**	**20.81**	**20.81**
	2	1.46	3.75	24.56
	3	1.13	2.91	27.47
	4	1.03	2.64	30.10
8	1	**9.95**	**25.52**	**25.52**
	2	1.39	3.57	29.09
	3	1.14	2.92	32.01
	4	1.04	2.66	34.67

Table 3.2. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	**9.01**	**28.17**	**28.17**
	2	1.38	4.30	32.47
	3	1.12	3.49	35.96
	4	1.05	3.27	39.23
4	1	**10.67**	**28.83**	**28.83**
	2	1.39	3.76	32.59
	3	1.10	2.98	35.56
	4	1.01	2.74	38.30
5	1	**9.66**	**26.11**	**26.11**
	2	1.77	4.79	30.91

Grade	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	3	1.17	3.17	34.08
6	1	**11.03**	**28.28**	**28.28**
	2	1.49	3.83	32.11
	3	1.07	2.73	34.84
7	1	**10.28**	**25.07**	**25.07**
	2	1.67	4.06	29.13
	3	1.36	3.31	32.44
	4	1.05	2.55	35.00
8	1	**10.26**	**25.01**	**25.01**
	2	1.38	3.37	28.39
	3	1.08	2.63	31.02

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 those obtained from the total population data. Evaluation of the magnitude of the 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 means 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 a 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 a 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 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 NWEA'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 (Jensen, 1980; Sandoval & Mille, 1980). Thus, empirical studies were conducted.

Statistical methods were employed to evaluate the amount of DIF in all test items: polytomous items were evaluated with standardized mean differences, and dichotomous items were analyzed using Mantel-Haenszel (MH) 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) and remained in the tests. Please refer to the *NYSTP 2024: ELA and Mathematics Grades 3–8 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 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* (*SAM*) located at https://www.nysed.gov/sites/default/files/programs/state-assessment/sam-g3-8-2024.pdf and the 2024 NYSTP *Grades 3–8 English Language Arts, Mathematics, and Science Tests Scoring Leader Handbook* located at https://www.nysed.gov/sites/default/files/programs/state-assessment/3-8-scoring-leader-handbook-2024.pdf.

4.1. Test Administration

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered to students in either a paper-based (PBT) or computer-based (CBT) testing mode in 2024, except for Grades 5 and 8, which were entirely CBT. The PBT testing window for the Grades 3–8 ELA Tests was April 10–12. The PBT testing window for the Grades 3–8 Mathematics Tests was May 7–9. The CBT testing window for both the Grades 3–8 ELA and Mathematics Tests was April 8–May 17.

Makeup-test administration windows allowed students who were ill or otherwise unable to test during their original assigned window to still take the tests. The makeup-test administration window for PBT was April 15–17 for the Grades 3–8 ELA Tests and May 10–14 for the Grades 3–8 Mathematics Tests. For CBT ELA and mathematics, makeup tests were to be administered within the testing window.

4.2. Scoring Models

For the 2023–2024 school year, schools and school districts were able to score Grades 3–8 ELA and/or Mathematics Tests regionally, multi-districtwide, 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 include either staff from three or more school districts or staff from two or more religious or independent schools in an affiliation group (religious, independent, or charter schools may participate in regional scoring with public school districts, and each religious, independent, or charter school may be counted as one district).
2. Schools from two districts—The scorers for the school's test papers include either staff from two school districts, two religious or independent schools, two charter schools, or a combination of two from those.
3. Three or more schools within a district—The scorers for the school's test papers include staff from at least three schools within a district.
4. Two schools within a district—The scorers for the school's test papers include staff from two schools within a district (not available for CBT schools).
5. One school (local scoring)—The scorers for the school's test papers include three or more staff for each grade, all from the same school (not available for CBT schools).
6. Private contractor—Scoring is conducted by a private contractor that does not belong to the Boards of Cooperative Educational Services (BOCES).

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

4.3. Scoring Procedures of Operational Tests

Operational tests contain multiple-choice (MC) and constructed-response (CR) items. All operational MC items are machine scored. This section describes the scoring of the operational CR items.

Qualified teachers and administrators performed the scoring of the NYSTP 2024 Grades 3–8 ELA and Mathematics Tests at designated sites. The number of personnel at a given site varies, as districts have the option of regional, districtwide, or schoolwide scoring (please refer to Section 4.2: Scoring Models for more details). Administrators are responsible for overseeing scoring operations, including preparation of the test site, security of the test materials, and supervision of the scoring process. At each site, designated trainers teach scoring committee members the basic criteria for scoring each item and monitor the scoring sessions in the room. Facilitators or leaders, who also help monitor the sessions and enforce scoring accuracy, assist the trainers.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, a Site Coordinator conducts oversight. A Scoring Leader trains the scoring committee members and monitors the sessions, and a Table Facilitator assists in monitoring the sessions. For each subject, oversight is structured in the same way for district- and schoolwide models. At the districtwide level, a School District Administrator oversees scoring. A district subject leader trains the scoring committee members and monitors the sessions, and a school subject leader assists in monitoring the sessions. For schoolwide scoring, oversight is provided by the principal; otherwise, titles for the schoolwide model are the same as those for the districtwide model. The general title "scoring committee members" includes scorers at every site. Details on titles and responsibilities for each scoring model can be found on page 17 of the 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/sam-g3-8-2024.pdf.

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

- Scoring Model 4 (two schools within a district) and Scoring Model 5 (one school) are not available for CBT. Please refer to page 16 of the 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* for descriptions of all the scoring models.
- For CBT, scorers use the ScorePoint system to score responses.

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 are scoring guides. These guides were created by NWEA from sets of actual field test student responses that were consensus scored by NYSED during the pre-score process or by NYS teachers during rangefinding. Trainers use these materials to train

scoring committee members on the criteria for scoring CR items and rubric application. Additionally, *Scoring Leader Handbooks* are distributed to provide guidelines, information, and procedures for both the scorers and Site Coordinators to facilitate scoring.

The constructed-response items are divided into three groups for scoring, and three separate scoring committee members score each constructed-response item in the group they are assigned. After scoring is completed, the Table Facilitator or subject (ELA or mathematics) leader conducts 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 2024 Grades 3–8 ELA and Mathematics Tests. Trainers use scoring guides to train scoring committee members on the criteria for scoring constructed-response items. Part of the training process is the administration of a consistency assurance set (CAS) that provides the State's scoring sites with information regarding strengths and weaknesses of their scorers. This tool allows trainers to retrain their scorers, if necessary. The CAS also acknowledges those scorers who grasp all aspects of the content area being scored and are 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 "Assigning Scorer Numbers and Questions to PBT Scoring Committee Members" section (page 26) of the 2024 NYSTP *Grades 3–8 English Language Arts, Mathematics, and Science Tests Scoring Leader Handbook* located at https://www.nysed.gov/sites/default/files/programs/state-assessment/3-8-scoring-leader-handbook-2024.pdf.

4.6. Quality Control Process

Test books and electronic responses are randomly distributed throughout each scoring room so that completed tests from each region, district, school, or class are evenly dispersed. Scoring teams are divided into groups of three to ensure that a variety of scorers grade each test. If a scorer and a facilitator cannot reach a decision after reviewing the scoring guides, they call the NWEA Scoring Helpline, a call center established to help teachers and administrators during scoring. The helpline staff consists of trained NWEA personnel who answer questions by phone. When a member of the staff is unable to resolve an issue, it is referred to NYSED for a scoring decision. A quality check is also performed to certify that all the items are scored and that the scoring committee members darkened each score on the answer document appropriately. The log of calls received by the scoring helpline is delivered to NYSED daily during the scoring window. To affirm that scoring guidelines and policies are followed, approximately 5% of student results for both ELA and mathematics are audited by NWEA. Details of the audit results can be found in the 2024 scoring audit report.

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 more than 95% of the student test records was received from the data warehouse and delivered to NWEA at the end of May 2024. During Phase 2, "straggler files" were submitted to NWEA in June 2024.

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 or 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, but more stringent data-cleaning procedures were applied for the analyses described in other sections (see details below).

Data processing 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. NWEA's psychometric team performed data cleaning on 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 does 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: Summary of Operational Test Results).

The largest group of cases excluded from the data set used for analyses (Sections 5, 6, and 7) was "Not Tested." These students were not tested for various reasons, including, for example, administrative error, not being enrolled at the time of the test, being medically excused, taking the New York State Alternate Assessment (NYSAA), being a first-year ELL, or not attempting any test items. For mathematics, students in the upper grades, especially Grade 8, who took the Regents exam were also part of this excluded group. Other deleted cases included students with missing school type information, incorrect or incomplete grade information, duplicate records, no-response records, or mismatched form codes.

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 Remaining
Initial Number of Cases	n/a	186,578
Wrong Subject	0	186,578

Exclusion Rule	#Deleted	#Cases Remaining
No Grade	23	186,555
Wrong Grade	49	186,506
Form Code Mismatch	391	186,115
School Type	3,899	182,216
Not Tested	18,161	164,055
Out-of-Range CR Score	0	164,055
Duplicated Record	4	164,051
Test Mode Discrepancy	0	164,051

Table 5.2. ELA Grade 4 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	193,568
Wrong Subject	0	193,568
No Grade	0	193,568
Wrong Grade	50	193,518
Form Code Mismatch	359	193,159
School Type	3,509	189,650
Not Tested	21,056	168,594
Out-of-Range CR Score	0	168,594
Duplicated Record	8	168,586
Test Mode Discrepancy	0	168,586

Table 5.3. ELA Grade 5 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	189,098
Wrong Subject	0	189,098
No Grade	0	189,098
Wrong Grade	11	189,087
Form Code Mismatch	17	189,070
School Type	3,727	185,343
Not Tested	26,579	158,764
Out-of-Range CR Score	0	158,764
Duplicated Record	0	158,764
Test Mode Discrepancy	0	158,764

Table 5.4. ELA Grade 6 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	196,303
Wrong Subject	0	196,303
No Grade	0	196,303
Wrong Grade	40	196,263
Form Code Mismatch	342	195,921
School Type	4,625	191,296

Exclusion Rule	#Deleted	#Cases Remaining
Not Tested	26,481	164,815
Out-of-Range CR Score	0	164,815
Duplicated Record	12	164,803
Test Mode Discrepancy	0	164,803

Table 5.5. ELA Grade 7 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	196,632
Wrong Subject	0	196,632
No Grade	0	196,632
Wrong Grade	46	196,586
Form Code Mismatch	249	196,337
School Type	4,806	191,531
Not Tested	31,700	159,831
Out-of-Range CR Score	0	159,831
Duplicated Record	14	159,817
Test Mode Discrepancy	0	159,817

Table 5.6. ELA Grade 8 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	195,778
Wrong Subject	0	195,778
No Grade	0	195,778
Wrong Grade	7	195,771
Form Code Mismatch	22	195,749
School Type	5,180	190,569
Not Tested	43,312	147,257
Out-of-Range CR Score	0	147,257
Duplicated Record	0	147,257
Test Mode Discrepancy	0	147,257

Table 5.7. Mathematics Grade 3 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	187,224
Wrong Subject	0	187,224
No Grade	0	187,224
Wrong Grade	49	187,175
Form Code Mismatch	9,665	177,510
School Type	3,607	173,903
Not Tested	17,221	156,682
Out-of-Range CR Score	0	156,682
Duplicated Record	9	156,673
Test Mode Discrepancy	0	156,673

Table 5.8. Mathematics Grade 4 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	193,550
Wrong Subject	0	193,550
No Grade	0	193,550
Wrong Grade	46	193,504
Form Code Mismatch	10,023	183,481
School Type	3,161	180,320
Not Tested	22,816	159,948
Out-of-Range CR Score	0	159,948
Duplicated Record	5	159,943
Test Mode Discrepancy	0	159,943

Table 5.9. Mathematics Grade 5 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	188,970
Wrong Subject	0	188,970
No Grade	0	188,970
Wrong Grade	10	188,960
Form Code Mismatch	7,409	181,551
School Type	3,680	177,871
Not Tested	25,732	152,139
Out-of-Range CR Score	2	152,139
Duplicated Record	0	152,139
Test Mode Discrepancy	0	152,139

Table 5.10. Mathematics Grade 6 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	197,297
Wrong Subject	0	197,297
No Grade	0	197,297
Wrong Grade	37	197,260
Form Code Mismatch	11,312	185,948
School Type	4,101	181,847
Not Tested	27,696	154,151
Out-of-Range CR Score	0	154,151
Duplicated Record	10	154,141
Test Mode Discrepancy	0	154,141

Table 5.11. Mathematics Grade 7 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	198,186
Wrong Subject	0	198,186
No Grade	0	198,186

Exclusion Rule	#Deleted	#Cases Remaining
Wrong Grade	47	198,139
Form Code Mismatch	10,942	187,197
School Type	4,257	182,940
Not Tested	34,194	148,746
Out-of-Range CR Score	0	148,746
Duplicated Record	10	148,736
Test Mode Discrepancy	0	148,736

Table 5.12. Mathematics Grade 8 Data Cleaning

Exclusion Rule	#Deleted	#Cases Remaining
Initial Number of Cases	n/a	186,458
Wrong Subject	0	186,458
No Grade	0	186,458
Wrong Grade	6	186,452
Form Code Mismatch	6,736	179,716
School Type	4,667	175,049
Not Tested [a]	84,154	90,895
Out-of-Range CR Score	0	90,895
Duplicated Record	0	90,895
Test Mode Discrepancy	0	90,895

[a] The number of students "Not Tested" was larger here than for other mathematics grades due to some students taking a Regents exam instead of the NYSTP Grade 8 mathematics test.

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, including gender, ethnicity, Needs Resource Capacity (NRC) category, English Language Learner (ELL) status, students with disabilities (SWDs), students using test accommodations (SUAs), 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). The NRC category is assigned at the district level and is an indicator of district and school socioeconomic status. The ethnicity and gender designations are based on student-level information.

Table 5.13. ELA Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	81,734	49.82
	Male	82,297	50.17
	Non-Binary	20	0.01
Ethnicity	Asian	17,152	10.47
	African American	22,924	14.00
	Hispanic	45,209	27.60
	American Indian	1,264	0.77

Demographic Category		N-Count	% of Total N-Count
	Multiracial	6,153	3.76
	Pacific Islander	317	0.19
	White	70,762	43.21
NRC	New York City	46,749	28.50
	Big 4 Cities	5,881	3.58
	Urban/Suburban	12,363	7.54
	Rural	9,132	5.57
	Average Needs	44,294	27.00
	Low Needs	20,698	12.62
	Charter	14,362	8.75
	Religious or Independent	10,572	6.44
SWD	No	140,199	85.46
	Yes	23,852	14.54
SUA	No	142,211	86.69
	Yes	21,840	13.31
ELL	No	142,263	86.72
	Yes	21,788	13.28
SWD/ SUA	No	145,690	88.81
	Yes	18,361	11.19
ELL/ SUA	No	160,973	98.12
	Yes	3,078	1.88

Note. The total n-count was 164,051.

Table 5.14. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	84,385	50.05
	Male	84,192	49.94
	Non-Binary	9	–
Ethnicity	Asian	17,945	10.66
	African American	24,114	14.33
	Hispanic	46,493	27.63
	American Indian	1,177	0.70
	Multiracial	6,177	3.67
	Pacific Islander	328	0.19
	White	72,035	42.81
NRC	New York City	47,861	28.39
	Big 4 Cities	6,012	3.57
	Urban/Suburban	12,831	7.61
	Rural	9,412	5.58
	Average Needs	44,176	26.20

Demographic Category		N-Count	% of Total N-Count
	Low Needs	20,796	12.34
	Charter	14,271	8.47
	Religious or Independent	13,227	7.85
SWD	No	143,912	85.36
	Yes	24,674	14.64
SUA	No	144,055	85.45
	Yes	24,531	14.55
ELL	No	148,255	87.94
	Yes	20,331	12.06
SWD/ SUA	No	148,228	87.92
	Yes	20,358	12.08
ELL/ SUA	No	165,359	98.09
	Yes	3,227	1.91

Note. The total n-count was 168,586.

Table 5.15. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,487	49.44
	Male	80,259	50.55
	Non-Binary	18	0.01
Ethnicity	Asian	17,272	10.92
	African American	24,605	15.55
	Hispanic	45,947	29.05
	American Indian	1,225	0.77
	Multiracial	5,807	3.67
	Pacific Islander	365	0.23
	White	62,968	39.81
NRC	New York City	49,162	30.97
	Big 4 Cities	6,170	3.89
	Urban/Suburban	12,328	7.76
	Rural	9,137	5.76
	Average Needs	42,790	26.95
	Low Needs	19,814	12.48
	Charter	14,451	9.10
	Religious or Independent	4,912	3.09
SWD	No	134,313	84.60
	Yes	24,451	15.40
SUA	No	133,606	84.15
	Yes	25,158	15.85

Demographic Category		N-Count	% of Total N-Count
ELL	No	144,895	91.26
	Yes	13,869	8.74
SWD/ SUA	No	138,150	87.02
	Yes	20,614	12.98
ELL/ SUA	No	155,738	98.09
	Yes	3,026	1.91

Note. The total n-count was 158,764.

Table 5.16. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	81,231	49.29
	Male	83,534	50.69
	Non-Binary	38	0.02
Ethnicity	Asian	18,447	11.22
	African American	25,323	15.40
	Hispanic	46,506	28.28
	American Indian	1,226	0.75
	Multiracial	5,522	3.36
	Pacific Islander	353	0.21
	White	67,055	40.78
NRC	New York City	48,873	29.66
	Big 4 Cities	6,033	3.66
	Urban/Suburban	12,418	7.54
	Rural	9,255	5.62
	Average Needs	41,430	25.14
	Low Needs	19,404	11.77
	Charter	15,097	9.16
	Religious or Independent	12,293	7.46
SWD	No	140,655	85.35
	Yes	24,148	14.65
SUA	No	140,331	85.15
	Yes	24,472	14.85
ELL	No	148,526	90.12
	Yes	16,277	9.88
SWD/ SUA	No	145,004	87.99
	Yes	19,799	12.01
ELL/ SUA	No	162,013	98.31
	Yes	2,790	1.69

Note. The total n-count was 164,803.

Table 5.17. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,684	49.23
	Male	81,087	50.74
	Non-Binary	46	0.03
Ethnicity	Asian	17,187	10.78
	African American	25,462	15.97
	Hispanic	45,833	28.75
	American Indian	1,165	0.73
	Multiracial	5,200	3.26
	Pacific Islander	360	0.23
	White	64,224	40.28
NRC	New York City	49,343	30.87
	Big 4 Cities	5,889	3.68
	Urban/Suburban	11,911	7.45
	Rural	9,168	5.74
	Average Needs	38,975	24.39
	Low Needs	19,215	12.02
	Charter	14,398	9.01
	Religious or Independent	10,918	6.83
SWD	No	137,028	85.74
	Yes	22,789	14.26
SUA	No	136,191	85.22
	Yes	23,626	14.78
ELL	No	145,876	91.28
	Yes	13,941	8.72
SWD/ SUA	No	141,042	88.25
	Yes	18,775	11.75
ELL/ SUA	No	157,372	98.47
	Yes	2,445	1.53

Note. The total n-count was 159,817.

Table 5.18. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	70,999	48.21
	Male	76,200	51.75
	Non-Binary	58	0.04
Ethnicity	Asian	16,461	11.23
	African American	24,382	16.63
	Hispanic	44,656	30.45

Demographic Category		N-Count	% of Total N-Count
	American Indian	1,084	0.74
	Multiracial	4,540	3.10
	Pacific Islander	299	0.20
	White	55,213	37.65
NRC	New York City	49,452	33.58
	Big 4 Cities	6,104	4.15
	Urban/Suburban	11,837	8.04
	Rural	8,677	5.89
	Average Needs	36,154	24.55
	Low Needs	17,401	11.82
	Charter	12,770	8.67
	Religious or Independent	4,862	3.30
SWD	No	125,886	85.49
	Yes	21,371	14.51
SUA	No	124,713	84.69
	Yes	22,544	15.31
ELL	No	136,381	92.61
	Yes	10,876	7.39
SWD/ SUA	No	129,448	87.91
	Yes	17,809	12.09
ELL/ SUA	No	145,015	98.48
	Yes	2,242	1.52

Note. The total n-count was 147,257.

Table 5.19. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,042	49.81
	Male	78,614	50.18
	Non-Binary	17	0.01
Ethnicity	Asian	16,825	10.76
	African American	22,311	14.27
	Hispanic	40,407	25.84
	American Indian	1,224	0.78
	Multiracial	5,906	3.78
	Pacific Islander	315	0.20
	White	69,378	44.37
NRC	New York City	44,475	28.39
	Big 4 Cities	5,767	3.68
	Urban/Suburban	10,953	6.99
	Rural	9,123	5.82

Demographic Category		N-Count	% of Total N-Count
	Average Needs	41,041	26.20
	Low Needs	20,444	13.05
	Charter	13,793	8.80
	Religious or Independent	11,077	7.07
SWD	No	134,754	86.01
	Yes	21,919	13.99
SUA	No	136,943	87.41
	Yes	19,730	12.59
ELL	No	137,978	88.07
	Yes	18,695	11.93
SWD/ SUA	No	140,114	89.43
	Yes	16,559	10.57
ELL/ SUA	No	154,202	98.42
	Yes	2,471	1.58

Note. The total n-count was 156,673.

Table 5.20. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,997	50.16
	Male	78,493	49.84
	Non-Binary	9	–
Ethnicity	Asian	17,234	10.97
	African American	22,925	14.59
	Hispanic	40,404	25.71
	American Indian	1,125	0.72
	Multiracial	5,845	3.72
	Pacific Islander	315	0.20
	White	69,324	44.11
NRC	New York City	45,238	28.72
	Big 4 Cities	5,319	3.38
	Urban/Suburban	10,848	6.89
	Rural	9,022	5.73
	Average Needs	40,435	25.67
	Low Needs	19,761	12.55
	Charter	13,447	8.54
	Religious or Independent	13,429	8.53
SWD	No	135,417	85.98
	Yes	22,082	14.02
SUA	No	135,812	86.23

Demographic Category		N-Count	% of Total N-Count
	Yes	21,687	13.77
ELL	No	140,682	89.32
	Yes	16,817	10.68
SWD/ SUA	No	139,550	88.60
	Yes	17,949	11.40
ELL/ SUA	No	154,915	98.36
	Yes	2,584	1.64

Note. The total n-count was 157,499.

Table 5.21. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	74,837	49.19
	Male	77,282	50.80
	Non-Binary	18	0.01
Ethnicity	Asian	16,911	11.17
	African American	24,261	16.02
	Hispanic	40,652	26.84
	American Indian	1,208	0.80
	Multiracial	5,715	3.77
	Pacific Islander	357	0.24
	White	62,360	41.17
NRC	New York City	46,627	30.65
	Big 4 Cities	6,008	3.95
	Urban/Suburban	10,694	7.03
	Rural	8,962	5.89
	Average Needs	41,421	27.23
	Low Needs	19,552	12.85
	Charter	14,252	9.37
	Religious or Independent	4,621	3.04
SWD	No	128,569	84.51
	Yes	23,568	15.49
SUA	No	127,626	83.89
	Yes	24,511	16.11
ELL	No	142,842	93.89
	Yes	9,295	6.11
SWD/ SUA	No	131,993	86.76
	Yes	20,144	13.24
ELL/ SUA	No	149,607	98.34
	Yes	2,530	1.66

Note. The total n-count was 152,137.

Table 5.22. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,158	49.41
	Male	77,951	50.57
	Non-Binary	32	0.02
Ethnicity	Asian	17,696	11.51
	African American	24,014	15.62
	Hispanic	39,688	25.82
	American Indian	1,188	0.77
	Multiracial	5,277	3.43
	Pacific Islander	340	0.22
	White	65,534	42.63
NRC	New York City	44,645	28.96
	Big 4 Cities	5,869	3.81
	Urban/Suburban	10,486	6.80
	Rural	8,956	5.81
	Average Needs	38,144	24.75
	Low Needs	18,956	12.30
	Charter	14,053	9.12
	Religious or Independent	13,032	8.45
SWD	No	133,020	86.30
	Yes	21,121	13.70
SUA	No	132,682	86.08
	Yes	21,459	13.92
ELL	No	141,347	91.70
	Yes	12,794	8.30
SWD/ SUA	No	136,915	88.82
	Yes	17,226	11.18
ELL/ SUA	No	152,133	98.70
	Yes	2,008	1.30

Note. The total n-count was 154,141.

Table 5.23. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	73,408	49.35
	Male	75,290	50.62
	Non-Binary	38	0.03
Ethnicity	Asian	16,305	10.99
	African American	24,121	16.26
	Hispanic	38,948	26.25

Demographic Category		N-Count	% of Total N-Count
	American Indian	1,125	0.76
	Multiracial	4,897	3.30
	Pacific Islander	350	0.24
	White	62,605	42.20
NRC	New York City	44,443	29.88
	Big 4 Cities	5,708	3.84
	Urban/Suburban	10,002	6.72
	Rural	8,899	5.98
	Average Needs	35,867	24.11
	Low Needs	18,370	12.35
	Charter	13,507	9.08
	Religious or Independent	11,940	8.03
SWD	No	128,963	86.71
	Yes	19,773	13.29
SUA	No	128,388	86.32
	Yes	20,348	13.68
ELL	No	138,028	92.80
	Yes	10,708	7.20
SWD/ SUA	No	132,713	89.23
	Yes	16,023	10.77
ELL/ SUA	No	147,101	98.90
	Yes	1,635	1.10

Note. The total n-count was 148,736.

Table 5.24. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	43,245	47.58
	Male	47,612	52.38
	Non-Binary	38	0.04
Ethnicity	Asian	7,260	8.04
	African American	16,936	18.75
	Hispanic	26,285	29.11
	American Indian	697	0.77
	Multiracial	2,919	3.23
	Pacific Islander	169	0.19
	White	36,043	39.91
NRC	New York City	26,291	28.92
	Big 4 Cities	4,872	5.36
	Urban/Suburban	7,899	8.69
	Rural	7,020	7.72

Demographic Category		N-Count	% of Total N-Count
	Average Needs	24,038	26.45
	Low Needs	8,237	9.06
	Charter	8,342	9.18
	Religious or Independent	4,196	4.62
SWD	No	74,845	82.34
	Yes	16,050	17.66
SUA	No	74,184	81.62
	Yes	16,711	18.38
ELL	No	86,277	94.92
	Yes	4,618	5.08
SWD/ SUA	No	77,490	85.25
	Yes	13,405	14.75
ELL/ SUA	No	89,636	98.61
	Yes	1,259	1.39

Note. The total n-count was 90,895.

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 the test reliability measures Cronbach's alpha (Cronbach, 1951) and the Feldt-Raju coefficient (Qualls, 1995). Additionally, classical DIF analysis is conducted at this stage. DIF analysis includes the 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 dichotomous item or the average proportion of the maximum score that students earned on each polytomous item. Point-biserial statistics are used to examine item-test correlations or item discrimination. Examining *p* values and point-biserial correlations can identify item flaws such as wrong keys. This procedure was used to check the operational data. 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

All Items	• Low *p* value (< 0.25 for dichotomous items and < 0.30 for polytomous items) • High *p* value (> 0.90 for dichotomous items and > 0.85 for polytomous items) • Low point-biserial correlation (< 0.20 for dichotomous items and < 0.40 for polytomous items) • High percent of students omitting the item (≥ 5%)
Dichotomous Items Only	• High percentage of students selecting distractor X (≥ 35%) • High point-biserial correlation for distractor X (> 0.05)
Polytomous Items Only	• High percentage of students earning 0 points (≥ 20%) • Low percentage of students earning the highest score point (≤ 5%)

The number of 2024 OP items flagged for each content area and grade is given in Table 5.26.

Table 5.26. Numbers of Flagged Items

			#Flagged Items	
Subject	**Grade**	**#Items**	***P* Value**	**Point-Biserial**
ELA	3	28	–	–
	4	29	–	–
	5	32	–	2
	6	32	–	–
	7	39	1	2
	8	39	–	–
Mathematics	3	32	1	–
	4	37	1	1
	5	37	5	–
	6	39	4	–
	7	41	1	1
	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 was not omitted) and to ensure that 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.

It is important to have a good range of *p* values to increase test reliability and 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 correlations to verify that items are functioning as intended.

The summary statistics of the item difficulty (*p* values) and item discrimination (point-biserial correlations) for the operational tests are shown in Table 5.27 and Table 5.28, respectively. The

data show a reasonably wide range of item difficulties for each test. The mean item difficulties ranged from 0.56 to 0.67 for ELA and 0.48 to 0.62 for mathematics.

Point-biserial correlations ranged from 0.14 to 0.67 for the ELA tests and 0.06 to 0.74 for the mathematics tests. The mean point-biserial correlations ranged from 0.38 to 0.46 for ELA and 0.45 to 0.49 for mathematics.

Table 5.27. Item Difficulty Distribution

Subject	Grade	N-Count	Mean	SD	Min	Max
ELA	3	164,051	0.56	0.11	0.38	0.83
	4	168,586	0.63	0.12	0.44	0.85
	5	158,764	0.56	0.11	0.34	0.81
	6	164,803	0.60	0.14	0.37	0.81
	7	159,817	0.63	0.13	0.30	0.91
	8	147,257	0.67	0.10	0.48	0.84
Mathematics	3	156,673	0.62	0.17	0.24	0.86
	4	157,499	0.62	0.14	0.32	0.91
	5	152,137	0.53	0.19	0.12	0.89
	6	154,141	0.52	0.16	0.20	0.91
	7	148,736	0.57	0.16	0.25	0.89
	8	90,895	0.48	0.16	0.11	0.79

Table 5.28. Item Discrimination Distribution

Subject	Grade	N-Count	Mean	SD	Min	Max
ELA	3	164,051	0.45	0.09	0.28	0.64
	4	168,586	0.44	0.11	0.28	0.67
	5	158,764	0.38	0.13	0.14	0.64
	6	164,803	0.40	0.13	0.20	0.66
	7	159,817	0.39	0.13	0.14	0.67
	8	147,257	0.46	0.10	0.26	0.66
Mathematics	3	156,673	0.48	0.11	0.28	0.69
	4	157,499	0.49	0.11	0.06	0.72
	5	152,137	0.46	0.10	0.24	0.70
	6	154,141	0.48	0.12	0.28	0.70
	7	148,736	0.45	0.11	0.15	0.68
	8	90,895	0.45	0.11	0.24	0.74

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

5.4.2. Omit Rates

Omit rates (i.e., the 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. The industry standard general

rule is that omit rates for multiple-choice items should be less than 5%; omit rates for 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. The psychometric phenomenon of DIF has been extensively investigated, and experts' judgments of bias were 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, 1992) for dichotomous items and the standardized mean difference method (Dorans et al., 1992) for polytomous items. Please refer to the *NYSTP 2024: ELA and Mathematics Grades 3–8 Field Test Technical Report* for details about these DIF methods and item-flagging 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

Item response theory (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 set of statistical models that attempt to relate observed responses to items on a test to latent traits. In the case of educational tests, the latent trait of interest is often students' mastery of a particular discipline, such as mathematics. Computer programs that implement IRT models use 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, two types of item parameters are estimated: the discrimination parameter and the difficulty parameter(s). 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.

Dichotomous and polytomous items differ in their number of score categories. To accommodate these differences, separate IRT models are used in item calibration. The two-parameter logistic (2PL) model (Lord, 1980; Lord & Novick, 1968) was adopted in 2023 for analyzing dichotomous items. In this model, the probability that a student with proficiency θ responds correctly to item i is:

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

where D is a scaling constant of 1.7, and a_i and b_i are the discrimination and difficulty parameters of item i, respectively.

For analysis of the polytomous items, the generalized partial-credit (GPC) model (Muraki, 1992) is used. In this model, the probability of an examinee with proficiency θ obtaining score k on item i is:

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

where:

- D, a_i, and b_i have been previously defined,
- m_i is the maximum number of score points of item i, and
- d_{ih} is the category parameter of item i on score h.

There are m_i -1 independent category parameters for item i because d_{i0} is usually arbitrarily set to 0, and the remaining category parameters satisfy $\sum_{k=1}^{m_i} d_{ik} = 0$.

6.2. Calibration Sample

The cleaned data were used to calibrate the NYSTP 2024 Grades 3–8 ELA and Mathematics Tests. Calibration sample sizes were adequate, as the calibration was performed using nearly all the New York State public and non-public school student population data in each grade. As shown in Tables 6.1–6.3 and Tables 6.4–6.6 for ELA and mathematics, respectively, the 2024 operational test samples were generally comparable to the 2023 populations in terms of student gender, race and ethnicity, NRC category, 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

Demographic Category		Grade 3		Grade 4	
		2023 Population	2024 Sample	2023 Population	2024 Sample
Gender	Female	49.96	49.82	49.62	50.05
	Male	50.04	50.17	50.37	49.94
	Non-Binary	–	–	–	–
Ethnicity	Asian	10.69	10.47	10.51	10.66
	African American	13.86	14.00	14.43	14.33
	Hispanic	27.06	27.60	27.35	27.63
	American Indian	0.71	0.77	0.73	0.70
	Multiracial	3.69	3.76	3.49	3.67
	Pacific Islander	0.19	0.19	0.22	0.19
	White	43.81	43.21	43.26	42.81
NRC	New York City	29.18	28.50	29.95	28.39
	Big 4 Cities	3.74	3.58	3.86	3.57
	Urban/Suburban	7.88	7.54	7.50	7.61
	Rural	5.77	5.57	5.67	5.58
	Average Needs	27.09	27.00	26.02	26.20
	Low Needs	12.69	12.62	12.02	12.34
	Charter	7.21	8.75	6.89	8.47
	Religious or Independent	6.44	6.44	8.09	7.85
SWD	No	86.07	85.46	85.58	85.36
	Yes	13.93	14.54	14.42	14.64
SUA	No	87.77	86.69	86.35	85.45
	Yes	12.23	13.31	13.65	14.55
ELL	No	86.84	86.72	88.51	87.94
	Yes	13.16	13.28	11.49	12.06
SWD/ SUA	No	89.49	88.81	88.47	87.92
	Yes	10.51	11.19	11.53	12.08
ELL/ SUA	No	98.31	98.12	98.26	98.09
	Yes	1.69	1.88	1.74	1.91

Table 6.2. ELA Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2023 Population	2024 Sample	2023 Population	2024 Sample
Gender	Female	49.54	49.44	49.33	49.29
	Male	50.45	50.55	50.65	50.69
	Non-Binary	–	–	–	–
Ethnicity	Asian	11.27	10.92	10.76	11.22
	African American	14.91	15.55	15.20	15.40
	Hispanic	27.32	29.05	27.70	28.28
	American Indian	0.73	0.77	0.71	0.75
	Multiracial	3.37	3.67	3.34	3.36
	Pacific Islander	0.21	0.23	0.23	0.21
	White	42.18	39.81	42.06	40.78
NRC	New York City	31.48	30.97	30.79	29.66
	Big 4 Cities	3.80	3.89	3.77	3.66
	Urban/Suburban	7.75	7.76	7.54	7.54
	Rural	5.78	5.76	5.79	5.62
	Average Needs	25.93	26.95	25.15	25.14
	Low Needs	12.17	12.48	11.89	11.77
	Charter	6.80	9.10	7.24	9.16
	Religious or Independent	6.30	3.09	7.83	7.46
SWD	No	85.22	84.60	85.87	85.35
	Yes	14.78	15.40	14.13	14.65
SUA	No	85.60	84.15	85.93	85.15
	Yes	14.40	15.85	14.07	14.85
ELL	No	89.07	91.26	89.85	90.12
	Yes	10.93	8.74	10.15	9.88
SWD/ SUA	No	87.77	87.02	88.58	87.99
	Yes	12.23	12.98	11.42	12.01
ELL/ SUA	No	98.12	98.09	98.16	98.31
	Yes	1.88	1.91	1.84	1.69

Table 6.3. ELA Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2023 Population	2024 Sample	2023 Population	2024 Sample
Gender	Female	49.07	49.23	48.71	48.21
	Male	50.90	50.74	51.25	51.75
	Non-Binary	–	–	–	–
Ethnicity	Asian	10.81	10.78	11.32	11.23
	African American	15.87	15.97	16.93	16.63
	Hispanic	28.78	28.75	28.93	30.45
	American Indian	0.72	0.73	0.71	0.74

Demographic Category		Grade 7		Grade 8	
		2023 Population	2024 Sample	2023 Population	2024 Sample
	Multiracial	3.01	3.26	2.81	3.10
	Pacific Islander	0.20	0.23	0.22	0.20
	White	40.61	40.28	39.08	37.65
NRC	New York City	32.46	30.87	34.64	33.58
	Big 4 Cities	3.90	3.68	4.23	4.15
	Urban/Suburban	7.51	7.45	7.85	8.04
	Rural	5.74	5.74	5.90	5.89
	Average Needs	23.65	24.39	22.95	24.55
	Low Needs	11.90	12.02	11.29	11.82
	Charter	7.80	9.01	7.70	8.67
	Religious or Independent	7.04	6.83	5.43	3.30
SWD	No	85.77	85.74	85.66	85.49
	Yes	14.23	14.26	14.34	14.51
SUA	No	85.97	85.22	86.15	84.69
	Yes	14.03	14.78	13.85	15.31
ELL	No	91.32	91.28	91.64	92.61
	Yes	8.68	8.72	8.36	7.39
SWD/ SUA	No	88.42	88.25	88.45	87.91
	Yes	11.58	11.75	11.55	12.09
ELL/ SUA	No	98.38	98.47	98.52	98.48
	Yes	1.62	1.53	1.48	1.52

Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2023 Population	2024 Sample	2023 Population	2024 Sample
Gender	Female	49.94	49.81	49.57	50.16
	Male	50.06	50.18	50.43	49.84
	Non-Binary	–	–	–	–
Ethnicity	Asian	10.86	10.76	10.68	10.97
	African American	14.08	14.27	14.62	14.59
	Hispanic	25.81	25.84	26.05	25.71
	American Indian	0.72	0.78	0.72	0.72
	Multiracial	3.73	3.78	3.54	3.72
	Pacific Islander	0.20	0.20	0.22	0.20
	White	44.61	44.37	44.17	44.11
NRC	New York City	28.74	28.39	29.33	28.72
	Big 4 Cities	3.71	3.68	3.87	3.38
	Urban/Suburban	7.45	6.99	7.19	6.89
	Rural	5.92	5.82	5.78	5.73
	Average Needs	27.38	26.20	26.32	25.67

Demographic Category		Grade 3		Grade 4	
		2023 Population	2024 Sample	2023 Population	2024 Sample
	Low Needs	12.92	13.05	12.37	12.55
	Charter	7.29	8.80	7.09	8.54
	Religious or Independent	6.59	7.07	8.04	8.53
SWD	No	86.61	86.01	86.20	85.98
	Yes	13.39	13.99	13.80	14.02
SUA	No	88.22	87.41	86.86	86.23
	Yes	11.78	12.59	13.14	13.77
ELL	No	87.92	88.07	89.56	89.32
	Yes	12.08	11.93	10.44	10.68
SWD/ SUA	No	89.94	89.43	89.04	88.60
	Yes	10.06	10.57	10.96	11.40
ELL/ SUA	No	98.41	98.42	98.30	98.36
	Yes	1.59	1.58	1.70	1.64

Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2023 Population	2024 Sample	2023 Population	2024 Sample
Gender	Female	49.40	49.19	49.23	49.41
	Male	50.60	50.80	50.76	50.57
	Non-Binary	–	–	–	–
Ethnicity	Asian	11.45	11.17	10.83	11.51
	African American	15.04	16.02	15.78	15.62
	Hispanic	26.19	26.84	26.10	25.82
	American Indian	0.74	0.80	0.72	0.77
	Multiracial	3.40	3.77	3.34	3.43
	Pacific Islander	0.21	0.24	0.24	0.22
	White	42.98	41.17	43.00	42.63
NRC	New York City	30.64	30.65	29.44	28.96
	Big 4 Cities	3.80	3.95	3.86	3.81
	Urban/Suburban	7.41	7.03	7.15	6.80
	Rural	5.80	5.89	5.86	5.81
	Average Needs	26.15	27.23	25.12	24.75
	Low Needs	12.42	12.85	12.24	12.30
	Charter	7.23	9.37	8.14	9.12
	Religious or Independent	6.55	3.04	8.20	8.45
SWD	No	86.09	84.51	86.78	86.30
	Yes	13.91	15.49	13.22	13.70
SUA	No	86.43	83.89	87.06	86.08
	Yes	13.57	16.11	12.94	13.92

Demographic Category		Grade 5		Grade 6	
		2023 Population	2024 Sample	2023 Population	2024 Sample
ELL	No	90.02	93.89	91.39	91.70
	Yes	9.98	6.11	8.61	8.30
SWD/ SUA	No	88.69	86.76	89.54	88.82
	Yes	11.31	13.24	10.46	11.18
ELL/ SUA	No	98.24	98.34	98.52	98.70
	Yes	1.76	1.66	1.48	1.30

Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2023 Population	2024 Sample	2023 Population	2024 Sample
Gender	Female	49.04	49.35	48.43	47.58
	Male	50.92	50.62	51.53	52.38
	Non-Binary	–	–	–	–
Ethnicity	Asian	10.87	10.99	7.96	8.04
	African American	16.22	16.26	18.31	18.75
	Hispanic	26.93	26.25	27.59	29.11
	American Indian	0.72	0.76	0.69	0.77
	Multiracial	3.04	3.30	2.80	3.23
	Pacific Islander	0.21	0.24	0.22	0.19
	White	42.01	42.20	42.44	39.91
NRC	New York City	30.94	29.88	29.20	28.92
	Big 4 Cities	4.03	3.84	5.26	5.36
	Urban/Suburban	6.97	6.72	7.88	8.69
	Rural	5.82	5.98	7.47	7.72
	Average Needs	24.04	24.11	24.61	26.45
	Low Needs	12.29	12.35	8.90	9.06
	Charter	8.24	9.08	8.39	9.18
	Religious or Independent	7.68	8.03	8.29	4.62
SWD	No	87.01	86.71	83.95	82.34
	Yes	12.99	13.29	16.05	17.66
SUA	No	87.43	86.32	84.79	81.62
	Yes	12.57	13.68	15.21	18.38
ELL	No	92.89	92.80	91.18	94.92
	Yes	7.11	7.20	8.82	5.08
SWD/ SUA	No	89.81	89.23	87.36	85.25
	Yes	10.19	10.77	12.64	14.75
ELL/ SUA	No	98.80	98.90	98.55	98.61
	Yes	1.20	1.10	1.45	1.39

6.2.1. Calibration Process

Item parameters were estimated using Scientific Software International (SSI) Inc.'s IRTPRO Version 6.0 (Cai et al., 2022) package. Dichotomous and polytomous items were calibrated simultaneously using marginal maximum likelihood procedures.

The calibration of NYSTP 2024 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 items were well within the prescribed parameter ranges except for a few, such as one Grade 4 mathematics item ($a = 0.11$). All calibration estimation results were reasonable for the Grades 3–8 ELA and Mathematics Tests. Table 6.7 and Table 6.8 present the summaries of the calibration results for ELA and mathematics, respectively. Additional details, including individual item parameter estimates, can be found in Appendix N in Tables N1–N12. The parameter estimates are expressed on the theta metric and are defined as:

- Dichotomous items:
  - a is a discrimination parameter.
  - b is a difficulty parameter.

- Polytomous items:
  - a is an item discrimination parameter.
  - b is an item location parameter.
  - d_k is the item category parameter for category k.

Table 6.7. ELA Calibration Results

Grade	N-Count	Range of *a*-Parameters		Range of *b*-Parameters	
3	164,051	0.38	1.03	-1.57	0.55
4	168,586	0.37	1.07	-1.90	0.34
5	158,764	0.18	0.95	-1.30	2.24
6	164,803	0.26	1.15	-1.41	1.03
7	159,817	0.19	1.20	-1.97	2.37
8	147,257	0.37	1.33	-1.74	0.14

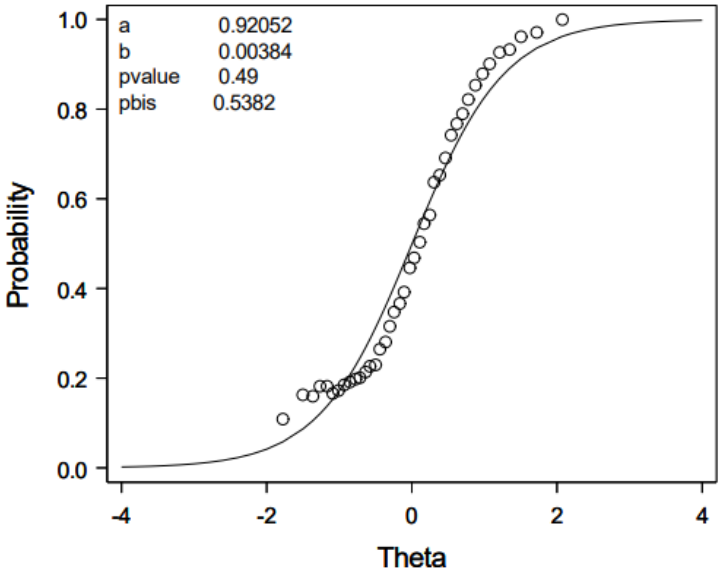
Table 6.8. Mathematics Calibration Results

Grade	N-Count	Range of *a*-Parameters		Range of *b*-Parameters	
3	156,673	0.41	1.40	-1.78	1.18
4	157,499	0.11	1.64	-1.75	0.92
5	152,137	0.35	1.43	-1.87	2.73
6	154,141	0.40	1.45	-2.02	1.36
7	148,736	0.22	1.22	-1.77	2.57
8	90,895	0.31	1.48	-1.03	2.23

6.3. Item-Model Fit

The *Standards for Educational and Psychological Testing* 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 by the examinees in different "buckets" (Sinharay, 2006). To make this judgment on each item, Hambleton and Swaminathan (1985) recommend the use of graphical plots comparing the estimated/predicted item-response function to the empirical student-response data for an item. An example item fit plot is shown in Figure 6.1.

Figure 6.1. Example Item Fit Plot



Fit plots were produced and closely examined for all operational items to visually examine the model-data fit for each item. All items showed adequate model-data fit except for one item in Grade 4 mathematics. This fact further supports the use of the chosen IRT models.

6.4. Equating

An IRT-based equating procedure was conducted for all ELA and mathematics tests to place the 2024 calibrated item parameter estimates on the base scale established in 2023. Multiple-choice items on a 2024 operational form that were field tested in an embedded setting in previous administrations were used as anchor items for equating. Anchor items consisted of 48–63% of the total score points and 59–75% of the operational items across subjects and grade levels. These substantial proportions of anchor items help stabilize equating results because a higher number of common items tends to result in a less random equating error (Budescu, 1985; Wingersky et al., 1987).

Once the item calibrations were completed, the anchor items were used to equate the form onto the base scale using a common-item nonequivalent groups design. The banked IRT parameters of the anchor items were used as reference parameters in the 2024 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) for the new and reference administrations. For a given item, *i*, the deviation called "*d*-squared" was calculated using the following equation:

$$d_i^2 = \sum_k \{[\Pr_{i,new}(\theta_k) - \Pr_{i,ref}(\theta_k)]^2 . g(\theta_k)\}$$

where i indexes anchor items, k indexes quadrature points for θ, $\Pr_{i,\text{ref}}(\cdot)$ is the probability of a correct response to item i under the previous administrations while $\Pr_{i,\text{new}}(\cdot)$ 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.

An anchor item with the largest $d_i^2 > 0.01$ would be flagged and considered for removal from an anchor set during an iterative equating process. Besides the d^2 criterion, the following item statistics were evaluated for anchor item stability: classical item statistics (*p* value and point-biserial correlations), item difficulty analysis, item fit evaluation, and ICC plot. After removing an anchor item, an impact analysis was also conducted based on equating results from the remaining anchor set. Following the stability checks and impact analysis, NWEA retained all anchor items for ELA tests and removed one anchor item from each mathematics test, with NYSED's approval.

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

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

$$b_i^E = (M_1^E b_i^C) + M_2^E$$

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

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	0.9369	-0.0839
4	1.0037	-0.1551

Grade	M_1^E	M_2^E
5	0.9878	-0.0825
6	0.9554	-0.1172
7	0.9930	-0.0268
8	1.0759	-0.0939

Table 6.10. Mathematics Equating Coefficients

Grade	M_1^E	M_2^E
3	0.9295	-0.0544
4	1.0869	0.1465
5	0.9897	0.0283
6	0.9861	0.0431
7	1.0520	0.1409
8	1.0226	0.0551

6.5. 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 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-score tables using a test characteristic curve (TCC) method. The obtained scaling transformation slope and intercept, M_1^S and M_2^S, from the 2023 standard setting 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.

The inverse TCC method was employed using POLYEQUATE (Kolen, 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 Grade 3–8 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 TCC is an excellent first-order approximation of the number of correct maximum likelihood estimates (MLEs). For tests with a mixture of MC and CR items, the MLE and TCC estimates are even more similar (Yen, 1984).

The inverse 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
- $\hat{\theta}$ is a trait estimate.

With the adoption of the 2PL model, the θ scores can be obtained for all raw score points, except for zero and perfect scores, using the inverse TCC method.

After the raw-score-to-theta-score conversion tables were produced, adjustments were made to the θ scores at the two ends of the scale based on the smoothing rules specified in Table 6.11. These adjustments were to smooth the unreliable θ scores with relatively large conditional standard errors of measurement (CSEMs).

Table 6.11. Smoothing Rules

Subject	Grade	Smoothing	
		Starting Point	Step Size
ELA	3–8	CSEM > 0.56	0.16
Math	3–8	CSEM > 0.44	0.12

At both ends of the scale, for any theta estimates with CSEMs greater than 0.56 for ELA and 0.44 for mathematics, 0.16 and 0.12 were subtracted (at the low end) or added (at the high end) from the preceding theta value for ELA and mathematics, respectively. Table 6.12 shows an example of smoothing at the two ends of the ELA and mathematics tests.

Table 6.12. Example of Smoothing in Raw-Score-to-Theta-Score Tables

ELA				Mathematics			
Raw Score	Estimated Theta	CSEM of Theta	Adjusted Theta	Raw Score	Estimated Theta	CSEM of Theta	Adjusted Theta
0	–	–	-3.4369	0	–	–	-2.5979
1	-4.4605	1.173	-3.2769	1	-3.1313	0.763	-2.4779
2	-3.6007	0.750	-3.1169	2	-2.5668	0.520	-2.3579
3	-3.1266	0.586	-2.9569	3	-2.2379	0.418	-2.2379
4	-2.7969	0.497	-2.7969	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	40	1.5144	0.399	1.5144
35	1.8043	0.517	1.8043	41	1.7576	0.462	1.6344
36	2.1182	0.582	1.9643	42	2.0993	0.573	1.7544
37	2.5189	0.684	2.1243	43	2.6909	0.834	1.8744
38	3.0882	0.874	2.2843	44	–	–	1.9944
39	4.1033	1.359	2.4443				
40	–	–	2.6043				

Note. Theta and CSEM values are not shown for zero and perfect scores because these values cannot be obtained using the inverse TCC method.

After smoothing the θ scores at the ends of the scale, the adjusted θ scores were obtained. The adjusted CSEMs were then computed. The scaling coefficients in Table 6.13 were then applied to the adjusted θ scores to obtain the corresponding scale scores using the equation below.

$$\text{Scale Score} = M_1^S\theta + M_2^S$$

The final raw-score-to-scale-score tables could then be established. Scaling coefficients, M_1^S and M_2^S, were determined during the 2023 standard setting and have been used in subsequent administrations. Note that comparing scale scores across tests of different subjects or grades is not appropriate, as each test has different content specifications and does not use the same scale.

Table 6.13. Operational Scaling Coefficients

Grade	Slope (M_1^S)	Intercept (M_2^S)
ELA		
3	22.27948	444.75274
4	21.80446	447.64882
5	21.68195	445.65776
6	22.11517	445.38191
7	22.04966	447.87882
8	22.26115	451.13510
Mathematics		
3	27.99291	453.39834
4	27.23092	453.38099
5	27.22817	451.23725
6	26.66825	451.10353
7	27.38364	454.35838
8	26.49501	446.10550

6.5.1. Raw-Score-to-Scale-Score Conversion Tables, TCCs, CSEMs, and Performance Levels

The scale score is the reported score for the NYSTP. The raw-score-to-scale-score (RSSS) conversion tables are presented in Appendix P, Tables P1–P12.

Test characteristic curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2024 TCCs were generated using final item parameters for all reporting test items administered in Spring 2024. TCCs are the summation of all the item characteristic curves (ICCs) contributing to the scale scores. The TCC plots for ELA and mathematics tests are presented in Appendix Q.

The CSEM of a scale score is calculated as follows and is included in the RSSS table:

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

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^S is the scaling coefficient in Table 6.13.

Scale score cuts were set in the summer of 2023 through standard setting and can be applied to the 2024 scale scores because the 2024 scale scores are on the 2023 base scale after equating and scaling. See Section 8 and Appendix Q of the 2023 operational technical report for more information on the standards review process.

The following procedure is conducted on an RSSS table to ensure that all cut scores are obtainable: If no rounded scale score matches a given scale score cut, the nearest available score below the cut is adjusted to match the cut score. For example, if the cut score of interest is 450 and only scale scores of 449 and 451 are obtainable (before adjustment), then the scale score of 449 would be adjusted to 450, and the scale score of 451 would remain unaltered.
The final element of the raw-score-to-scale-score tables is the application of the performance level cut scores.

Table 6.14 and Table 6.15 present scale score ranges associated with each performance level for ELA and mathematics, respectively.

Table 6.14. ELA Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	383–431	432–449	450–473	474–494
4	371–430	431–449	450–470	471–493
5	373–431	432–449	450–473	474–504
6	368–430	431–449	450–469	470–498
7	363–432	433–449	450–471	472–502
8	363–429	430–449	450–471	472–500

Table 6.15. Mathematics Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	381–423	424–449	450–486	487–505
4	376–430	431–449	450–485	486–512
5	384–431	432–449	450–482	483–521
6	387–430	431–449	450–484	485–517
7	378–429	430–449	450–476	477–523
8	382–435	436–449	450–481	482–521

6.6. CSEMs

Conditional standard error of measurement (CSEM) curves graphically show the amount of measurement error at different ability levels. The CSEM curves are presented in Figures 6.2–6.13.

Figure 6.2. ELA Grade 3 CSEM Curve

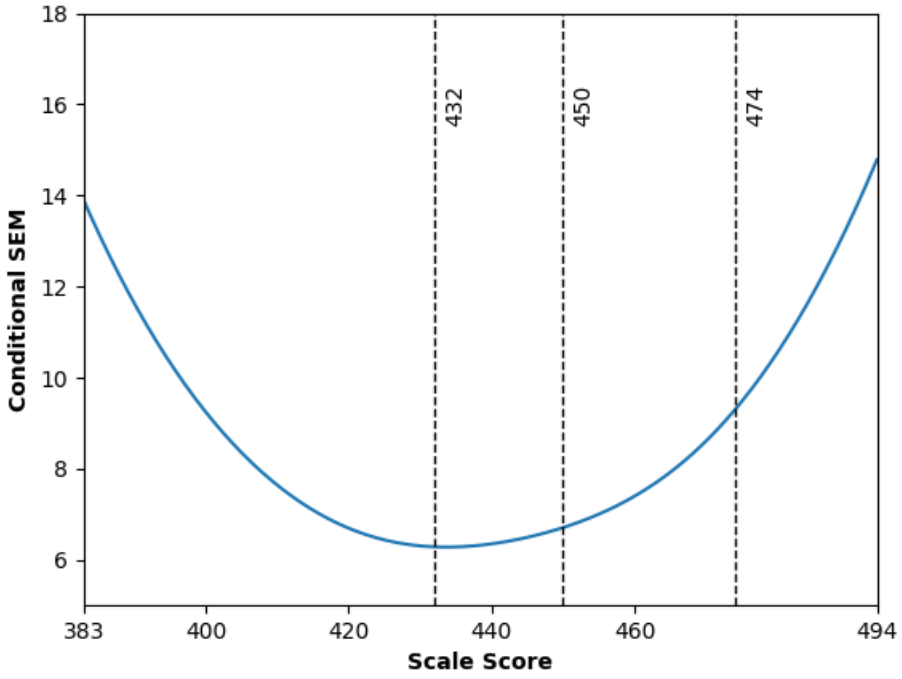


Figure 6.3. ELA Grade 4 CSEM Curve

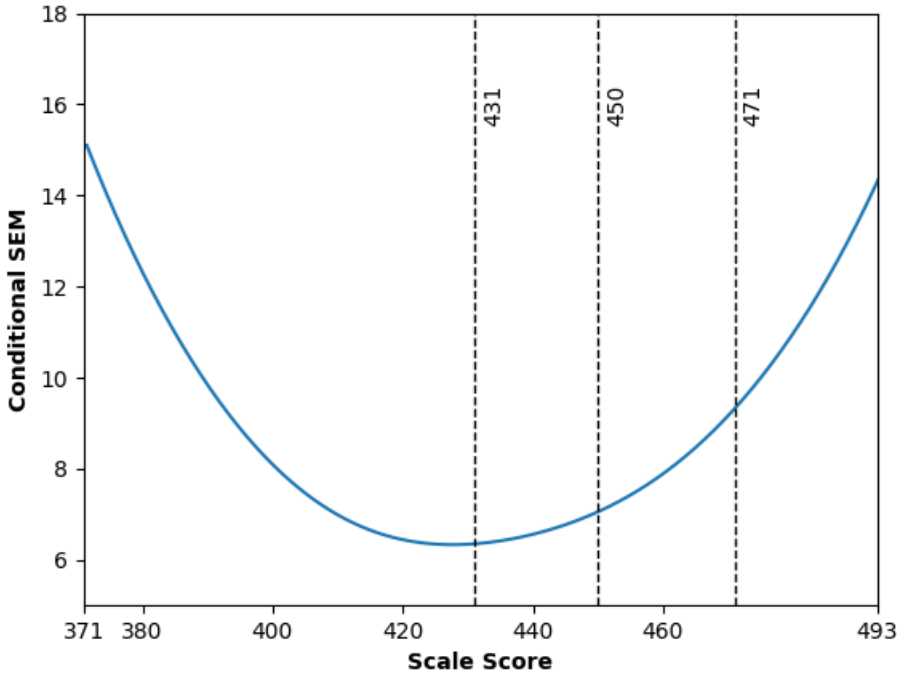


Figure 6.4. ELA Grade 5 CSEM Curve

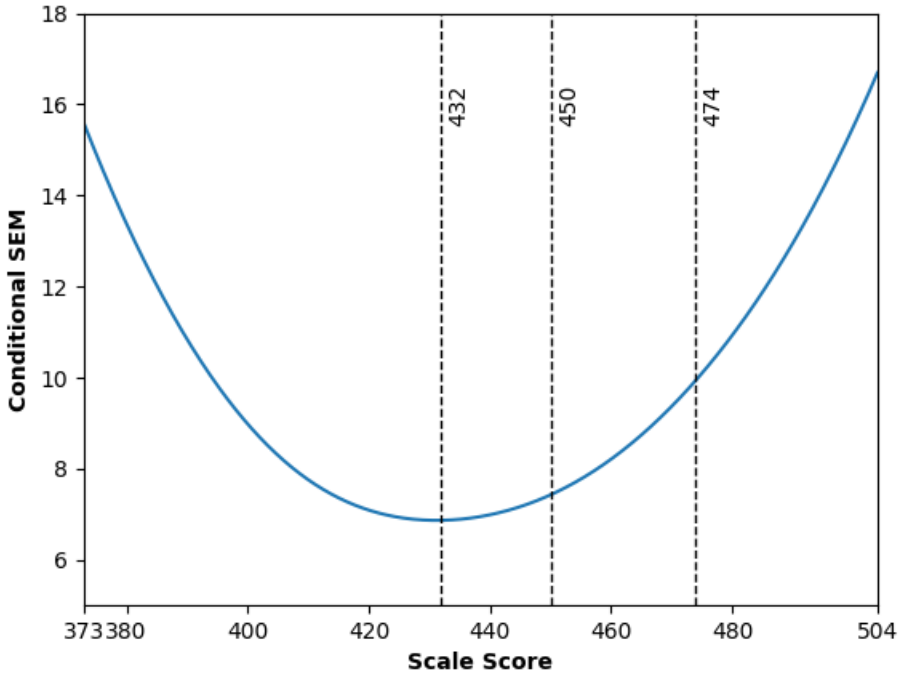


Figure 6.5. ELA Grade 6 CSEM Curve

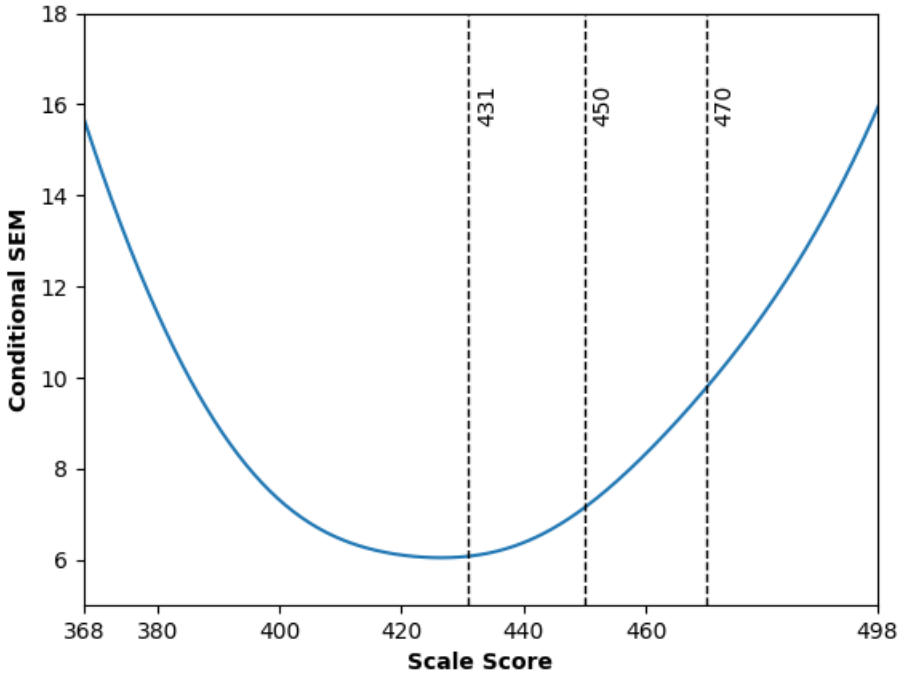


Figure 6.6. ELA Grade 7 CSEM Curve

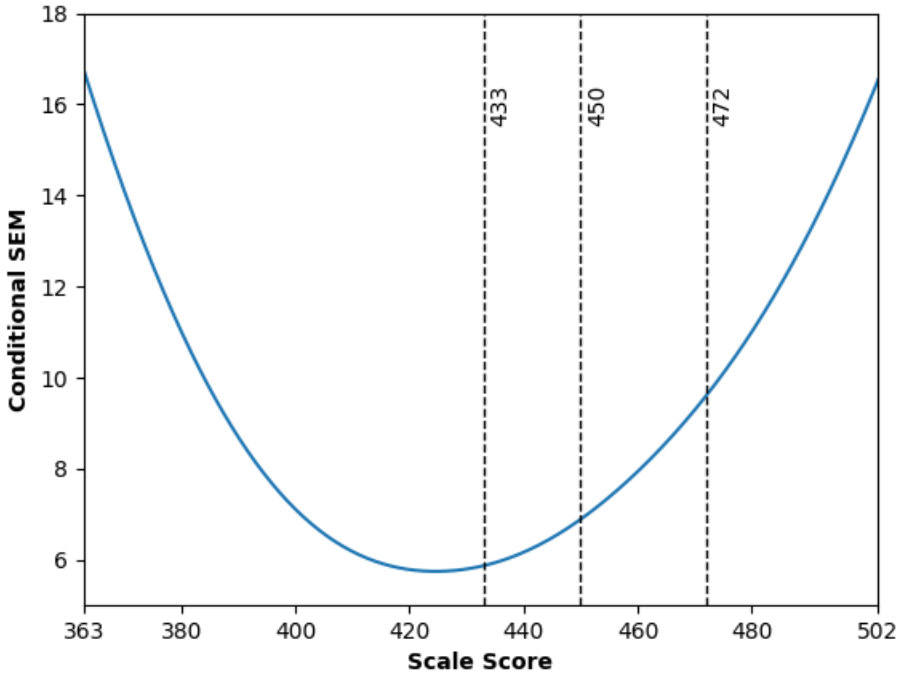


Figure 6.7. ELA Grade 8 CSEM Curve

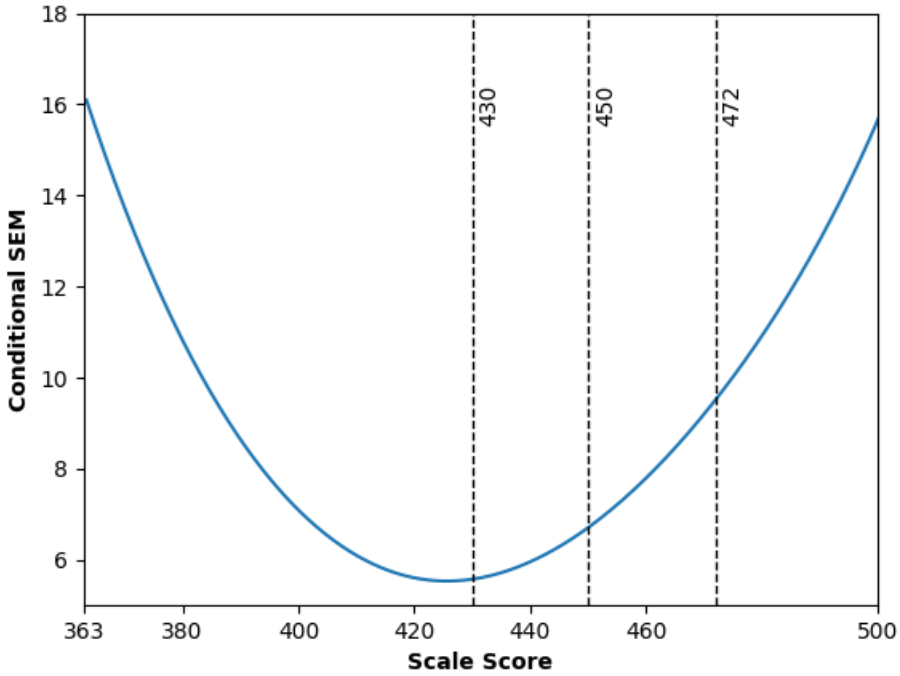


Figure 6.8. Mathematics Grade 3 CSEM Curve

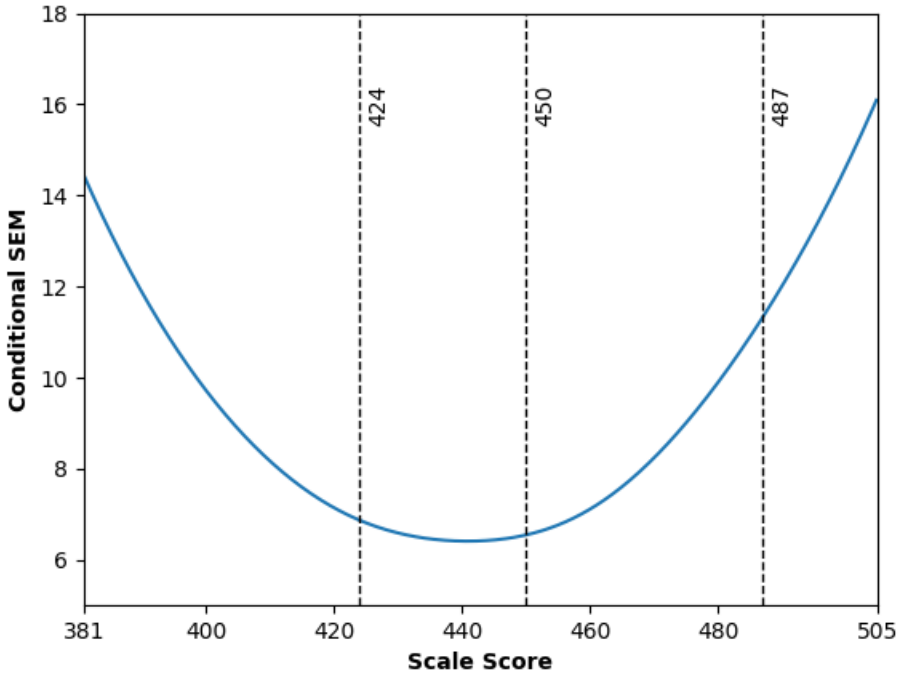


Figure 6.9. Mathematics Grade 4 CSEM Curve

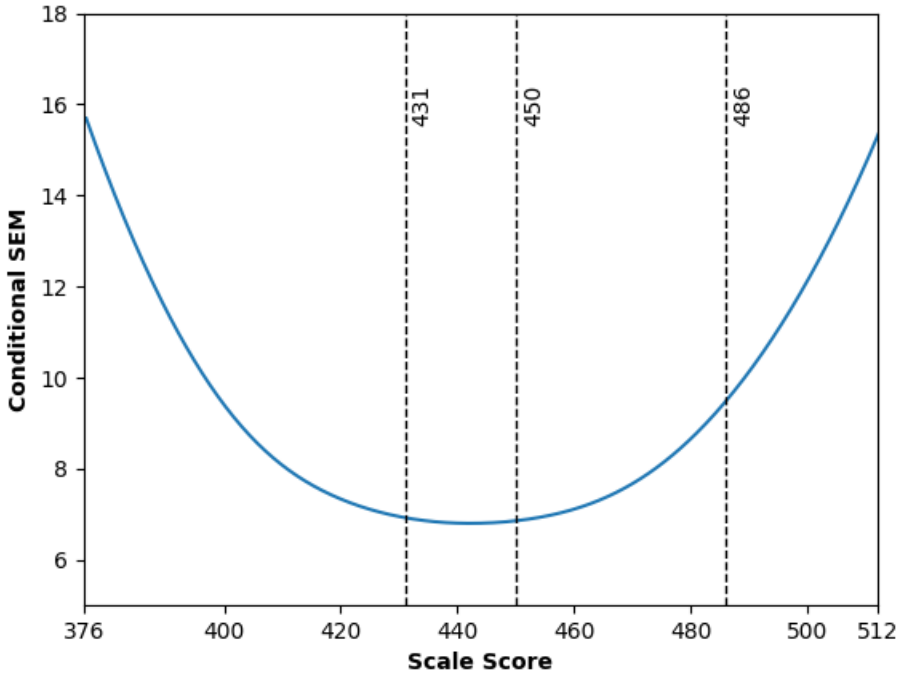


Figure 6.10. Mathematics Grade 5 CSEM Curve

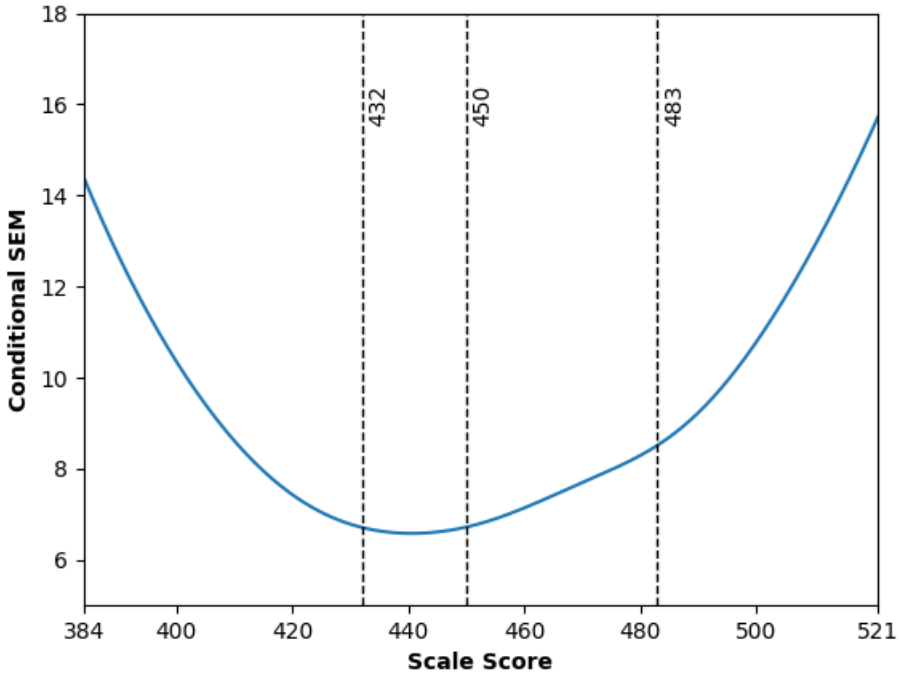


Figure 6.11. Mathematics Grade 6 CSEM Curve

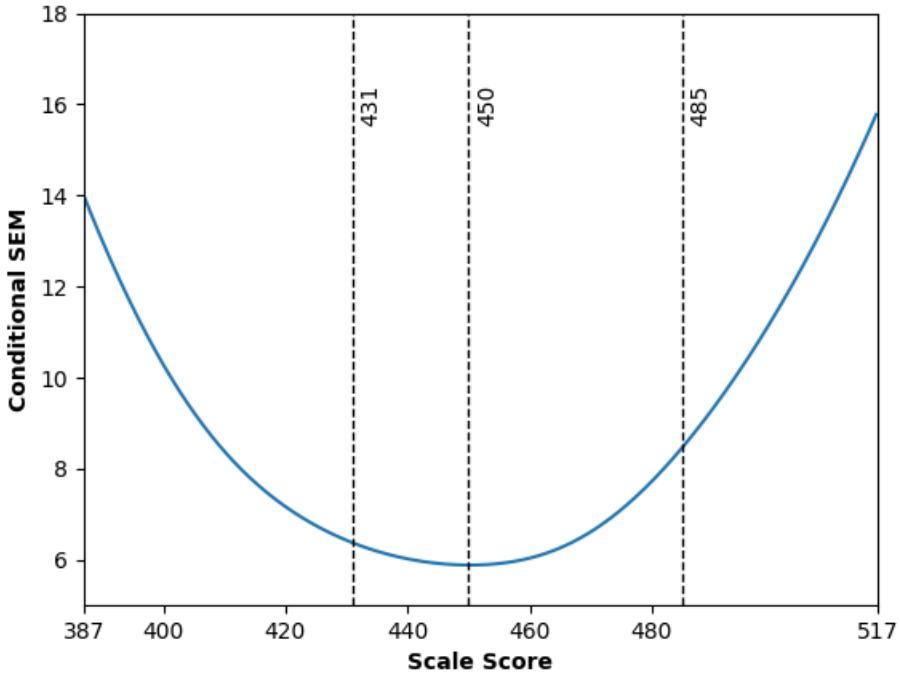


Figure 6.12. Mathematics Grade 7 CSEM Curve

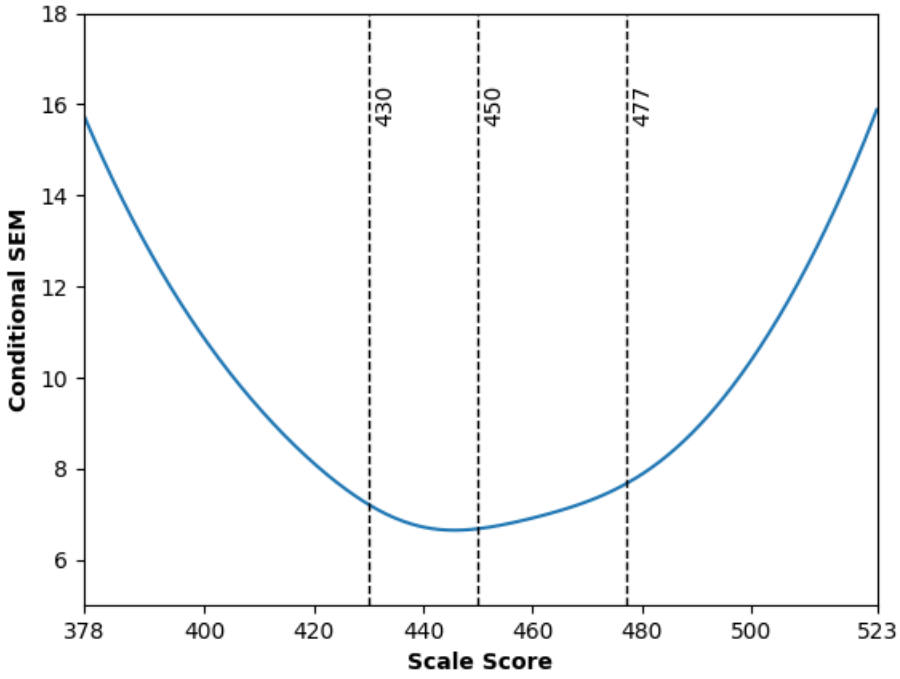
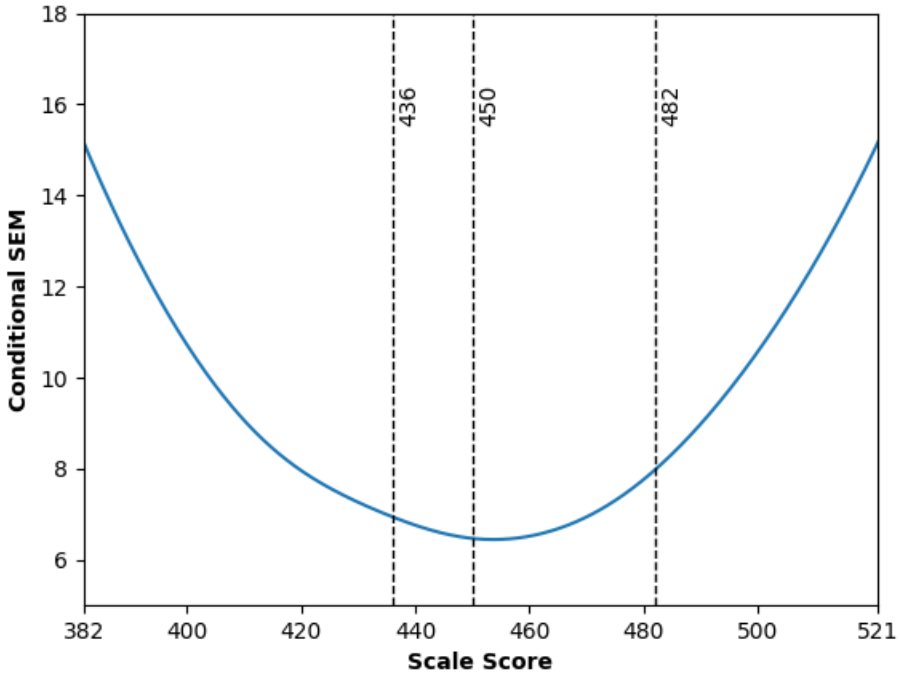


Figure 6.13. Mathematics Grade 8 CSEM Curve



Section 7: Reliability and Standard Error of Measurement

This section presents specific information on various test reliability statistics, standard errors of measurement (SEMs), and the results of performance level classification accuracy and consistency analyses. The data set for these analyses includes New York State students who were tested and 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, and the alpha statistic can be used to derive the SEM. For the Grades 3–8 ELA and Mathematics Tests, NWEA calculated two types of reliability statistics: Cronbach's alpha (Cronbach, 1951) and the Feldt-Raju coefficient (Qualls, 1995). These two measures are appropriate for assessing the internal consistency of a test 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 reflecting 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 the Feldt-Raju coefficient measures are appropriate for tests consisting 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. Table 7.2 and Table 7.4 present the case counts ("N-Count"), number of test items ("#Items"), Cronbach's alpha and associated SEM, and the 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. The total 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 coefficient) ranged from 0.87 to 0.92. Grades 3–8 mathematics reliability estimates (Cronbach's alpha and Feldt-Raju coefficient) 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 0.87 across all grades and both subjects, which is a good indication that the NYSTP Grades 3–8 ELA and Mathematics Tests are acceptably reliable.

Table 7.1. ELA Test Form Statistics

	Item-Level			Student-Level			
	P Value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.56	0.38	0.83	164,051	33	18.35	7.52
4	0.63	0.44	0.85	168,586	37	22.59	7.91
5	0.56	0.34	0.81	158,764	40	22.44	7.88
6	0.60	0.37	0.81	164,803	40	24.04	7.98
7	0.63	0.30	0.91	159,817	47	29.94	9.00

Grade	Item-Level			Student-Level			
	P Value				Raw Score		
	Mean	Min.	Max.	N-Count	Max.	Mean	SD
8	0.67	0.48	0.84	147,257	47	31.19	9.96

Table 7.2. ELA Test Reliability and Standard Error of Measurement

Grade	N-Count	#Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	164,051	28	33	0.89	2.48	0.90	2.42
4	168,586	29	37	0.89	2.63	0.90	2.50
5	158,764	32	40	0.87	2.87	0.88	2.76
6	164,803	32	40	0.88	2.75	0.89	2.64
7	159,817	39	47	0.89	2.97	0.90	2.86
8	147,257	39	47	0.92	2.85	0.92	2.76

Table 7.3. Mathematics Test Form Statistics

Grade	Item-Level			Student-Level			
	P Value				Raw Score		
	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.62	0.24	0.86	156,673	38	22.13	9.00
4	0.62	0.32	0.91	157,499	44	25.90	10.60
5	0.53	0.12	0.89	152,137	44	21.67	9.90
6	0.52	0.20	0.91	154,141	47	23.36	11.50
7	0.57	0.25	0.89	148,736	49	27.15	11.30
8	0.48	0.11	0.79	90,895	49	22.25	11.20

Table 7.4. Mathematics Test Reliability and Standard Error of Measurement

Grade	N-Count	#Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	156,673	32	38	0.91	2.69	0.92	2.55
4	157,499	37	44	0.92	2.91	0.93	2.76
5	152,137	37	44	0.92	2.86	0.92	2.75
6	154,141	39	47	0.93	3.12	0.93	2.94
7	148,736	41	49	0.92	3.18	0.93	3.03
8	90,895	41	49	0.92	3.16	0.93	3.00

7.1.2. Reliability of MC Items

In addition to overall test reliability, Cronbach's alpha and the 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	164,051	23	0.85	2.07	0.85	2.07
4	168,586	23	0.83	2.02	0.83	2.01
5	158,764	26	0.80	2.30	0.80	2.29
6	164,803	26	0.82	2.24	0.82	2.23
7	159,817	33	0.84	2.49	0.84	2.49
8	147,257	33	0.89	2.37	0.89	2.36

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	156,673	24	0.87	1.96	0.87	1.94
4	157,499	28	0.89	2.13	0.89	2.11
5	152,137	28	0.88	2.20	0.88	2.19
6	154,141	29	0.89	2.26	0.89	2.24
7	148,736	31	0.87	2.37	0.87	2.36
8	90,895	31	0.88	2.44	0.88	2.43

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	164,051	5	10	0.83	1.15	0.83	1.14
4	168,586	6	14	0.84	1.43	0.86	1.37
5	158,764	6	14	0.83	1.46	0.84	1.42
6	164,803	6	14	0.85	1.32	0.86	1.27
7	159,817	6	14	0.85	1.32	0.87	1.25
8	147,257	6	14	0.85	1.34	0.86	1.28

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	156,673	8	14	0.82	1.72	0.84	1.63
4	157,499	9	16	0.85	1.84	0.86	1.74
5	152,137	9	16	0.81	1.75	0.83	1.66
6	154,141	10	18	0.85	2.00	0.87	1.88
7	148,736	10	18	0.86	1.96	0.87	1.87
8	90,895	10	18	0.85	1.89	0.87	1.73

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 subgroups include the following: gender, ethnicity, Needs Resource Capacity (NRC) category, English Language Learner (ELL) status, all students with disabilities (SWDs), all students using test accommodations (SUAs), 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 Flexibility in Scheduling/Timing, Flexibility in Setting, Method of Presentation (excluding braille), Method of Response, Braille and Large-type, and others (IEP or 504 Plan). Accommodations available to ELLs are Separate Location and Bilingual Dictionary.

As shown in Tables 7.9–7.14 and Tables 7.15–7.20 for ELA and mathematics, respectively, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Except for the non-binary group, Cronbach's alpha reliability coefficients were all at least 0.74, and the Feldt-Raju reliability coefficients, which tend to be larger than Cronbach's alpha estimates for the same group, were at least 0.76. These indicate a very good internal test consistency (reliability) for the 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	164,051	0.89	2.48	0.90	2.42
Gender	Female	81,734	0.89	2.47	0.89	2.42
	Male	82,297	0.89	2.48	0.90	2.43
	Non-Binary	20	0.93	2.20	0.94	2.01
Ethnicity	Asian	17,152	0.89	2.37	0.90	2.33
	African American	22,924	0.88	2.52	0.89	2.47
	Hispanic	45,209	0.88	2.52	0.89	2.47
	American Indian	1,264	0.88	2.52	0.89	2.46
	Multiracial	6,153	0.90	2.41	0.91	2.36
	Pacific Islander	317	0.89	2.45	0.89	2.41
	White	70,762	0.89	2.45	0.89	2.40
NRC	New York City	46,749	0.90	2.46	0.91	2.39
	Big 4 Cities	5,881	0.88	2.54	0.89	2.46
	Urban/Suburban	12,363	0.87	2.53	0.87	2.48
	Rural	9,132	0.87	2.49	0.88	2.45
	Average Needs	44,294	0.88	2.46	0.88	2.42
	Low Needs	20,698	0.87	2.37	0.88	2.34
	Charter	14,362	0.88	2.48	0.89	2.44
	Religious or Independent	10,572	0.89	2.50	0.90	2.44
SWD	All Codes	23,852	0.86	2.53	0.86	2.48
SUA	All Codes	21,840	0.85	2.52	0.86	2.48
ELL	ELL=Y	21,788	0.82	2.56	0.83	2.50

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SWD/SUA	SWD & SUA codes	18,361	0.84	2.52	0.85	2.48
ELL/SUA	SUA & ELL codes	3,078	0.79	2.52	0.80	2.47

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	168,586	0.89	2.63	0.90	2.50
Gender	Female	84,385	0.89	2.61	0.90	2.49
	Male	84,192	0.89	2.64	0.90	2.51
	Non-Binary	9	–	–	–	–
Ethnicity	Asian	17,945	0.88	2.46	0.89	2.35
	African American	24,114	0.89	2.66	0.90	2.54
	Hispanic	46,493	0.88	2.67	0.89	2.55
	American Indian	1,177	0.88	2.65	0.89	2.52
	Multiracial	6,177	0.89	2.59	0.90	2.46
	Pacific Islander	328	0.88	2.63	0.89	2.52
	White	72,035	0.89	2.61	0.90	2.49
NRC	New York City	47,861	0.90	2.58	0.91	2.44
	Big 4 Cities	6,012	0.90	2.74	0.91	2.57
	Urban/Suburban	12,831	0.87	2.70	0.88	2.60
	Rural	9,412	0.87	2.68	0.88	2.59
	Average Needs	44,176	0.87	2.62	0.88	2.53
	Low Needs	20,796	0.85	2.51	0.86	2.42
	Charter	14,271	0.87	2.54	0.88	2.46
	Religious or Independent	13,227	0.91	2.69	0.92	2.52
SWD	All Codes	24,674	0.87	2.72	0.88	2.62
SUA	All Codes	24,531	0.86	2.71	0.87	2.62
ELL	ELL=Y	20,331	0.83	2.74	0.85	2.62
SWD/SUA	SWD & SUA codes	20,358	0.86	2.71	0.86	2.62
ELL/SUA	SUA & ELL codes	3,227	0.80	2.70	0.81	2.62

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	158,764	0.87	2.87	0.88	2.76
Gender	Female	78,487	0.86	2.85	0.87	2.75
	Male	80,259	0.87	2.88	0.88	2.77
	Non-Binary	18	0.84	2.88	0.86	2.66

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Ethnicity	Asian	17,272	0.86	2.73	0.87	2.64
	African American	24,605	0.86	2.92	0.87	2.81
	Hispanic	45,947	0.86	2.92	0.87	2.80
	American Indian	1,225	0.86	2.91	0.87	2.80
	Multiracial	5,807	0.88	2.85	0.89	2.73
	Pacific Islander	365	0.88	2.84	0.89	2.74
	White	62,968	0.86	2.83	0.86	2.74
NRC	New York City	49,162	0.88	2.86	0.89	2.74
	Big 4 Cities	6,170	0.87	2.93	0.88	2.78
	Urban/Suburban	12,328	0.85	2.91	0.86	2.81
	Rural	9,137	0.85	2.88	0.86	2.80
	Average Needs	42,790	0.86	2.86	0.86	2.77
	Low Needs	19,814	0.84	2.76	0.85	2.68
	Charter	14,451	0.85	2.86	0.86	2.78
	Religious or Independent	4,912	0.84	2.84	0.85	2.75
SWD	All Codes	24,451	0.84	2.91	0.85	2.80
SUA	All Codes	25,158	0.84	2.90	0.85	2.80
ELL	ELL=Y	13,869	0.76	2.87	0.78	2.79
SWD/SUA	SWD & SUA codes	20,614	0.83	2.90	0.84	2.80
ELL/SUA	SUA & ELL codes	3,026	0.74	2.83	0.76	2.76

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	164,803	0.88	2.75	0.89	2.64
Gender	Female	81,231	0.88	2.70	0.88	2.61
	Male	83,534	0.88	2.77	0.89	2.66
	Non-Binary	38	0.87	2.59	0.88	2.46
Ethnicity	Asian	18,447	0.87	2.56	0.88	2.48
	African American	25,323	0.87	2.78	0.88	2.68
	Hispanic	46,506	0.87	2.79	0.88	2.68
	American Indian	1,226	0.87	2.77	0.88	2.66
	Multiracial	5,522	0.89	2.70	0.90	2.60
	Pacific Islander	353	0.87	2.72	0.88	2.63
	White	67,055	0.88	2.73	0.89	2.63
NRC	New York City	48,873	0.89	2.72	0.90	2.60
	Big 4 Cities	6,033	0.88	2.85	0.89	2.72
	Urban/Suburban	12,418	0.87	2.81	0.88	2.71
	Rural	9,255	0.86	2.78	0.87	2.70

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Average Needs	41,430	0.86	2.74	0.87	2.66
	Low Needs	19,404	0.85	2.60	0.85	2.54
	Charter	15,097	0.87	2.67	0.87	2.60
	Religious or Independent	12,293	0.91	2.84	0.92	2.66
SWD	All Codes	24,148	0.85	2.84	0.86	2.74
SUA	All Codes	24,472	0.86	2.84	0.87	2.74
ELL	ELL=Y	16,277	0.80	2.86	0.81	2.75
SWD/SUA	SWD & SUA codes	19,799	0.84	2.84	0.85	2.75
ELL/SUA	SUA & ELL codes	2,790	0.75	2.81	0.77	2.74

Table 7.13. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	159,817	0.89	2.97	0.90	2.86
Gender	Female	78,684	0.88	2.92	0.89	2.83
	Male	81,087	0.90	3.00	0.90	2.88
	Non-Binary	46	0.84	2.87	0.86	2.74
Ethnicity	Asian	17,187	0.88	2.69	0.88	2.61
	African American	25,462	0.88	3.02	0.89	2.92
	Hispanic	45,833	0.89	3.03	0.89	2.92
	American Indian	1,165	0.89	3.04	0.90	2.91
	Multiracial	5,200	0.90	2.92	0.90	2.81
	Pacific Islander	360	0.89	2.87	0.90	2.77
	White	64,224	0.89	2.95	0.90	2.84
NRC	New York City	49,343	0.89	2.91	0.90	2.81
	Big 4 Cities	5,889	0.89	3.11	0.90	2.97
	Urban/Suburban	11,911	0.88	3.06	0.89	2.96
	Rural	9,168	0.87	3.03	0.88	2.95
	Average Needs	38,975	0.88	2.97	0.88	2.88
	Low Needs	19,215	0.86	2.78	0.87	2.72
	Charter	14,398	0.86	2.88	0.87	2.82
	Religious or Independent	10,918	0.92	3.05	0.93	2.86
SWD	All Codes	22,789	0.87	3.13	0.88	3.03
SUA	All Codes	23,626	0.88	3.12	0.88	3.01
ELL	ELL=Y	13,941	0.84	3.15	0.85	3.03
SWD/SUA	SWD & SUA codes	18,775	0.86	3.13	0.87	3.04
ELL/SUA	SUA & ELL codes	2,445	0.79	3.11	0.80	3.04

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	147,257	0.92	2.85	0.92	2.76
Gender	Female	70,999	0.92	2.78	0.92	2.70
	Male	76,200	0.92	2.90	0.92	2.80
	Non-Binary	58	0.87	2.65	0.89	2.50
Ethnicity	Asian	16,461	0.91	2.54	0.92	2.46
	African American	24,382	0.91	2.96	0.92	2.86
	Hispanic	44,656	0.92	2.94	0.92	2.85
	American Indian	1,084	0.91	2.92	0.92	2.82
	Multiracial	4,540	0.92	2.83	0.93	2.71
	Pacific Islander	299	0.92	2.80	0.92	2.71
	White	55,213	0.91	2.79	0.92	2.71
NRC	New York City	49,452	0.92	2.82	0.93	2.73
	Big 4 Cities	6,104	0.92	3.05	0.92	2.94
	Urban/Suburban	11,837	0.92	3.00	0.92	2.89
	Rural	8,677	0.91	2.95	0.91	2.87
	Average Needs	36,154	0.91	2.86	0.92	2.77
	Low Needs	17,401	0.90	2.63	0.91	2.55
	Charter	12,770	0.90	2.79	0.90	2.73
	Religious or Independent	4,862	0.90	2.73	0.91	2.65
SWD	All Codes	21,371	0.89	3.08	0.90	3.00
SUA	All Codes	22,544	0.90	3.06	0.91	2.97
ELL	ELL=Y	10,876	0.82	3.07	0.83	3.00
SWD/SUA	SWD & SUA codes	17,809	0.89	3.09	0.89	3.01
ELL/SUA	SUA & ELL codes	2,242	0.80	3.09	0.81	3.03

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	156,673	0.91	2.69	0.92	2.55
Gender	Female	78,042	0.91	2.71	0.91	2.58
	Male	78,614	0.92	2.66	0.92	2.53
	Non-Binary	17	0.93	2.65	0.94	2.37
Ethnicity	Asian	16,825	0.90	2.53	0.92	2.37
	African American	22,311	0.92	2.67	0.92	2.55
	Hispanic	40,407	0.91	2.70	0.91	2.58
	American Indian	1,224	0.91	2.70	0.92	2.55
	Multiracial	5,906	0.92	2.66	0.93	2.51
	Pacific Islander	315	0.91	2.67	0.92	2.53

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	White	69,378	0.90	2.69	0.91	2.57
NRC	New York City	44,475	0.92	2.67	0.93	2.51
	Big 4 Cities	5,767	0.91	2.61	0.92	2.52
	Urban/Suburban	10,953	0.90	2.68	0.91	2.59
	Rural	9,123	0.90	2.67	0.91	2.58
	Average Needs	41,041	0.90	2.69	0.91	2.58
	Low Needs	20,444	0.88	2.63	0.89	2.50
	Charter	13,793	0.91	2.59	0.92	2.43
	Religious or Independent	11,077	0.90	2.77	0.90	2.68
SWD	All Codes	21,919	0.91	2.63	0.92	2.54
SUA	All Codes	19,730	0.90	2.62	0.91	2.55
ELL	ELL=Y	18,695	0.90	2.71	0.90	2.63
SWD/SUA	SWD & SUA codes	16,559	0.90	2.60	0.91	2.54
ELL/SUA	SUA & ELL codes	2,471	0.90	2.61	0.90	2.54

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	157,499	0.92	2.91	0.93	2.76
Gender	Female	78,997	0.92	2.93	0.93	2.78
	Male	78,493	0.93	2.88	0.94	2.72
	Non-Binary	9	–	–	–	–
Ethnicity	Asian	17,234	0.92	2.67	0.93	2.49
	African American	22,925	0.93	2.91	0.93	2.77
	Hispanic	40,404	0.92	2.94	0.93	2.80
	American Indian	1,125	0.92	2.93	0.93	2.77
	Multiracial	5,845	0.93	2.88	0.94	2.70
	Pacific Islander	315	0.93	2.86	0.94	2.67
	White	69,324	0.92	2.92	0.92	2.78
NRC	New York City	45,238	0.93	2.87	0.94	2.69
	Big 4 Cities	5,319	0.93	2.83	0.93	2.73
	Urban/Suburban	10,848	0.92	2.91	0.92	2.80
	Rural	9,022	0.91	2.93	0.92	2.82
	Average Needs	40,435	0.91	2.92	0.92	2.80
	Low Needs	19,761	0.90	2.81	0.91	2.66
	Charter	13,447	0.93	2.82	0.94	2.65
	Religious or Independent	13,429	0.92	2.98	0.92	2.88
SWD	All Codes	22,082	0.92	2.85	0.92	2.76
SUA	All Codes	21,687	0.91	2.85	0.92	2.77

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL	ELL=Y	16,817	0.90	2.92	0.91	2.84
SWD/SUA	SWD & SUA codes	17,949	0.91	2.83	0.91	2.75
ELL/SUA	SUA & ELL codes	2,584	0.89	2.79	0.89	2.74

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	152,137	0.92	2.86	0.92	2.75
Gender	Female	74,837	0.91	2.85	0.92	2.76
	Male	77,282	0.92	2.86	0.93	2.75
	Non-Binary	18	0.93	2.90	0.94	2.76
Ethnicity	Asian	16,911	0.92	2.79	0.93	2.64
	African American	24,261	0.91	2.81	0.92	2.73
	Hispanic	40,652	0.91	2.84	0.91	2.76
	American Indian	1,208	0.92	2.83	0.92	2.73
	Multiracial	5,715	0.92	2.86	0.93	2.75
	Pacific Islander	357	0.92	2.84	0.92	2.74
	White	62,360	0.91	2.87	0.91	2.77
NRC	New York City	46,627	0.93	2.84	0.93	2.73
	Big 4 Cities	6,008	0.90	2.74	0.90	2.66
	Urban/Suburban	10,694	0.90	2.81	0.90	2.73
	Rural	8,962	0.89	2.83	0.90	2.76
	Average Needs	41,421	0.90	2.86	0.91	2.78
	Low Needs	19,552	0.90	2.86	0.91	2.75
	Charter	14,252	0.92	2.85	0.93	2.74
	Religious or Independent	4,621	0.91	2.90	0.91	2.80
SWD	All Codes	23,568	0.90	2.73	0.90	2.66
SUA	All Codes	24,511	0.89	2.74	0.90	2.67
ELL	ELL=Y	9,295	0.87	2.72	0.87	2.66
SWD/SUA	SWD & SUA codes	20,144	0.88	2.71	0.89	2.65
ELL/SUA	SUA & ELL codes	2,530	0.84	2.65	0.85	2.60

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	154,141	0.93	3.12	0.93	2.94
Gender	Female	76,158	0.92	3.14	0.93	2.96
	Male	77,951	0.93	3.09	0.94	2.92

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Non-Binary	32	0.91	3.20	0.92	2.98
Ethnicity	Asian	17,696	0.93	2.96	0.94	2.76
	African American	24,014	0.92	3.08	0.93	2.93
	Hispanic	39,688	0.91	3.12	0.92	2.97
	American Indian	1,188	0.91	3.13	0.92	2.97
	Multiracial	5,277	0.93	3.09	0.94	2.89
	Pacific Islander	340	0.92	3.15	0.93	2.95
	White	65,534	0.92	3.13	0.93	2.97
NRC	New York City	44,645	0.93	3.12	0.94	2.91
	Big 4 Cities	5,869	0.91	2.98	0.92	2.84
	Urban/Suburban	10,486	0.91	3.03	0.92	2.91
	Rural	8,956	0.90	3.07	0.91	2.95
	Average Needs	38,144	0.91	3.13	0.92	2.98
	Low Needs	18,956	0.91	3.08	0.92	2.92
	Charter	14,053	0.93	3.10	0.94	2.92
	Religious or Independent	13,032	0.92	3.09	0.92	2.95
SWD	All Codes	21,121	0.90	2.90	0.90	2.81
SUA	All Codes	21,459	0.90	2.91	0.91	2.82
ELL	ELL=Y	12,794	0.86	2.89	0.87	2.83
SWD/SUA	SWD & SUA codes	17,226	0.88	2.86	0.89	2.78
ELL/SUA	SUA & ELL codes	2,008	0.84	2.76	0.84	2.71

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	148,736	0.92	3.18	0.93	3.03
Gender	Female	73,408	0.92	3.19	0.93	3.03
	Male	75,290	0.92	3.17	0.93	3.02
	Non-Binary	38	0.89	3.28	0.90	3.09
Ethnicity	Asian	16,305	0.93	2.86	0.93	2.70
	African American	24,121	0.91	3.23	0.92	3.08
	Hispanic	38,948	0.90	3.23	0.91	3.09
	American Indian	1,125	0.92	3.21	0.93	3.05
	Multiracial	4,897	0.93	3.17	0.93	3.00
	Pacific Islander	350	0.92	3.18	0.93	3.01
	White	62,605	0.92	3.17	0.92	3.03
NRC	New York City	44,443	0.93	3.13	0.94	2.96
	Big 4 Cities	5,708	0.90	3.16	0.91	3.02
	Urban/Suburban	10,002	0.89	3.22	0.90	3.09

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Rural	8,899	0.89	3.22	0.90	3.09
	Average Needs	35,867	0.91	3.20	0.91	3.07
	Low Needs	18,370	0.91	3.04	0.92	2.91
	Charter	13,507	0.92	3.16	0.93	3.01
	Religious or Independent	11,940	0.92	3.22	0.93	3.08
SWD	All Codes	19,773	0.88	3.17	0.89	3.05
SUA	All Codes	20,348	0.89	3.18	0.90	3.05
ELL	ELL=Y	10,708	0.85	3.16	0.86	3.06
SWD/SUA	SWD & SUA codes	16,023	0.87	3.15	0.88	3.04
ELL/SUA	SUA & ELL codes	1,635	0.77	3.06	0.78	2.99

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	90,895	0.92	3.16	0.93	3.00
Gender	Female	43,245	0.92	3.18	0.93	3.02
	Male	47,612	0.92	3.13	0.93	2.98
	Non-Binary	38	0.93	3.21	0.94	3.07
Ethnicity	Asian	7,260	0.94	3.06	0.95	2.87
	African American	16,936	0.92	3.11	0.93	2.96
	Hispanic	26,285	0.92	3.14	0.92	2.99
	American Indian	697	0.91	3.16	0.92	3.00
	Multiracial	2,919	0.92	3.14	0.93	2.99
	Pacific Islander	169	0.93	3.16	0.94	2.98
	White	36,043	0.91	3.18	0.92	3.05
NRC	New York City	26,291	0.93	3.15	0.94	2.97
	Big 4 Cities	4,872	0.91	2.95	0.92	2.83
	Urban/Suburban	7,899	0.88	3.05	0.89	2.95
	Rural	7,020	0.90	3.13	0.90	3.02
	Average Needs	24,038	0.90	3.18	0.90	3.06
	Low Needs	8,237	0.91	3.17	0.92	3.03
	Charter	8,342	0.93	3.15	0.94	2.96
	Religious or Independent	4,196	0.93	3.15	0.94	2.98
SWD	All Codes	16,050	0.88	2.96	0.89	2.87
SUA	All Codes	16,711	0.89	2.99	0.90	2.89
ELL	ELL=Y	4,618	0.87	2.93	0.87	2.86
SWD/SUA	SWD & SUA codes	13,405	0.87	2.93	0.88	2.86
ELL/SUA	SUA & ELL codes	1,259	0.80	2.79	0.80	2.75

7.2. Standard Error of Measurement (SEM)

Table 7.2 and Table 7.4 presented the SEMs computed from Cronbach's alpha and the Feldt-Raju reliability statistics for ELA and mathematics, respectively. The SEMs ranged from 2.43 to 3.17 across subjects, grades, and the two estimation methods, which were 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 the SEMs were expected to be low.

The SEMs for the subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, were presented in Tables 7.9–7.14 and Tables 7.15–7.20 for ELA and mathematics, respectively. The SEMs associated with all reliability estimates for all subjects, grades, estimation methods, and subpopulations, except for the non-binary group, ranged from 2.33 to 3.23, which were close to those for the entire population. This narrow range indicates that all students' test scores are reasonably reliable across the Grades 3–8 ELA and Mathematics Tests with minimal error.

7.3. Performance Level Classification Consistency and Accuracy

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 & 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 the extent to which a student's performance classification is expected to remain the same over repeated measurements.

Classification consistency is most relevant for students whose performance is near the proficiency cut score. For example, consider the cut score delineating Levels 2 and 3, or simply the "Level 3 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. Students 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, as these numbers show the number of students who are at risk of being misclassified.

Scoring tables with SEMs and student scale score frequency distributions are located in Appendix P. Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee et al. (2002) and Wang et al. (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 those based solely on the Level 3 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 elements in the contingency table. Kappa is a similar measure but corrects for chance agreement. The inconsistency index is equal to the "1 - agreement index."

Table 7.21 depicts the ELA and mathematics consistency study results based on the range of performance levels for all grades. For ELA, 64–68% of students were estimated to be classified consistently into one of the four performance categories following a hypothetical second administration. Kappa coefficients—which correct for chance agreement—ranged from 0.51 to 0.56. These values are between "moderate" and "substantial" agreement per Landis and Koch's (1977) rules of thumb for kappa. For mathematics, 72–75% of students were estimated to be classified consistently into one of the four performance categories, and kappa coefficients ranged from 0.62 to 0.64. These values are all considered "substantial" agreement by Landis and Koch's (1977) rules of thumb for the kappa statistic.

As mentioned above, all scores contain an acceptable measurement error for all tests. For example, by random chance, students testing twice may be classified first as Level 3 and second as Level 4. This is expected to occur more often for students scoring around a specific cut score and less often for students scoring closer to the middle of a 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	164,051	68%	32%	0.56
4	168,586	65%	35%	0.53
5	158,764	64%	36%	0.51
6	164,803	65%	35%	0.53
7	159,817	65%	35%	0.53
8	147,257	68%	32%	0.56
Mathematics				
3	156,673	74%	26%	0.63
4	157,499	73%	27%	0.62
5	152,137	73%	27%	0.62
6	154,141	75%	25%	0.64
7	148,736	72%	28%	0.62
8	90,895	73%	27%	0.62

Table 7.22 depicts the ELA and mathematics consistency study results based on two performance levels (NYS Level 2 and NYS Level 3) as defined by the Level 3 cut. For ELA, 85–87% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.69 to 0.74. These values are considered "substantial" agreement per Landis and Koch's (1977) rules of thumb for kappa. For mathematics, 89–90% of the classifications were estimated consistently, and kappa coefficients ranged from 0.77 to 0.79. These values indicate at least "substantial" agreement (where kappa > 0.60) for all tests and "almost perfect" agreement (where kappa > 0.80) for a few tests per Landis and Koch's (1977) rules of thumb for kappa.

Table 7.22. Decision Consistency (Level 3 Cut)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	164,051	87%	13%	0.74
4	168,586	86%	14%	0.72
5	158,764	85%	15%	0.69
6	164,803	86%	14%	0.71
7	159,817	86%	14%	0.72
8	147,257	87%	13%	0.74
Mathematics				
3	156,673	90%	10%	0.79
4	157,499	89%	11%	0.77
5	152,137	89%	11%	0.78
6	154,141	90%	10%	0.79
7	148,736	89%	11%	0.78
8	90,895	89%	11%	0.78

7.3.2. Accuracy

Table 7.23 presents the classification accuracy results 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 3 cut score. By definition, accuracy associated with the Level 3 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 agrees with the location of their underlying proficiency 72% to 75% of the time across all performance levels and 89% to 91% of the time regarding the Level 3 cut score. For mathematics, the estimated accuracy rates indicate that the categorization of a student's observed performance agrees with the location of their true proficiency 79% to 81% of the time across all performance levels and 92% to 93% of the time regarding the Level 3 cut score.

Table 7.23. Decision Agreement (Accuracy) Estimates

Grade	N-Count	Accuracy	
		All Cuts	Level 3 Cut
ELA			
3	164,051	73%	89%
4	168,586	72%	89%
5	158,764	73%	89%
6	164,803	74%	90%
7	159,817	74%	89%
8	147,257	75%	91%
Mathematics			
3	156,673	81%	92%
4	157,499	81%	92%
5	152,137	81%	92%
6	154,141	81%	93%
7	148,736	79%	92%
8	90,895	79%	92%

Section 8: Summary of Operational Test Results

This section summarizes the distribution of scale-score results on the NYSTP 2024 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 subgroups. Demographic subgroups include gender, ethnicity, Needs Resource Capacity (NRC) category, English Language Learner (ELL) status, students with disabilities (SWDs), and students using test accommodations (SUAs). Furthermore, the ELL/SUA subgroup is defined as ELLs who use one or more ELL-related accommodation, and the SWD/SUA subgroup is defined as SWDs who use one or more disability-related accommodation. For the mathematics analyses, the test translation language is also indicated (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

The following two subsections present ELA and mathematics scale scores and subscore statistics by grade and selected subgroups. (Caution is advised when interpreting the statistics for subgroups with small counts.)

8.1.1. ELA Scale Score and Subscore Distributions

Table 8.1 and Table 8.2 show the summary of scale scores and raw subscores, respectively, for each ELA grade. Tables 8.3–8.8 show the summary of scale scores by subgroup for ELA. Some general observations from these tables include:

- Female students outperformed Male students.
- Asian students scored considerably higher than other reported ethnic groups.
- Students from Low Needs (as identified by NRC category) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter).
- ELLs, SWDs, and/or SUAs tended to underperform the State population (All Students).

Table 8.1. ELA Scale Score Distribution Summary

Grade	N-Count	Scale Score	
		Mean	SD
3	164,461	443.27	22.17
4	168,811	444.42	22.95
5	158,765	444.11	22.99
6	164,913	443.16	22.65
7	159,902	447.62	22.82
8	147,257	449.25	24.63

Table 8.2. ELA Subscore Summary

Grade	Subscore Category	N-Count	Subscore		
			Max	Mean	SD
3	Reading	164,461	23	13.12	5.37
	Writing	164,461	10	5.25	2.76
4	Reading	168,811	23	14.67	4.93
	Writing	168,811	14	7.92	3.61
5	Reading	158,765	26	14.40	5.08
	Writing	158,765	14	8.05	3.52
6	Reading	164,913	26	15.09	5.25
	Writing	164,913	14	8.96	3.43
7	Reading	159,902	33	20.33	6.26
	Writing	159,902	14	9.62	3.44
8	Reading	147,257	33	21.98	7.19
	Writing	147,257	14	9.21	3.43

8.1.1.1. ELA Grade 3

Table 8.3 presents the Grade 3 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 443.27, with a standard deviation of 22.17. Female students tended to outperform Male students by around 3 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 (10 scale-score points above the State population), and Hispanic students earned the lowest mean score (5 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (9 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (12 scale-score points below the State population). The students with disabilities (SWD), students tested under accommodations (SUA), and ELL subgroups scored about 14–20 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring 22 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	164,461	443.27	22.17
Gender	Female	81,973	445.00	21.97
	Male	82,468	441.55	22.24
	Non-Binary	20	458.00	25.76
Ethnicity	Asian	17,179	453.47	22.35
	African American	22,990	439.84	21.44
	Hispanic	45,287	437.98	21.13
	American Indian	1,265	442.15	21.29

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Multiracial	6,165	446.39	23.38
	Pacific Islander	317	446.11	21.56
	White	70,988	445.09	21.76
NRC	New York City	46,749	445.75	23.39
	Big 4 Cities	5,879	431.55	21.59
	Urban/Suburban	12,341	434.03	19.63
	Rural	9,142	436.86	19.95
	Average Needs	44,346	442.54	20.60
	Low Needs	20,818	452.28	20.21
	Charter	14,484	446.49	21.29
	Religious or Independent	10,702	436.23	22.36
SWD	All Codes	24,063	429.61	19.31
SUA	All Codes	21,818	428.92	18.76
ELL	ELL=Y	12,858	423.34	16.56
ELL/SUA	SUA & ELL codes	1,378	421.36	15.42
SWD/SUA	SWD & SUA codes	18,344	427.36	18.09

8.1.1.2. ELA Grade 4

Table 8.4 presents Grade 4 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 444.42, with a standard deviation of 22.95. Female students tended to outperform Male students by around 4 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 (11 scale-score points above the State population), and Hispanic students earned the lowest mean score (5 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (10 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (14 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored about 16–24 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 25 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	SD
State	All Students	168,811	444.42	22.95
Gender	Female	84,493	446.61	22.75
	Male	84,309	442.23	22.95
	Non-Binary	9	–	–
Ethnicity	Asian	17,957	455.78	21.76
	African American	24,167	441.03	22.52

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Hispanic	46,542	439.66	22.27
	American Indian	1,181	442.85	22.00
	Multiracial	6,177	447.31	23.51
	Pacific Islander	328	446.64	21.72
	White	72,142	445.65	22.61
NRC	New York City	47,857	447.44	23.85
	Big 4 Cities	6,007	430.34	23.79
	Urban/Suburban	12,817	435.16	21.04
	Rural	9,418	437.36	20.44
	Average Needs	44,158	443.44	20.94
	Low Needs	20,794	454.00	19.76
	Charter	14,384	449.53	21.06
	Religious or Independent	13,376	436.65	25.22
SWD	All Codes	24,805	428.82	20.43
SUA	All Codes	24,509	428.74	20.24
ELL	ELL=Y	12,761	420.06	17.94
ELL/SUA	SUA & ELL codes	1,621	419.87	15.66
SWD/SUA	SWD & SUA codes	20,344	426.77	19.48

8.1.1.3. ELA Grade 5

Table 8.5 presents the Grade 5 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 444.11, with a standard deviation of 22.99. Female students tended to outperform Male students by around 4 scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those enrolled in New York City, Low Needs districts, Charter, and Religious or Independent schools. Across all ethnic groups, Asian students earned the highest mean score (11 scale-score points above the State population), and Hispanic students earned the lowest mean score (6 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (10 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (15 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored about 17–24 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 27 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	158,765	444.11	22.99
Gender	Female	78,487	446.05	22.73
	Male	80,260	442.20	23.09
	Non-Binary	18	454.78	21.50

Demographic Category		N-Count	Scale Score	
			Mean	SD
Ethnicity	Asian	17,273	455.31	22.36
	African American	24,605	438.63	22.17
	Hispanic	45,947	437.76	22.27
	American Indian	1,225	442.07	22.45
	Multiracial	5,807	446.77	23.85
	Pacific Islander	365	445.63	23.84
	White	62,968	447.71	21.90
NRC	New York City	49,162	444.93	23.94
	Big 4 Cities	6,170	429.16	23.21
	Urban/Suburban	12,328	434.59	21.31
	Rural	9,137	439.02	21.13
	Average Needs	42,790	443.47	21.75
	Low Needs	19,814	453.89	20.86
	Charter	14,452	446.17	21.79
	Religious or Independent	4,912	448.10	21.14
SWD	All Codes	24,452	426.92	20.41
SUA	All Codes	25,159	427.53	20.68
ELL	ELL=Y	13,869	420.31	16.56
ELL/SUA	SUA & ELL codes	3,026	417.60	15.97
SWD/SUA	SWD & SUA codes	20,615	425.18	19.60

8.1.1.4. ELA Grade 6

Table 8.6 presents Grade 6 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 443.16, with a standard deviation of 22.65. Female students tended to outperform Male students by around 5 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 (13 scale-score points above the State population), and Hispanic students earned the lowest mean score (5 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (9 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (12 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored about 15–26 scale-score points below the mean scale score for the population. ELLs were the lowest-performing subgroup analyzed, scoring about 26 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	164,913	443.16	22.65
Gender	Female	81,294	445.91	22.17

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Male	83,581	440.47	22.79
	Non-Binary	38	450.32	20.15
Ethnicity	Asian	18,440	456.02	22.06
	African American	25,365	439.24	21.19
	Hispanic	46,506	438.02	21.57
	American Indian	1,225	441.43	21.18
	Multiracial	5,518	445.64	23.15
	Pacific Islander	354	445.34	22.09
	White	67,135	444.56	22.44
NRC	New York City	48,846	445.36	23.67
	Big 4 Cities	6,027	431.45	21.76
	Urban/Suburban	12,418	433.52	20.53
	Rural	9,249	437.78	20.21
	Average Needs	41,406	442.38	20.64
	Low Needs	19,382	452.61	20.02
	Charter	15,166	446.87	21.27
	Religious or Independent	12,419	437.12	25.66
SWD	All Codes	24,177	426.43	19.15
SUA	All Codes	24,420	427.68	19.73
ELL	ELL=Y	10,627	416.70	16.42
ELL/SUA	SUA & ELL codes	1,588	417.00	14.21
SWD/SUA	SWD & SUA codes	19,760	425.06	18.49

8.1.1.5. ELA Grade 7

Table 8.7 presents the Grade 7 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 447.62, with a standard deviation of 22.82. Female students tended to outperform Male students by 6 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 (14 scale-score points above the State population), and Hispanic students earned the lowest mean score (5 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (8 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (12 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored about 15–29 scale-score points below the mean scale score for the population. ELLs were the lowest-performing subgroup analyzed, scoring about 29 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	159,902	447.62	22.82
Gender	Female	78,716	450.41	22.01
	Male	81,140	444.90	23.25
	Non-Binary	46	456.26	20.42
Ethnicity	Asian	17,193	461.52	21.43
	African American	25,491	444.67	21.48
	Hispanic	45,845	443.05	22.01
	American Indian	1,165	444.79	22.47
	Multiracial	5,199	449.28	23.35
	Pacific Islander	360	452.39	22.61
	White	64,263	448.31	22.62
NRC	New York City	49,335	451.49	23.35
	Big 4 Cities	5,878	435.44	22.64
	Urban/Suburban	11,904	436.47	20.74
	Rural	9,165	440.42	20.25
	Average Needs	38,948	445.16	21.06
	Low Needs	19,185	455.96	19.97
	Charter	14,443	452.40	20.23
	Religious or Independent	11,044	442.70	26.76
SWD	All Codes	22,802	431.29	19.78
SUA	All Codes	23,595	432.74	20.53
ELL	ELL=Y	8,904	418.28	17.75
ELL/SUA	SUA & ELL codes	1,449	419.68	14.44
SWD/SUA	SWD & SUA codes	18,758	430.03	19.33

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 449.25, with a standard deviation of 24.63. Female students tended to outperform Male students by 6 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, Charter, and Religious or Independent schools. Across ethnic groups, Asian students earned the highest mean score (14 scale-score points above the State population), and Hispanic students earned the lowest mean score (6 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (10 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (13 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored about 16–31 scale-score points below the mean scale score for the population. ELLs were the lowest-performing subgroup analyzed, scoring about 31 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	147,257	449.25	24.63
Gender	Female	70,999	452.42	24.29
	Male	76,200	446.28	24.57
	Non-Binary	58	461.47	20.91
Ethnicity	Asian	16,461	462.97	23.76
	African American	24,382	444.23	23.31
	Hispanic	44,656	443.33	24.08
	American Indian	1,084	446.37	23.60
	Multiracial	4,540	450.82	25.77
	Pacific Islander	299	451.06	24.24
	White	55,213	452.22	23.62
NRC	New York City	49,452	450.73	25.12
	Big 4 Cities	6,104	435.79	23.94
	Urban/Suburban	11,837	438.30	23.55
	Rural	8,677	442.60	22.77
	Average Needs	36,154	447.42	23.59
	Low Needs	17,401	459.44	22.83
	Charter	12,770	453.36	22.45
	Religious or Independent	4,862	455.79	22.91
SWD	All Codes	21,371	430.82	20.27
SUA	All Codes	22,544	432.77	21.72
ELL	ELL=Y	10,876	418.42	15.03
ELL/SUA	SUA & ELL codes	2,242	419.58	14.06
SWD/SUA	SWD & SUA codes	17,809	429.49	19.72

8.1.2. Mathematics Scale Score and Subscore Distributions

Table 8.9 and Table 8.10 show the summary of scale scores and raw subscores, respectively, for mathematics. Tables 8.11–8.16 summarize the scale scores by subgroup for mathematics. Some general observations from the mathematics data include:

- Male students outperformed Female students in Grades 3–5 but underperformed Female students in Grade 8, while performance among the groups was roughly equal in Grades 6 and 7.
- Asian students scored considerably higher than other reported ethnic groups.
- Students from Low Needs districts (as identified by NRC category) outperformed students from High Needs (New York City, Big 4 Cities, Urban/Suburban, and Rural) and Average Needs districts.
- ELLs, SWDs, and/or SUAs tended to underperform the State population (All Students).
- Students taking the Chinese (Simplified) translation tended to outperform those taking other translations.

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. Therefore, the summary of mean scale scores for the translation subgroups was limited to subgroups with minimum n-counts of 50.

Table 8.9. Mathematics Scale Score Distribution Summary

Grade	N-Count	Scale Score	
		Mean	SD
3	167,123	451.21	26.34
4	167,995	456.15	29.27
5	155,787	451.58	27.52
6	164,601	451.37	26.87
7	157,751	457.06	29.01
8	94,063	447.18	27.79

Table 8.10. Mathematics Subscore Summary

Grade	Subscore Category	N-Count	Subscore		
			Max	Mean	SD
3	Operations and Algebraic Thinking	167,123	13	8.30	3.48
	Number and Operations—Fractions	167,123	10	4.72	2.49
	Measurement and Data	167,123	9	5.62	2.45
4	Operations and Algebraic Thinking	167,995	9	4.62	2.52
	Number and Operations in Base Ten	167,995	11	6.73	3.08
	Number and Operations—Fractions	167,995	11	7.08	2.88
5	Number and Operations in Base Ten	155,787	14	7.06	3.13
	Number and Operations—Fractions	155,787	17	7.71	4.05
	Measurement and Data	155,787	11	5.90	3.13
6	Ratios and Proportional Relationships	164,601	10	5.81	2.79
	The Number System	164,601	11	4.89	3.00
	Expressions, Equations, and Inequalities	164,601	17	9.03	4.42
7	Ratios and Proportional Relationships	157,751	15	7.51	3.83
	The Number System	157,751	11	7.22	2.58
	Expressions, Equations, and Inequalities	157,751	14	6.98	3.69
8	Expressions, Equations, and Inequalities	94,063	16	7.32	4.35
	Functions	94,063	10	5.26	2.60
	Geometry	94,063	16	6.92	3.92

8.1.2.1. Mathematics Grade 3

Table 8.11 presents the Grade 3 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 451.21, with a standard deviation of 26.34. Male students outperformed Female students by 2 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 (15 scale-score points above the State population), and Hispanic students earned the lowest mean score (6 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (11 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (16 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored 16–21 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 25 scale-score points below the State mean. Among ELLs taking the language translation forms, the mean scale score was the highest for the Chinese (Simplified) subgroup and the lowest for the Haitian-Creole subgroup.

Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	167,123	451.21	26.34
Gender	Female	83,036	450.08	25.53
	Male	84,068	452.33	27.07
	Non-Binary	19	462.58	33.26
Ethnicity	Asian	17,822	466.65	25.45
	African American	23,037	446.75	26.95
	Hispanic	46,466	445.31	25.43
	American Indian	1,267	450.02	26.95
	Multiracial	6,119	453.31	27.55
	Pacific Islander	330	453.93	26.60
	White	71,734	452.55	25.04
NRC	New York City	49,288	453.68	27.61
	Big 4 Cities	5,950	435.09	25.54
	Urban/Suburban	12,158	440.69	24.14
	Rural	9,153	442.44	23.94
	Average Needs	44,286	449.70	24.19
	Low Needs	20,886	462.01	23.05
	Charter	14,249	460.36	26.35
	Religious or Independent	11,153	441.61	24.33
SWD	All Codes	24,124	434.76	25.81
SUA	All Codes	14,419	430.57	23.86
ELL	ELL=Y	24,436	435.52	23.54
ELL/SUA	SUA & ELL codes	1,341	425.92	21.71
SWD/SUA	SWD & SUA codes	11,344	427.63	22.93

Demographic Category		N-Count	Scale Score	
			Mean	SD
ELL Test Language	Arabic	174	432.21	24.81
	Bengali	80	432.41	24.83
	Chinese (Traditional)	57	450.96	25.78
	Chinese (Simplified)	447	460.25	22.43
	Haitian-Creole	91	429.93	23.37
	Korean	48	469.19	24.60
	Russian	309	446.11	23.79
	Spanish	4,994	431.60	21.75
	All Translations	6,200	434.86	23.71

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 456.15, with a standard deviation of 29.27. Male students outperformed Female students by 3 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 (19 scale-score points above the State population), and Hispanic and African-American students earned the lowest mean score (7 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (14 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (21 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored 20–22 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 30 scale-score points below the State mean. Among ELLs taking the language translation forms, the mean scale score was the highest for the Chinese (Simplified) subgroup and the lowest for the Haitian-Creole subgroup.

Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	167,995	456.15	29.27
Gender	Female	83,828	454.77	28.69
	Male	84,158	457.52	29.77
	Non-Binary	9	–	–
Ethnicity	Asian	18,435	475.16	27.51
	African American	23,689	449.27	29.36
	Hispanic	46,496	448.66	28.20
	American Indian	1,146	453.56	29.03
	Multiracial	5,995	458.85	30.08
	Pacific Islander	329	461.12	29.77
	White	71,541	458.30	27.64

Demographic Category		N-Count	Scale Score	
			Mean	SD
NRC	New York City	50,081	458.50	31.01
	Big 4 Cities	5,477	434.69	28.10
	Urban/Suburban	12,168	443.79	26.87
	Rural	9,037	447.80	25.87
	Average Needs	43,169	455.42	26.35
	Low Needs	20,529	469.93	25.01
	Charter	14,003	464.07	29.20
	Religious or Independent	13,531	446.06	27.70
SWD	All Codes	24,185	436.44	27.15
SUA	All Codes	16,329	434.18	25.45
ELL	ELL=Y	22,703	434.91	24.99
ELL/SUA	SUA & ELL codes	1,546	425.84	21.26
SWD/SUA	SWD & SUA codes	12,662	430.47	24.17
ELL Test Language	Arabic	140	431.85	26.16
	Bengali	72	433.79	25.20
	Chinese (Traditional)	52	450.21	33.71
	Chinese (Simplified)	451	471.24	25.06
	Haitian-Creole	110	430.85	23.44
	Korean	34	479.97	25.92
	Russian	401	448.61	26.98
	Spanish	5,107	431.53	23.72
	All Translations	6,367	435.85	26.69

8.1.2.3. Mathematics Grade 5

Table 8.13 presents the Grade 5 n-counts and scale-score statistics for key demographic subgroups. The population scale-score mean was 451.58, with a standard deviation of 27.52. Male students outperformed Female students by 4 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, Charter, and Religious or Independent schools. Across ethnic groups, Asian students earned the highest mean score (19 scale-score points above the State population), and African-American students earned the lowest mean score (9 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (12 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (19 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored 18–20 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 25 scale-score points below the State mean. Among ELLs taking the language translation forms, the mean scale score was the highest for the Chinese (Simplified) subgroup and the lowest for the Haitian-Creole subgroup.

Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	155,787	451.58	27.52
Gender	Female	76,559	449.56	26.35
	Male	79,210	453.53	28.46
	Non-Binary	18	452.44	30.19
Ethnicity	Asian	17,249	470.67	28.27
	African American	24,332	442.57	25.87
	Hispanic	43,587	444.55	25.20
	American Indian	1,218	448.01	27.24
	Multiracial	5,718	453.00	28.69
	Pacific Islander	361	453.38	27.52
	White	62,584	454.88	26.05
NRC	New York City	48,070	454.19	29.07
	Big 4 Cities	6,066	432.94	23.86
	Urban/Suburban	11,785	438.83	23.63
	Rural	9,017	443.20	22.98
	Average Needs	42,096	450.13	25.22
	Low Needs	19,764	463.46	25.77
	Charter	14,368	454.27	28.41
	Religious or Independent	4,621	451.79	25.87
SWD	All Codes	24,004	433.07	23.36
SUA	All Codes	25,002	433.15	23.01
ELL	ELL=Y	12,161	431.90	20.33
ELL/SUA	SUA & ELL codes	2,969	426.55	18.68
SWD/SUA	SWD & SUA codes	20,531	430.86	21.93
ELL Test Language	Arabic	111	432.64	20.00
	Bengali	30	432.33	17.14
	Chinese (Traditional)	25	474.28	25.10
	Chinese (Simplified)	179	465.18	24.20
	Haitian-Creole	51	423.98	17.98
	Korean	28	465.07	23.96
	Russian	198	446.75	25.86
	Spanish	3,026	431.60	19.83
	All Translations	3,648	434.55	22.34

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 451.37, with a standard deviation of 26.87. 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, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (20 scale-score points above the State population), and Hispanic and African-American students earned the lowest mean score (8 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (15 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (15 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored 18–21 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 26 scale-score points below the State mean. Among ELLs taking the language translation forms, the mean scale score was the highest for the Chinese (Simplified) subgroup and the lowest for the Haitian-Creole subgroup.

Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	164,601	451.37	26.87
Gender	Female	80,939	451.18	26.57
	Male	83,630	451.56	27.15
	Non-Binary	32	456.97	24.32
Ethnicity	Asian	18,979	471.71	27.91
	African American	24,841	443.69	24.96
	Hispanic	46,248	443.69	24.18
	American Indian	1,224	447.28	24.41
	Multiracial	5,415	454.06	28.25
	Pacific Islander	359	453.33	26.37
	White	67,092	453.74	25.42
NRC	New York City	50,648	452.86	28.91
	Big 4 Cities	6,025	436.13	23.62
	Urban/Suburban	11,905	438.77	22.50
	Rural	9,009	443.40	22.42
	Average Needs	40,020	451.09	23.96
	Low Needs	19,193	465.96	24.61
	Charter	14,580	455.65	27.34
	Religious or Independent	13,221	444.33	24.75
SWD	All Codes	23,236	432.28	20.99
SUA	All Codes	17,040	433.36	21.96
ELL	ELL=Y	19,065	430.75	18.76
ELL/SUA	SUA & ELL codes	1,561	424.91	16.11
SWD/SUA	SWD & SUA codes	12,883	429.24	19.29

Demographic Category		N-Count	Scale Score	
			Mean	SD
ELL Test Language	Arabic	285	431.67	19.48
	Bengali	189	439.79	22.60
	Chinese (Traditional)	59	449.51	31.04
	Chinese (Simplified)	753	465.97	23.88
	Haitian-Creole	151	426.71	17.07
	Korean	28	467.61	25.60
	Russian	414	446.48	22.93
	Spanish	5,562	429.25	17.00
	All Translations	7,441	434.54	21.98

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 457.06, with a standard deviation of 29.01. 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, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (23 scale-score points above the State population), and Hispanic and African-American students earned the lowest mean score (8 scale-score points below the State population). Across NRC subgroups, students from Low Needs districts earned the highest mean scale score (14 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (19 scale-score points below the State population). The SWD, SUA, and ELL subgroups scored 19–24 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 30 scale-score points below the State mean. Among ELLs taking the language translation forms, the mean scale score was the highest for the Chinese (Simplified) subgroup and the lowest for the Haitian-Creole subgroup.

Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	157,751	457.06	29.01
Gender	Female	77,504	457.02	28.38
	Male	80,209	457.09	29.60
	Non-Binary	38	462.24	24.54
Ethnicity	Asian	17,345	479.62	29.33
	African American	24,734	448.79	26.91
	Hispanic	45,161	449.47	25.90
	American Indian	1,155	451.73	28.13
	Multiracial	4,962	458.85	29.88
	Pacific Islander	359	460.48	29.06
	White	63,617	459.60	28.00
	New York City	50,307	459.99	30.60

Demographic Category		N-Count	Scale Score	
			Mean	SD
NRC	Big 4 Cities	5,821	437.87	24.60
	Urban/Suburban	11,203	443.10	23.51
	Rural	8,939	448.32	24.02
	Average Needs	36,985	456.21	26.05
	Low Needs	18,556	471.20	26.65
	Charter	13,896	462.46	29.05
	Religious or Independent	12,044	448.16	29.09
SWD	All Codes	21,622	436.56	23.05
SUA	All Codes	16,385	437.79	24.35
ELL	ELL=Y	16,506	432.98	21.19
ELL/SUA	SUA & ELL codes	1,310	426.69	15.14
SWD/SUA	SWD & SUA codes	12,199	432.90	21.27
ELL Test Language	Arabic	226	436.40	22.22
	Bengali	164	448.16	26.29
	Chinese (Traditional)	46	453.65	33.04
	Chinese (Simplified)	591	474.99	28.37
	Haitian-Creole	142	431.47	17.21
	Korean	28	469.14	33.45
	Russian	459	455.12	25.24
	Spanish	5,441	432.68	18.70
	All Translations	7,097	438.39	24.25

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 447.18, with a standard deviation of 27.79. Female students tended to outperform Male students by 2 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, Low Needs districts, Charter, and Religious or Independent schools. Across ethnic groups, Asian students earned the highest mean score (20 scale-score points above the State population), and African-American students earned the lowest mean score (7 scale-score points below the State population). Across NRC subgroups, students from Religious or Independent schools earned the highest mean scale score (12 scale-score points above the State population), and students from Big 4 Cities districts earned the lowest mean score (17 scale-score points below the State population). 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 26 scale-score points below the State mean. Among ELLs taking the language translation forms, the mean scale score was the highest for the Chinese (Simplified) subgroup and the lowest for the Haitian-Creole subgroup.

Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	94,063	447.18	27.79
Gender	Female	44,681	448.25	27.26
	Male	49,344	446.20	28.22
	Non-Binary	38	455.39	30.19
Ethnicity	Asian	7,598	466.76	31.23
	African American	17,063	440.54	26.77
	Hispanic	28,585	442.44	26.38
	American Indian	715	442.31	26.35
	Multiracial	2,921	444.64	27.58
	Pacific Islander	174	448.57	29.16
	White	36,377	450.48	26.17
NRC	New York City	28,004	450.32	29.69
	Big 4 Cities	4,918	429.88	24.67
	Urban/Suburban	8,593	433.38	21.53
	Rural	7,058	440.91	23.41
	Average Needs	24,503	445.30	23.86
	Low Needs	8,368	458.29	26.57
	Charter	8,421	454.79	30.15
	Religious or Independent	4,198	458.78	29.85
SWD	All Codes	16,627	429.41	21.86
SUA	All Codes	17,286	430.71	22.58
ELL	ELL=Y	6,942	428.84	21.12
ELL/SUA	SUA & ELL codes	1,729	421.38	16.44
SWD/SUA	SWD & SUA codes	13,921	427.94	21.02
ELL Test Language	Arabic	161	432.93	23.26
	Bengali	73	442.42	23.73
	Chinese (Traditional)	14	462.21	28.62
	Chinese (Simplified)	175	470.86	29.24
	Haitian-Creole	92	425.13	16.63
	Korean	11	464.09	28.22
	Russian	249	454.34	27.52
	Spanish	2,385	429.16	19.66
	All Translations	3,160	434.10	24.26

8.2. Performance Level Distribution Summary

Students under the New York State Testing Program are classified into performance levels as Level 1, Level 2, Level 3, or Level 4. The cut scores for these performance levels were established during the standards review in the summer of 2023. While vertical articulation helps apply 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. Students are considered proficient if they are classified as Level 3 or Level 4.

8.2.1. ELA Test Performance Level Distributions

Table 8.17 shows the performance level distributions for all examinees for each ELA grade. Tables 8.18 through 8.23 show the performance level distributions by subgroup for each ELA grade. The percentage of proficient students at a subgroup level reflected the mean scale-score distributions for the subgroup. Therefore, similar achievement trends were observed for the percentage of proficient students:

- Female students outperformed Male students.
- Asian students outperformed other ethnic groups.
- Students from Low Needs (as identified by NRC category) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter).
- ELLs, SWDs, and/or SUAs tended to underperform the State population (All Students).

Table 8.17. ELA Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level 1	Level 2	Level 3	Level 4	Level 3 & 4
3	164,461	29.88	27.82	28.99	13.31	42.30
4	168,811	27.19	26.13	28.88	17.80	46.68
5	158,765	28.14	27.39	31.14	13.33	44.47
6	164,913	28.89	27.20	28.46	15.45	43.91
7	159,902	23.68	26.18	31.92	18.21	50.13
8	147,257	21.66	25.15	28.98	24.22	53.20

8.2.1.1. ELA Grade 3

Table 8.18 presents the ELA Grade 3 performance level distributions and n-counts for key demographic subgroups. The percentage of proficient students was 42.30% for the State population. That percentage was 6% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (19% above the State population) and the lowest for Hispanic students (10% below the state population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (18% above the State population) and the lowest for Big 4 Cities districts (19% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 25–34%

below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 36% below that for the State population.

Table 8.18. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	164,461	29.88	27.82	28.99	13.31	42.30
Gender	Female	81,973	26.97	27.92	30.27	14.84	45.11
	Male	82,468	32.78	27.72	27.72	11.78	39.50
	Non-Binary	20	20.00	15.00	35.00	30.00	65.00
Ethnicity	Asian	17,179	16.33	22.85	34.47	26.35	60.82
	African American	22,990	35.48	28.52	26.21	9.79	36.00
	Hispanic	45,287	38.38	29.38	24.00	8.24	32.24
	American Indian	1,265	31.07	29.96	27.43	11.54	38.97
	Multiracial	6,165	27.10	24.66	30.09	18.15	48.24
	Pacific Islander	317	23.97	25.87	35.96	14.20	50.16
	White	70,988	26.08	28.05	31.70	14.17	45.87
NRC	New York City	46,749	28.02	25.76	28.23	17.98	46.21
	Big 4 Cities	5,879	53.16	23.32	17.64	5.89	23.53
	Urban/Suburban	12,341	45.34	30.09	19.57	4.99	24.56
	Rural	9,142	39.51	30.41	24.05	6.03	30.08
	Average Needs	44,346	29.05	30.63	29.90	10.43	40.33
	Low Needs	20,818	14.25	25.82	39.13	20.80	59.93
	Charter	14,484	23.62	27.71	33.64	15.03	48.67
	Religious or Independent	10,702	41.52	26.84	23.83	7.82	31.65
SWD	All Codes	24,063	55.72	26.61	14.03	3.64	17.67
SUA	All Codes	21,818	56.79	26.99	13.18	3.04	16.22
ELL	ELL=Y	12,858	68.25	23.56	7.28	0.92	8.20
ELL/SUA	SUA & ELL codes	1,378	73.51	20.17	5.88	0.44	6.32
SWD/SUA	SWD & SUA codes	18,344	60.14	25.99	11.58	2.29	13.87

8.2.1.2. ELA Grade 4

Table 8.19 presents the ELA Grade 4 performance level distributions and n-counts for key demographic subgroups. The percentage of proficient students was 46.68% for the State population. That percentage was 8% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (21% above the State population) and the lowest for Hispanic students (9% below the state population). Across NRC subgroups, the

percentage of proficient students was the highest for Low Needs districts (18% above the State population) and the lowest for Big 4 Cities districts (22% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 28–40% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 42% below that for the State population.

Table 8.19. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	168,811	27.19	26.13	28.88	17.80	46.68
Gender	Female	84,493	24.23	25.30	30.04	20.42	50.46
	Male	84,309	30.15	26.96	27.72	15.16	42.88
	Non-Binary	9	–	–	–	–	–
Ethnicity	Asian	17,957	13.20	19.39	33.15	34.27	67.42
	African American	24,167	32.44	27.31	26.83	13.43	40.26
	Hispanic	46,542	34.24	28.18	25.52	12.06	37.58
	American Indian	1,181	28.37	29.64	26.59	15.41	42.00
	Multiracial	6,177	24.20	24.74	28.30	22.76	51.06
	Pacific Islander	328	24.39	24.70	31.10	19.82	50.92
	White	72,142	24.48	26.18	30.81	18.52	49.33
NRC	New York City	47,857	24.61	23.72	28.21	23.46	51.67
	Big 4 Cities	6,007	52.57	22.36	17.21	7.86	25.07
	Urban/Suburban	12,817	42.10	29.04	21.53	7.33	28.86
	Rural	9,418	36.94	31.25	23.69	8.12	31.81
	Average Needs	44,158	26.69	29.34	30.04	13.92	43.96
	Low Needs	20,794	12.07	23.05	37.22	27.66	64.88
	Charter	14,384	18.40	25.28	34.56	21.76	56.32
	Religious or Independent	13,376	38.46	25.12	24.39	12.03	36.42
SWD	All Codes	24,805	55.13	26.20	14.20	4.47	18.67
SUA	All Codes	24,509	55.17	26.56	13.99	4.28	18.27
ELL	ELL=Y	12,761	70.63	22.19	6.55	0.63	7.18
ELL/SUA	SUA & ELL codes	1,621	73.66	21.28	4.75	0.31	5.06
SWD/SUA	SWD & SUA codes	20,344	59.12	25.73	12.03	3.11	15.14

8.2.1.3. ELA Grade 5

Table 8.20 presents the ELA Grade 5 performance level distributions and n-counts for key demographic subgroups. The percentage of proficient students was 44.47% for the State population. That percentage was 6% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in

New York City, Low Needs districts, Charter schools, or Religious or Independent schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (21% above the State population) and the lowest for Hispanic students (12% below the state population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (18% above the State population) and the lowest for Big 4 Cities districts (22% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 28–39% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 41% below that for the State population.

Table 8.20. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	158,765	28.14	27.39	31.14	13.33	44.47
Gender	Female	78,487	25.05	27.28	32.66	15.01	47.67
	Male	80,260	31.16	27.50	29.65	11.69	41.34
	Non-Binary	18	11.11	22.22	38.89	27.78	66.67
Ethnicity	Asian	17,273	13.83	21.00	38.14	27.03	65.17
	African American	24,605	36.47	29.10	26.36	8.08	34.44
	Hispanic	45,947	37.68	29.34	25.40	7.57	32.97
	American Indian	1,225	30.94	29.55	28.00	11.51	39.51
	Multiracial	5,807	26.40	24.54	31.20	17.86	49.06
	Pacific Islander	365	26.58	25.75	30.14	17.53	47.67
	White	62,968	21.75	27.29	35.46	15.49	50.95
NRC	New York City	49,162	27.86	26.38	30.21	15.55	45.76
	Big 4 Cities	6,170	54.88	22.80	17.34	4.98	22.32
	Urban/Suburban	12,328	44.09	28.95	21.60	5.36	26.96
	Rural	9,137	35.19	30.77	26.81	7.23	34.04
	Average Needs	42,790	27.82	29.34	31.63	11.21	42.84
	Low Needs	19,814	13.36	23.87	40.49	22.27	62.76
	Charter	14,452	23.52	28.00	35.06	13.42	48.48
	Religious or Independent	4,912	20.22	28.48	36.20	15.11	51.31
SWD	All Codes	24,452	59.64	24.52	13.16	2.68	15.84
SUA	All Codes	25,159	58.42	24.90	13.61	3.08	16.69
ELL	ELL=Y	13,869	72.88	21.57	5.25	0.30	5.55
ELL/SUA	SUA & ELL codes	3,026	78.88	17.32	3.64	0.17	3.81
SWD/SUA	SWD & SUA codes	20,615	63.09	23.86	11.07	1.98	13.05

8.2.1.4. ELA Grade 6

Table 8.21 presents the ELA Grade 6 performance level distributions and n-counts for key demographic subgroups. The percentage of proficient students was 43.91% for the State population. That percentage was 9% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (24% above the State population) and the lowest for Hispanic students (10% below the state population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (18% above the State population) and the lowest for Big 4 Cities districts (20% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 27–40% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 41% below that for the State population.

Table 8.21. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	164,913	28.89	27.20	28.46	15.45	43.91
Gender	Female	81,294	24.51	27.06	30.34	18.09	48.43
	Male	83,581	33.16	27.34	26.61	12.88	39.49
	Non-Binary	38	18.42	18.42	47.37	15.79	63.16
Ethnicity	Asian	18,440	13.02	19.31	34.03	33.64	67.67
	African American	25,365	34.50	29.32	26.18	10.00	36.18
	Hispanic	46,506	36.76	29.13	24.50	9.61	34.11
	American Indian	1,225	29.80	29.22	29.96	11.02	40.98
	Multiracial	5,518	26.60	25.28	28.87	19.25	48.12
	Pacific Islander	354	24.01	29.10	27.40	19.49	46.89
	White	67,135	25.74	27.35	30.54	16.37	46.91
NRC	New York City	48,846	27.15	25.33	27.93	19.59	47.52
	Big 4 Cities	6,027	50.51	25.25	18.02	6.22	24.24
	Urban/Suburban	12,418	45.56	28.60	19.79	6.04	25.83
	Rural	9,249	36.43	30.59	25.16	7.83	32.99
	Average Needs	41,406	28.17	30.56	29.02	12.25	41.27
	Low Needs	19,382	13.27	24.95	37.24	24.54	61.78
	Charter	15,166	22.18	27.03	33.22	17.57	50.79
	Religious or Independent	12,419	37.99	24.10	25.30	12.61	37.91
SWD	All Codes	24,177	61.11	24.24	11.74	2.91	14.65
SUA	All Codes	24,420	58.57	24.80	12.91	3.72	16.63
ELL	ELL=Y	10,627	79.63	16.23	3.71	0.43	4.14
ELL/SUA	SUA & ELL codes	1,588	82.93	14.11	2.71	0.25	2.96

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
SWD/SUA	SWD & SUA codes	19,760	63.97	23.55	10.21	2.26	12.47

8.2.1.5. ELA Grade 7

Table 8.22 presents the ELA Grade 7 performance level distributions and n-counts for key demographic subgroups. The percentage of proficient students was 50.13% for the State population. That percentage was 10% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (25% above the State population) and the lowest for Hispanic students (9% below the state population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (16% above the State population) and the lowest for Urban/Suburban districts (22% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 28–45% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 47% below that for the State population.

Table 8.22. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	159,902	23.68	26.18	31.92	18.21	50.13
Gender	Female	78,716	19.34	25.66	34.15	20.85	55.00
	Male	81,140	27.90	26.69	29.75	15.65	45.40
	Non-Binary	46	10.87	26.09	39.13	23.91	63.04
Ethnicity	Asian	17,193	9.20	15.44	35.99	39.36	75.35
	African American	25,491	27.00	28.54	31.27	13.19	44.46
	Hispanic	45,845	29.53	29.04	29.08	12.35	41.43
	American Indian	1,165	27.64	27.64	30.13	14.59	44.72
	Multiracial	5,199	23.04	24.49	30.78	21.70	52.48
	Pacific Islander	360	18.06	22.78	34.17	25.00	59.17
	White	64,263	21.91	26.23	33.32	18.54	51.86
NRC	New York City	49,335	19.81	23.43	32.44	24.32	56.76
	Big 4 Cities	5,878	44.83	25.81	21.33	8.03	29.36
	Urban/Suburban	11,904	40.67	31.06	21.71	6.57	28.28
	Rural	9,165	33.45	31.31	26.95	8.28	35.23
	Average Needs	38,948	25.50	29.83	31.35	13.31	44.66
	Low Needs	19,185	10.85	23.01	39.57	26.57	66.14
	Charter	14,443	14.92	26.17	37.71	21.20	58.91
	Religious or Independent	11,044	30.68	21.82	31.51	15.99	47.50

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
SWD	All Codes	22,802	52.21	28.30	15.60	3.90	19.50
SUA	All Codes	23,595	49.57	28.24	17.10	5.10	22.20
ELL	ELL=Y	8,904	76.55	17.91	5.15	0.38	5.53
ELL/SUA	SUA & ELL codes	1,449	78.40	18.63	2.90	0.07	2.97
SWD/SUA	SWD & SUA codes	18,758	54.71	27.84	14.26	3.19	17.45

8.2.1.6. ELA Grade 8

Table 8.23 presents the ELA Grade 8 performance level distributions and n-counts for key demographic subgroups. The percentage of proficient students was 53.20% for the State population. That percentage was 10% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, Charter schools, or Religious or Independent schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (22% above the State population) and the lowest for Hispanic students (10% below the state population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (18% above the State population) and the lowest for Big 4 Cities districts (23% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 29–49% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 50% below that for the State population.

Table 8.23. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	147,257	21.66	25.15	28.98	24.22	53.20
Gender	Female	70,999	17.70	24.16	29.88	28.26	58.14
	Male	76,200	25.36	26.07	28.13	20.44	48.57
	Non-Binary	58	5.17	24.14	32.76	37.93	70.69
Ethnicity	Asian	16,461	9.69	14.85	28.92	46.55	75.47
	African American	24,382	26.78	28.68	27.92	16.62	44.54
	Hispanic	44,656	29.00	27.39	26.91	16.69	43.60
	American Indian	1,084	24.82	26.66	28.51	20.02	48.53
	Multiracial	4,540	20.99	24.27	26.83	27.91	54.74
	Pacific Islander	299	18.39	23.75	30.77	27.09	57.86
	White	55,213	16.81	24.89	31.42	26.88	58.30
NRC	New York City	49,452	20.79	23.86	28.40	26.96	55.36
	Big 4 Cities	6,104	41.83	27.72	19.59	10.86	30.45
	Urban/Suburban	11,837	37.10	28.40	22.51	12.00	34.51

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Rural	8,677	28.73	29.99	27.07	14.21	41.28
	Average Needs	36,154	22.05	27.84	29.81	20.30	50.11
	Low Needs	17,401	10.17	18.75	32.69	38.40	71.09
	Charter	12,770	14.49	24.38	33.77	27.35	61.12
	Religious or Independent	4,862	11.91	23.45	33.69	30.95	64.64
SWD	All Codes	21,371	49.21	30.34	15.74	4.70	20.44
SUA	All Codes	22,544	46.46	29.69	16.79	7.06	23.85
ELL	ELL=Y	10,876	75.77	20.30	3.59	0.34	3.93
ELL/SUA	SUA & ELL codes	2,242	74.58	22.30	2.94	0.18	3.12
SWD/SUA	SWD & SUA codes	17,809	51.82	30.01	14.27	3.90	18.17

8.2.2. Mathematics Test Performance Level Distributions

Table 8.24 shows the performance level distributions for all examinees for each mathematics grade. Tables 8.25 through 8.30 show the performance level distributions by subgroup for each mathematics grade. The percentage of proficient students at a subgroup level reflected the mean scale-score distributions for the subgroup. Therefore, similar achievement trends were observed for the percentage of proficient students:

- Gender performance was mixed. Male students outperformed Female students in Grades 3–5 but underperformed Female students in Grade 8, while performance among the groups was roughly equal in Grades 6 and 7.
- Asian students outperformed other ethnic groups.
- Students from Low Needs (as identified by NRC category) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter).
- ELLs, SWDs, and/or SUAs tended to underperform the State population (All Students).
- The subgroup that used the Chinese (Simplified) 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. Therefore, the summary of proficient percentages for the translation subgroups was limited to subgroups with a minimum n-count of 100.

Table 8.24. Mathematics Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level 1	Level 2	Level 3	Level 4	Level 3 & 4
3	167,123	13.78	32.61	39.50	14.10	53.60
4	167,995	19.54	22.53	38.27	19.65	57.92
5	155,787	26.11	23.18	33.52	17.19	50.71
6	164,601	24.02	25.14	36.29	14.54	50.83
7	157,751	17.50	24.96	31.04	26.51	57.55
8	94,063	35.80	20.70	29.51	13.99	43.50

8.2.2.1. Mathematics Grade 3

Table 8.25 presents the mathematics Grade 3 performance level summaries and n-counts for key demographic subgroups. The percentage of proficient students was 53.60% for the State population. That percentage was 4% lower for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (23% above the State population) and the lowest for Hispanic students (10% below the State population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (19% above the State population) and the lowest for Big 4 Cities districts (about 24% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 25–32% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 38% below that for the State population. Among ELLs taking the language translation forms, the percentage of proficient students was the lowest for the Spanish subgroup and the highest for the Chinese (Simplified) subgroup.

Table 8.25. Mathematics Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	167,123	13.78	32.61	39.50	14.10	53.60
Gender	Female	83,036	13.57	34.81	39.31	12.31	51.62
	Male	84,068	13.99	30.45	39.69	15.86	55.55
	Non-Binary	19	15.79	21.05	26.32	36.84	63.16
Ethnicity	Asian	17,822	5.18	18.58	44.28	31.96	76.24
	African American	23,037	18.75	35.00	34.32	11.93	46.25
	Hispanic	46,466	18.43	37.70	34.69	9.17	43.86
	American Indian	1,267	14.84	35.12	35.75	14.29	50.04
	Multiracial	6,119	13.60	30.38	38.47	17.55	56.02
	Pacific Islander	330	10.91	31.52	40.00	17.58	57.58
	White	71,734	11.20	32.17	43.34	13.29	56.63
NRC	New York City	49,288	13.57	29.72	38.50	18.20	56.70
	Big 4 Cities	5,950	33.43	37.28	24.32	4.97	29.29

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Urban/Suburban	12,158	22.46	41.03	30.92	5.59	36.51
	Rural	9,153	19.41	40.61	34.41	5.56	39.97
	Average Needs	44,286	12.27	35.97	41.65	10.11	51.76
	Low Needs	20,886	4.20	23.16	51.38	21.26	72.64
	Charter	14,249	7.85	25.41	42.27	24.46	66.73
	Religious or Independent	11,153	21.70	40.75	31.24	6.31	37.55
SWD	All Codes	24,124	33.99	37.90	22.44	5.68	28.12
SUA	All Codes	14,419	38.61	39.34	18.84	3.20	22.04
ELL	ELL=Y	24,436	29.79	42.72	23.23	4.26	27.49
ELL/SUA	SUA & ELL codes	1,341	46.31	38.33	13.57	1.79	15.36
SWD/SUA	SWD & SUA codes	11,344	43.48	38.48	15.67	2.37	18.04
ELL Test Language	Arabic	174	35.63	40.23	20.11	4.02	24.13
	Bengali	80	35.00	41.25	20.00	3.75	23.75
	Chinese (Traditional)	57	12.28	35.09	36.84	15.79	52.63
	Chinese (Simplified)	447	4.03	24.38	53.69	17.90	71.59
	Haitian-Creole	91	39.56	36.26	23.08	1.10	24.18
	Korean	48	6.25	8.33	52.08	33.33	85.41
	Russian	309	12.94	43.37	35.92	7.77	43.69
	Spanish	4,994	35.08	43.79	18.60	2.52	21.12
	All Translations	6,200	31.39	41.77	22.55	4.29	26.84

8.2.2.2. Mathematics Grade 4

Table 8.26 presents the mathematics Grade 4 performance level summaries and n-counts for key demographic subgroups. The percentage of proficient students was 57.92% for the State population. That percentage was 5% lower for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (23% above the State population) and the lowest for Hispanic students (11% below the State population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (21% above the State population) and the lowest for Big 4 Cities districts (about 30% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 29–32% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 44% below that for the State population. Among ELLs taking the language translation forms, the percentage of proficient students was the lowest for the Spanish subgroup and the highest for the Chinese (Simplified) subgroup.

Table 8.26. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	167,995	19.54	22.53	38.27	19.65	57.92
Gender	Female	83,828	20.13	24.26	37.73	17.87	55.60
	Male	84,158	18.96	20.81	38.81	21.43	60.24
	Non-Binary	9	–	–	–	–	–
Ethnicity	Asian	18,435	7.02	11.73	36.78	44.48	81.26
	African American	23,689	27.81	24.51	33.22	14.47	47.69
	Hispanic	46,496	26.43	26.46	34.71	12.40	47.11
	American Indian	1,146	22.86	23.04	38.05	16.06	54.11
	Multiracial	5,995	18.42	21.03	36.48	24.07	60.55
	Pacific Islander	329	13.98	21.28	39.82	24.92	64.74
	White	71,541	15.48	22.22	42.87	19.43	62.30
NRC	New York City	50,081	19.80	20.42	35.16	24.62	59.78
	Big 4 Cities	5,477	49.04	23.01	21.12	6.83	27.95
	Urban/Suburban	12,168	31.66	28.29	31.75	8.31	40.06
	Rural	9,037	24.69	29.19	37.00	9.12	46.12
	Average Needs	43,169	16.58	24.82	43.51	15.10	58.61
	Low Needs	20,529	5.92	15.37	46.69	32.01	78.70
	Charter	14,003	13.59	19.02	38.55	28.84	67.39
	Religious or Independent	13,531	28.63	27.74	33.65	9.98	43.63
SWD	All Codes	24,185	45.14	25.50	22.97	6.40	29.37
SUA	All Codes	16,329	47.10	26.72	21.97	4.21	26.18
ELL	ELL=Y	22,703	44.22	29.19	22.42	4.17	26.59
ELL/SUA	SUA & ELL codes	1,546	60.28	25.42	13.52	0.78	14.30
SWD/SUA	SWD & SUA codes	12,662	53.47	25.45	18.17	2.90	21.07
ELL Test Language	Arabic	140	53.57	22.86	20.00	3.57	23.57
	Bengali	72	48.61	26.39	19.44	5.56	25.00
	Chinese (Traditional)	52	30.77	19.23	28.85	21.15	50.00
	Chinese (Simplified)	451	6.21	12.64	46.12	35.03	81.15
	Haitian-Creole	110	53.64	21.82	20.91	3.64	24.55
	Korean	34	5.88	5.88	38.24	50.00	88.24
	Russian	401	27.43	24.19	36.66	11.72	48.38
	Spanish	5,107	49.76	28.00	19.62	2.62	22.24
	All Translations	6,367	45.01	26.24	22.77	5.97	28.74

8.2.2.3. Mathematics Grade 5

Table 8.27 presents the mathematics Grade 5 performance level summaries and n-counts for key demographic subgroups. The percentage of proficient students was 50.71% for the State population. That percentage was 6% lower for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, Charter schools, or Religious or Independent schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (26% above the State population) and the lowest for African-American students (14% below the State population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (20% above the State population) and the lowest for Big 4 Cities districts (about 28% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 28–32% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 39% below that for the State population. Among ELLs taking the language translation forms, the percentage of proficient students was the lowest for the Spanish subgroup and the highest for the Chinese (Simplified) subgroup.

Table 8.27. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	155,787	26.11	23.18	33.52	17.19	50.71
Gender	Female	76,559	27.21	24.96	33.50	14.32	47.82
	Male	79,210	25.04	21.45	33.54	19.96	53.50
	Non-Binary	18	22.22	33.33	22.22	22.22	44.44
Ethnicity	Asian	17,249	9.75	13.88	36.12	40.25	76.37
	African American	24,332	38.53	25.22	26.28	9.97	36.25
	Hispanic	43,587	34.02	26.04	29.76	10.19	39.95
	American Indian	1,218	30.95	23.23	31.53	14.29	45.82
	Multiracial	5,718	26.09	22.07	31.78	20.06	51.84
	Pacific Islander	361	22.71	24.10	34.07	19.11	53.18
	White	62,584	19.82	23.11	38.63	18.44	57.07
NRC	New York City	48,070	24.80	21.56	32.43	21.21	53.64
	Big 4 Cities	6,066	56.58	20.51	17.71	5.21	22.92
	Urban/Suburban	11,785	43.38	25.79	24.38	6.46	30.84
	Rural	9,017	34.09	27.99	30.49	7.43	37.92
	Average Needs	42,096	25.07	25.52	35.73	13.68	49.41
	Low Needs	19,764	10.75	18.80	43.03	27.42	70.45
	Charter	14,368	23.62	22.86	33.09	20.43	53.52
	Religious or Independent	4,621	23.07	25.84	35.43	15.67	51.10
SWD	All Codes	24,004	55.63	22.17	17.13	5.07	22.20
SUA	All Codes	25,002	55.11	22.52	17.60	4.77	22.37

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
ELL	ELL=Y	12,161	55.52	26.08	15.47	2.93	18.40
ELL/SUA	SUA & ELL codes	2,969	67.09	21.05	10.17	1.68	11.85
SWD/SUA	SWD & SUA codes	20,531	59.40	21.72	15.18	3.70	18.88
ELL Test Language	Arabic	111	53.15	27.03	18.02	1.80	19.82
	Bengali	30	53.33	30.00	16.67	.	16.67
	Chinese (Traditional)	25	4.00	16.00	24.00	56.00	80.00
	Chinese (Simplified)	179	8.38	17.88	44.69	29.05	73.74
	Haitian-Creole	51	70.59	23.53	3.92	1.96	5.88
	Korean	28	7.14	17.86	57.14	17.86	75.00
	Russian	198	32.83	23.23	31.31	12.63	43.94
	Spanish	3,026	55.58	25.88	16.29	2.25	18.54
	All Translations	3,648	51.43	25.25	18.75	4.58	23.33

8.2.2.4. Mathematics Grade 6

Table 8.28 presents the mathematics Grade 6 performance level summaries and n-counts for key demographic subgroups. The percentage of proficient students was 50.83% for the State population. That percentage was 1% lower for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (27% above the State population) and the lowest for African-American students (13% below the State population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (25% above the State population) and the lowest for Big 4 Cities districts (about 24% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 29–35% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 43% below that for the State population. Among ELLs taking the language translation forms, the percentage of proficient students was the lowest for the Haitian-Creole subgroup and the highest for the Chinese (Simplified) subgroup.

Table 8.28. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	164,601	24.02	25.14	36.29	14.54	50.83
Gender	Female	80,939	23.74	26.11	35.92	14.23	50.15
	Male	83,630	24.30	24.20	36.66	14.84	51.50
	Non-Binary	32	9.38	34.38	37.50	18.75	56.25
Ethnicity	Asian	18,979	8.58	13.33	38.92	39.17	78.09
	African American	24,841	33.90	27.83	29.75	8.52	38.27

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Hispanic	46,248	32.52	29.04	30.86	7.58	38.44
	American Indian	1,224	25.90	30.15	34.07	9.89	43.96
	Multiracial	5,415	23.25	21.98	36.27	18.50	54.77
	Pacific Islander	359	20.89	25.63	38.72	14.76	53.48
	White	67,092	18.72	24.96	41.88	14.44	56.32
NRC	New York City	50,648	24.94	23.99	32.76	18.30	51.06
	Big 4 Cities	6,025	48.61	24.43	21.91	5.05	26.96
	Urban/Suburban	11,905	40.13	29.48	25.84	4.55	30.39
	Rural	9,009	30.55	30.64	33.30	5.52	38.82
	Average Needs	40,020	20.11	27.61	41.29	10.99	52.28
	Low Needs	19,193	7.42	17.06	48.90	26.61	75.51
	Charter	14,580	19.33	23.79	38.12	18.76	56.88
	Religious or Independent	13,221	31.43	28.00	32.40	8.16	40.56
SWD	All Codes	23,236	53.79	26.46	16.80	2.95	19.75
SUA	All Codes	17,040	51.88	26.52	18.10	3.50	21.60
ELL	ELL=Y	19,065	54.70	29.23	14.37	1.70	16.07
ELL/SUA	SUA & ELL codes	1,561	67.58	24.22	7.75	0.45	8.20
SWD/SUA	SWD & SUA codes	12,883	59.66	25.31	13.27	1.76	15.03
ELL Test Language	Arabic	285	53.68	27.72	16.49	2.11	18.60
	Bengali	189	36.51	33.86	24.34	5.29	29.63
	Chinese (Traditional)	59	37.29	10.17	37.29	15.25	52.54
	Chinese (Simplified)	753	5.71	17.66	50.73	25.90	76.63
	Haitian-Creole	151	62.25	26.49	11.26	.	11.26
	Korean	28	14.29	7.14	46.43	32.14	78.57
	Russian	414	24.64	32.61	35.27	7.49	42.76
	Spanish	5,562	57.44	29.29	12.55	0.72	13.27
	All Translations	7,441	49.48	28.06	18.42	4.03	22.45

8.2.2.5. Mathematics Grade 7

Table 8.29 presents the mathematics Grade 7 performance level summaries and n-counts for key demographic subgroups. The percentage of proficient students was 57.55% for the State population. Female students and Male students performed similarly. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, or Charter schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (26% above the State population) and the lowest for African-American students (12% below the State population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (22% above the State population) and

the lowest for Big 4 Cities districts (28% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 30–38% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 50% below that for the State population. Among ELLs taking the language translation forms, the percentage of proficient students was the lowest for the Haitian-Creole subgroup and the highest for the Chinese (Simplified) subgroup.

Table 8.29. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	157,751	17.50	24.96	31.04	26.51	57.55
Gender	Female	77,504	16.64	25.91	31.70	25.75	57.45
	Male	80,209	18.33	24.03	30.41	27.24	57.65
	Non-Binary	38	5.26	28.95	36.84	28.95	65.79
Ethnicity	Asian	17,345	5.19	11.60	26.19	57.02	83.21
	African American	24,734	24.82	30.08	28.30	16.80	45.10
	Hispanic	45,161	22.64	30.75	30.21	16.40	46.61
	American Indian	1,155	22.51	28.92	27.88	20.69	48.57
	Multiracial	4,962	16.71	24.24	29.71	29.34	59.05
	Pacific Islander	359	15.88	23.68	30.64	29.81	60.45
	White	63,617	14.12	22.47	34.28	29.13	63.41
NRC	New York City	50,307	16.21	24.52	29.00	30.26	59.26
	Big 4 Cities	5,821	42.04	28.86	20.58	8.52	29.10
	Urban/Suburban	11,203	29.27	34.36	26.25	10.12	36.37
	Rural	8,939	21.97	31.73	32.13	14.17	46.30
	Average Needs	36,985	15.28	25.78	35.11	23.83	58.94
	Low Needs	18,556	5.87	14.93	34.83	44.37	79.20
	Charter	13,896	12.00	23.42	32.56	32.01	64.57
	Religious or Independent	12,044	27.77	25.81	28.19	18.23	46.42
SWD	All Codes	21,622	42.36	32.80	18.06	6.78	24.84
SUA	All Codes	16,385	41.13	31.71	18.83	8.33	27.16
ELL	ELL=Y	16,506	47.35	33.02	15.33	4.30	19.63
ELL/SUA	SUA & ELL codes	1,310	58.93	33.59	6.79	0.69	7.48
SWD/SUA	SWD & SUA codes	12,199	48.47	32.31	14.75	4.48	19.23
ELL Test Language	Arabic	226	40.27	30.97	23.89	4.87	28.76
	Bengali	164	23.17	28.05	32.32	16.46	48.78
	Chinese (Traditional)	46	32.61	13.04	26.09	28.26	54.35
	Chinese (Simplified)	591	5.58	13.71	29.78	50.93	80.71
	Haitian-Creole	142	50.70	31.69	16.90	0.70	17.60
	Korean	28	10.71	17.86	32.14	39.29	71.43

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Russian	459	12.85	31.15	35.73	20.26	55.99
	Spanish	5,441	47.11	34.70	15.31	2.89	18.20
	All Translations	7,097	40.50	32.18	18.67	8.65	27.32

8.2.2.6. Mathematics Grade 8

Table 8.30 presents the mathematics Grade 8 performance level summaries and n-counts for key demographic subgroups. The percentage of proficient students was 43.50% for the State population. That percentage was 4% higher for Female students than for Male students. Compared with the State population, the percentages of proficient students were higher for Asian, Multiracial, Pacific Islander, and White students; the same is true for students enrolled in New York City, Low Needs districts, Charter schools, or Religious or Independent schools. Across ethnic groups, the percentage of proficient students was the highest for Asian students (25% above the State population) and the lowest for African-American students (10% below the State population). Across NRC subgroups, the percentage of proficient students was the highest for Low Needs districts (19% above the State population) and the lowest for Big 4 Cities districts (about 25% below the State population). The percentages of proficient students for SWD, SUA, and ELL subgroups were about 25–28% below that for the State population. ELLs tested under accommodations had the lowest percentage of proficient students, 37% below that for the State population. Among ELLs taking the language translation forms, the percentage of proficient students was the lowest for the Spanish subgroup and the highest for the Chinese (Simplified) subgroup.

Table 8.30. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	94,063	35.80	20.70	29.51	13.99	43.50
Gender	Female	44,681	33.52	21.15	31.19	14.15	45.34
	Male	49,344	37.86	20.31	27.99	13.84	41.83
	Non-Binary	38	31.58	13.16	31.58	23.68	55.26
Ethnicity	Asian	7,598	16.77	14.53	32.31	36.39	68.70
	African American	17,063	46.80	19.72	23.74	9.74	33.48
	Hispanic	28,585	42.86	21.24	25.60	10.31	35.91
	American Indian	715	44.20	20.84	25.59	9.37	34.96
	Multiracial	2,921	39.61	19.62	28.38	12.39	40.77
	Pacific Islander	174	35.63	19.54	28.16	16.67	44.83
	White	36,377	28.14	22.21	35.13	14.53	49.66
NRC	New York City	28,004	33.57	19.85	28.26	18.32	46.58
	Big 4 Cities	4,918	66.80	14.25	13.40	5.55	18.95
	Urban/Suburban	8,593	56.87	21.15	18.67	3.32	21.99
	Rural	7,058	41.61	23.96	27.76	6.67	34.43

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Average Needs	24,503	33.76	24.11	33.38	8.75	42.13
	Low Needs	8,368	18.59	18.50	40.62	22.29	62.91
	Charter	8,421	28.45	17.88	30.91	22.75	53.66
	Religious or Independent	4,198	22.32	17.70	34.37	25.61	59.98
SWD	All Codes	16,627	66.04	17.36	13.08	3.52	16.60
SUA	All Codes	17,286	63.57	17.59	14.86	3.99	18.85
ELL	ELL=Y	6,942	67.00	17.76	12.27	2.97	15.24
ELL/SUA	SUA & ELL codes	1,729	81.26	12.38	5.73	0.64	6.37
SWD/SUA	SWD & SUA codes	13,921	68.93	16.53	11.66	2.88	14.54
ELL Test Language	Arabic	161	57.14	21.74	16.77	4.35	21.12
	Bengali	73	39.73	23.29	28.77	8.22	36.99
	Chinese (Traditional)	14	14.29	14.29	35.71	35.71	71.42
	Chinese (Simplified)	175	10.86	12.57	36.00	40.57	76.57
	Haitian-Creole	92	70.65	18.48	10.87	.	10.87
	Korean	11	9.09	9.09	45.45	36.36	81.81
	Russian	249	24.90	18.47	37.35	19.28	56.63
	Spanish	2,385	66.21	18.87	13.00	1.93	14.93
	All Translations	3,160	58.51	18.67	16.90	5.92	22.82

References

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. AERA. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Budescu, D. (1985). Efficiency of linear equating as a function of the length of the anchor test. *Journal of Educational Measurement, 22*(1), 13–20. http://www.jstor.org/stable/1434562

Cai, L., Thissen, D. J., & du Toit, S. (2022). *IRTPRO* (Version 6.0) [Computer software]. Vector Psychometric Group (VPG). Available from https://store.vpgcentral.com/

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dorans, N. J., & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 33–66). Lawrence Erlbaum Associates, Inc. https://doi.org/10.1002/j.2333-8504.1992.tb01440.x

Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement, 29*(4), 309–319. http://www.jstor.org/stable/1435087

Every Student Succeeds Act (ESSA), 20 U.S.C. § 6301 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer. https://doi.org/10.1007/978-94-017-1988-9

Individuals with Disabilities Education Act (IDEA). Pub L. No. 108–446 (2004). https://www.congress.gov/108/plaws/publ446/PLAW-108publ446.pdf

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

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kim S., & Kolen M. J. (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models* (Version 1.0). [Computer software]. Iowa Testing Programs, University of Iowa.

Kolen, M. J. (2004). *POLYEQUATE: A computer program for IRT true and observed scoring equating for dichotomously and polytomously scored tests*. [Computer software]. University of Iowa.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lee, W.-C. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement*, *47*(1), 1–17. https://doi.org/10.1111/j.1745-3984.2009.00096.x

Lee, W.-C., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement*, *26*(4), 412–432. https://doi.org/10.1177/014662102237797

Lee, W.-C., & Kolen, M. J. (2006, Revised 2008). *IRT-CLASS: A computer program for item response theory classification consistency and accuracy* (Version 2.0) [Computer software]. Center for Advanced Studies in Measurement and Assessment, The University of Iowa. Available at http://www.education.uiowa.edu/casma

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*(2), 179–197. http://www.jstor.org/stable/1435147

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge. https://doi.org/10.4324/9780203056615

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (3rd Ed.). Holt, Rinehart, and Winston.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176. https://doi.org/10.1177/014662169201600206

Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats. *Applied Measurement in Education, 8*(2), 111–120. https://doi.org/10.1207/s15324818ame0802_1

Sandoval, J., & Mille, M. W. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology*, *48*(2), 249–253. https://doi.org/10.1037/0022-006X.48.2.249

Sinharay, S. (2006). Bayesian item fit analysis for unidimensional item response theory models. *British Journal of Mathematical and Statistical Psychology*, *59*(2), 429–449. https://doi.org/10.1348/000711005X66888

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210. https://doi.org/10.1177/014662168300700208

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes. https://nceo.umn.edu/docs/onlinepubs/synth44.pdf

Wang, T., Kolen, M. J., & Harris, D. J. (2000). Psychometric properties of scale scores and performance levels for performance assessment using polytomous IRT. *Journal of Educational Measurement*, *37*(2), 141–162. https://psycnet.apa.org/doi/10.1111/j.1745-3984.2000.tb01080.x

Wingersky, M. S., Cook, L. L., & Eignor, D. R. (1987). *Specifying the characteristics of linking items used for item response theory item calibration* (Research Report No. RR-87-24). Educational Testing Service (ETS). https://doi.org/10.1002/j.2330-8516.1987.tb00228.x

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement, 21*(2), 93–111. https://www.jstor.org/stable/1434536

Appendix A: ELA and Mathematics Test Configurations and Testing Times

Table A1. ELA Test Configuration

Grade	Day	Session	Number of Items				
			Multiple-Choice		Constructed-Response		
			Operational	Embedded	Operational	Embedded	Total
3	1	1	17	6	2	0	25
	2	2	6	0	3	0	9
		Total	**23**	**6**	**5**	**0**	**34**
4	1	1	17	6	2	0	25
	2	2	6	0	4	0	10
		Total	**23**	**6**	**6**	**0**	**35**
5	1	1	19	7	2	0	28
	2	2	7	0	4	0	11
		Total	**26**	**7**	**6**	**0**	**39**
6	1	1	19	7	2	0	28
	2	2	7	0	4	0	11
		Total	**26**	**7**	**6**	**0**	**39**
7	1	1	19	7	2	0	28
	2	2	14	0	4	0	18
		Total	**33**	**7**	**6**	**0**	**46**
8	1	1	19	7	2	0	28
	2	2	14	0	4	0	18
		Total	**33**	**7**	**6**	**0**	**46**

Table A2. Mathematics Test Configuration

Grade	Day	Session	Number of Items				
			Multiple-Choice		Constructed-Response		
			Operational	Embedded	Operational	Embedded	Total
3	1	1	19	6	0	0	25
	2	2	5	0	8	0	13
		Total	**24**	**6**	**8**	**0**	**38**
4	1	1	23	7	0	0	30
	2	2	5	0	9	0	14
		Total	**28**	**7**	**9**	**0**	**44**
5	1	1	23	7	0	0	30
	2	2	5	0	9	0	14
		Total	**28**	**7**	**9**	**0**	**44**
6	1	1	23	7	0	0	30
	2	2	6	0	10	0	16
		Total	**29**	**7**	**10**	**0**	**46**
7	1	1	25	7	0	0	32
	2	2	6	0	10	0	16
		Total	**31**	**7**	**10**	**0**	**48**
8	1	1	25	7	0	0	32
	2	2	6	0	10	0	16
		Total	**31**	**7**	**10**	**0**	**48**

Table A3. ELA Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)
3	1	1	38
	2	2	25
		Total	**63**
4	1	1	38
	2	2	50
		Total	**88**
5	1	1	40
	2	2	51
		Total	**91**
6	1	1	40
	2	2	51
		Total	**91**
7	1	1	40
	2	2	63
		Total	**103**
8	1	1	40
	2	2	63
		Total	**103**

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 item—1 minute
- Average time to respond to a 2-point constructed-response item—3 minutes
- Average time to respond to a 4-point constructed-response item—20 minutes

Table A4. Mathematics Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)
3	1	1	28.5
	2	2	45.5
		Total	**74**
4	1	1	34.5
	2	2	50.5
		Total	**85**
5	1	1	34.5
	2	2	50.5
		Total	**85**
6	1	1	34.5
	2	2	57
		Total	**91.5**
7	1	1	37.5
	2	2	57
		Total	**94.5**

Grade	Day	Session	Estimated Time on Task (min.)
8	1	1	37.5
	2	2	57
		Total	**94.5**

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

- Average time to respond to a multiple-choice item—1.5 minutes
- Average time to respond to a 1-point constructed-response item – 3 minutes
- Average time to respond to a 2-point constructed-response item—5 minutes
- Average time to respond to a 3-point constructed-response item—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 2024 Teacher's Directions manuals located at https://www.nysed.gov/state-assessment/archive-2024-grades-3-8-ela-math-and-science-test-manuals and the 2024 *NYSTP English Language Arts, Mathematics, and Science Tests School Administrator's Manual* (*SAM*) located at https://www.nysed.gov/sites/default/files/programs/state-assessment/sam-g3-8-2024.pdf.

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Strand	Point Range		% of Test	
			Target	Actual	Target	Actual
3	33	Literature	18–21	21	55%–64%	64%
		Informational Text	12–15	12	36%–45%	36%
4	37	Literature	17–20	17	47%–53%	46%
		Informational Text	17–20	20	47%–53%	54%
5	40	Literature	18–22	24	45%–55%	60%
		Informational Text	18–22	16	45%–55%	40%
6	40	Literature	18–22	20	45%–55%	50%
		Informational Text	18–22	20	45%–55%	50%
7	47	Literature	20–25	23	43%–57%	49%
		Informational Text	20–25	24	43%–57%	51%
8	47	Literature	20–25	22	43%–57%	47%
		Informational Text	20–25	25	43%–57%	53%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
3	38	Operations and Algebraic Thinking	12–16	13	31%–43%	34%
		Number and Operations in Base Ten	3–5	4	7%–14%	11%
		Number and Operations—Fractions	7–11	10	18%–29%	26%
		Measurement and Data	8–12	9	21%–32%	24%
		Geometry [a]	1–3	2	2%–8%	5%
4	44	Operations and Algebraic Thinking	7–11	9	15%–25%	20%
		Number and Operations in Base Ten	9–13	11	20%–30%	25%
		Number and Operations—Fractions	9–13	11	20%–30%	25%
		Measurement and Data	4–6	5	9%–14%	11%
		Geometry	6–10	8	13%–23%	18%
5	44	Number and Operations in Base Ten	11–15	14	25%–35%	32%
		Number and Operations—Fractions	15–19	17	34%–44%	38%
		Measurement and Data	10–14	11	22%–32%	25%
		Geometry [a]	1–3	2	2%–7%	5%
6	47	Ratios and Proportional Relationships	10–14	10	21%–30%	21%
		The Number System	8–12	11	17%–26%	23%
		Expressions, Equations, and Inequalities	12–20	17	25%–43%	36%
		Geometry	6–11	9	14%–24%	19%
7	49	Ratios and Proportional Relationships	12–16	15	24%–33%	31%
		The Number System	8–12	11	16%–25%	22%
		Expressions, Equations, and Inequalities	13–19	14	26%–39%	29%
		Geometry	1–3	2	2%–7%	4%
		Statistics and Probability	6–10	7	12%–21%	14%

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
8	49	The Number System	1–4	3	2%–9%	6%
		Expressions, Equations, and Inequalities	14–20	16	28%–41%	33%
		Functions	8–12	10	16%–25%	20%
		Geometry	14–20	16	28%–41%	33%
		Statistics and Probability	2–5	4	4%–11%	8%

[a] There is a slight difference between the "Target % of Test" shown in these tables and the tables presented in the guides to the 2024 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 Next Generation 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 they graduate, students should be prepared to successfully read and analyze the types of complex texts 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 Next Generation 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 they will encounter in college and in the workplace. The array of passages selected for assessment in 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 for selecting passages for assessment, attention should also be dedicated to three other considerations:

- Text complexity
- Text types
- Text suitability for specific standards

These guidelines should inform the training of passage finders 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.
√	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.
√	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.
√	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 Grades 3–8 ELA and Mathematics Test items.

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 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 being tested
- presented at a reading level suitable for the grade 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 2024 operational test. The psychometrics department provides support, as necessary, and reviews the final item selection. The psychometrics department provides data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2024 item selection included 2019–2023 embedded and stand-alone field test items.

Here are the general guidelines for item selection:

- Satisfy the content specifications in terms of objective coverage and the number and percentage of MC and CR items on the test. An often-used criterion for objective coverage is within 5% of the percentages of score points and items per objective.
- To the extent possible, select both easy and difficult items to provide good measurement information at both ends of the performance scale.
- Avoid selecting items with too high/low *p* values, items with flagged point-biserials, and poorly fitting items.
- Minimize the number of items flagged for DIF (gender, ethnic, and High/Low Needs schools). Flagged items should be reviewed for content again. Keep in mind 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 provides 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

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2024 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

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-3.L.4	Language Standards	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.6	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.5	Reading Standards for Literature	Reading
13	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.4	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.8	Reading Standards for Informational Text	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.5	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.6	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.5	Reading Standards for Literature	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.8	Reading Standards for Literature	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.9	Reading Standards for Literature	Reading
24	Constructed Response	2	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Writing to Sources
25	Constructed Response	2	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Writing to Sources
Session 2					
26	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
27	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.4	Reading Standards for Literature	Reading
28	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.6	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
29	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
32	Constructed Response	2	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Informational Text	Writing to Sources
33	Constructed Response	2	NGLS.ELA.Content.NY-3.R.9	Reading Standards for Informational Text	Writing to Sources
34	Constructed Response	2	NGLS.ELA.Content.NY-3.R.7	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2024 operational ELA test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point constructed-response items 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 items, please refer to the rubrics shown in the Educator Guides.

Table G2. ELA Grade 4 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-4.L.4	Language Standards	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Informational Text	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.7	Reading Standards for Informational Text	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.8	Reading Standards for Informational Text	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
14	Multiple Choice	1	NGLS.ELA.Content.NY-4.L.4	Reading Standards for Literature	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Literature	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Literature	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.9	Reading Standards for Literature	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Literature	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Literature	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.8	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
23	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.9	Reading Standards for Literature	Reading
24	Constructed Response	2	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Literature	Writing to Sources
25	Constructed Response	2	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Writing to Sources
Session 2					
26	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Reading
27	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Reading
28	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Informational Text	Reading
29	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.7	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.9	Reading Standards for Informational Text	Reading
32	Constructed Response	2	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Informational Text	Writing to Sources
33	Constructed Response	2	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Writing to Sources
34	Constructed Response	2	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Informational Text	Writing to Sources
35	Constructed Response	4	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2024 operational ELA test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items 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 items, please refer to the rubrics shown in the Educator Guides.

Table G3. ELA Grade 5 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-5.L.4	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.5	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.5	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.9	Reading Standards for Literature	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.7	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Informational Text	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.9	Reading Standards for Informational Text	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.8	Reading Standards for Informational Text	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-5.R.7	Reading Standards for Informational Text	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.6	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.5	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Literature	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
34	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Reading
36	Constructed Response	2	NGLS.ELA.Content.NY-5.R.5	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Writing to Sources
39	Constructed Response	4	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2024 operational ELA test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items 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 items, please refer to the rubrics shown in the Educator Guides.

Table G4. ELA Grade 6 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-6.L.4	Language Standards	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.9	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Literature	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-6.L.4	Language Standards	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.8	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.6	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.9	Reading Standards for Informational Text	Reading

Item	Type	Points	Standard	Strand	Subscore Category
24	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Informational Text	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.6	Reading Standards for Informational Text	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Informational Text	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-6.R.8	Reading Standards for Informational Text	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Informational Text	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-6.L.4	Language Standards	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Literature	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.9	Reading Standards for Literature	Reading
36	Constructed Response	2	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Writing to Sources
39	Constructed Response	4	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Literature	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2024 operational ELA test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items 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 items, please refer to the rubrics shown in the Educator Guides.

Table G5. ELA Grade 7 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
5	Multiple Choice	1	NGLS.ELA.Content.NY-7.L.4	Language Standards	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.5	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Literature	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.9	Reading Standards for Literature	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-7.L.4	Language Standards	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Literature	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
36	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Reading
37	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.8	Reading Standards for Informational Text	Reading
38	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.8	Reading Standards for Informational Text	Reading
39	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.5	Reading Standards for Informational Text	Reading

Item	Type	Points	Standard	Strand	Subscore Category
40	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Informational Text	Reading
41	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Reading
42	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
43	Constructed Response	2	NGLS.ELA.Content.NY-7.R.8	Reading Standards for Informational Text	Writing to Sources
44	Constructed Response	2	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Writing to Sources
45	Constructed Response	2	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Informational Text	Writing to Sources
46	Constructed Response	4	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2024 operational ELA test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items 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 items, please refer to the rubrics shown in the Educator Guides.

Table G6. ELA Grade 8 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.8	Reading Standards for Literature	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
23	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Literature	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Literature	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.8	Reading Standards for Literature	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-8.R.8	Reading Standards for Literature	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.9	Reading Standards for Informational Text	Reading
36	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
37	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Reading
38	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
39	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
40	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Reading
41	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
42	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
43	Constructed Response	2	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Writing to Sources
44	Constructed Response	2	NGLS.ELA.Content.NY-8.R.9	Reading Standards for Literature	Writing to Sources
45	Constructed Response	2	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Writing to Sources
46	Constructed Response	4	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Literature	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2024 operational ELA test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items 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 items, please refer to the rubrics shown in the Educator Guides.

Table G7. Mathematics Grade 3 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
2	Multiple Choice	1	NGLS.Math.Content.NY-3.NBT.1	Number and Operations in Base Ten	
4	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.2b	Number and Operations—Fractions	Number and Operations—Fractions
5	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
7	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.1	Measurement and Data	Measurement and Data
8	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
9	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.7d	Measurement and Data	Measurement and Data
11	Multiple Choice	1	NGLS.Math.Content.NY-3. OA.8a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
12	Multiple Choice	1	NGLS.Math.Content.NY-3. NF.3c	Number and Operations—Fractions	Number and Operations—Fractions
13	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.5b	Measurement and Data	Measurement and Data
14	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.2a	Number and Operations—Fractions	Number and Operations—Fractions
16	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.9	Operations and Algebraic Thinking	Operations and Algebraic Thinking
17	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
18	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
20	Multiple Choice	1	NGLS.Math.Content.NY-3. MD.2b	Measurement and Data	Measurement and Data
21	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.1	Number and Operations—Fractions	Number and Operations—Fractions
23	Multiple Choice	1	NGLS.Math.Content.NY-3. MD.7a	Measurement and Data	Measurement and Data
24	Multiple Choice	1	NGLS.Math.Content.NY-3.G.2	Geometry	
25	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.6	Operations and Algebraic Thinking	Operations and Algebraic Thinking
Session 2					
26	Multiple Choice	1	NGLS.Math.Content.NY-3.G.2	Geometry	
27	Multiple Choice	1	NGLS.Math.Content.NY-3. OA.9a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
28	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.6	Measurement and Data	Measurement and Data
29	Multiple Choice	1	NGLS.Math.Content.NY-3. OA.8a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
30	Multiple Choice	1	NGLS.Math.Content.NY-3. NF.3b	Number and Operations—Fractions	Number and Operations—Fractions
31	Constructed Response	1	NGLS.Math.Content.NY-3.MD.1	Measurement and Data	Measurement and Data
32	Constructed Response	1	NGLS.Math.Content.NY-3.NBT.4b	Number and Operations in Base Ten	
33	Constructed Response	1	NGLS.Math.Content.NY-3. OA.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking

Item	Type	Points	Standard	Cluster	Subscore Category
34	Constructed Response	2	NGLS.Math.Content.NY-3. MD.7b	Measurement and Data	Measurement and Data
35	Constructed Response	2	NGLS.Math.Content.NY-3.NF.3.b	Number and Operations—Fractions	Number and Operations—Fractions
36	Constructed Response	2	NGLS.Math.Content.NY-3.NBT.3	Number and Operations in Base Ten	
37	Constructed Response	2	NGLS.Math.Content.NY-3. OA.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
38	Constructed Response	3	NGLS.Math.Content.NY-3.NF.3d	Number and Operations—Fractions	Number and Operations—Fractions

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

Table G8. Mathematics Grade 4 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
2	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.1	Number and Operations—Fractions	Number and Operations—Fractions
3	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
4	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.3d	Number and Operations—Fractions	Number and Operations—Fractions
5	Multiple Choice	1	NGLS.Math.Content.NY-3.G.1	Geometry	
7	Multiple Choice	1	NGLS.Math.Content.NY-4.MD.7	Measurement and Data	
8	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.5	Number and Operations in Base Ten	Number and Operations in Base Ten
10	Multiple Choice	1	NGLS.Math.Content.NY-4.G.1	Geometry	
11	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.3a	Number and Operations—Fractions	Number and Operations—Fractions
12	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.3	Number and Operations in Base Ten	Number and Operations in Base Ten
14	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
15	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.4c	Number and Operations—Fractions	Number and Operations—Fractions
16	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
18	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.2	Number and Operations—Fractions	Number and Operations—Fractions
19	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
21	Multiple Choice	1	NGLS.Math.Content.NY-4.G.3	Geometry	
22	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.3c	Number and Operations—Fractions	Number and Operations—Fractions
23	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.1	Number and Operations in Base Ten	Number and Operations in Base Ten
24	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.4b	Number and Operations—Fractions	Number and Operations—Fractions
26	Multiple Choice	1	NGLS.Math.Content.NY-4.G.2c	Geometry	

Item	Type	Points	Standard	Cluster	Subscore Category
27	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
29	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.5	Number and Operations in Base Ten	Number and Operations in Base Ten
30	Multiple Choice	1	NGLS.Math.Content.NY-4.MD.3	Measurement and Data	
Session 2					
31	Multiple Choice	1	NGLS.Math.Content.NY-4.G.3	Geometry	
32	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.3a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
33	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.2	Number and Operations—Fractions	Number and Operations—Fractions
34	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.5	Number and Operations in Base Ten	Number and Operations in Base Ten
35	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.3b	Number and Operations—Fractions	Number and Operations—Fractions
36	Constructed Response	1	NGLS.Math.Content.NY-4.MD.6	Measurement and Data	
37	Constructed Response	1	NGLS.Math.Content.NY-4.OA.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
38	Constructed Response	1	NGLS.Math.Content.NY-4.G.2a	Geometry	
39	Constructed Response	2	NGLS.Math.Content.NY-4.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
40	Constructed Response	2	NGLS.Math.Content.NY-4.NBT.2b	Number and Operations in Base Ten	Number and Operations in Base Ten
41	Constructed Response	2	NGLS.Math.Content.NY-4.MD.4	Measurement and Data	
42	Constructed Response	2	NGLS.Math.Content.NY-4.NF.4a	Number and Operations—Fractions	Number and Operations—Fractions
43	Constructed Response	2	NGLS.Math.Content.NY-4.G.1	Geometry	
44	Constructed Response	3	NGLS.Math.Content.NY-4.OA.3b	Operations and Algebraic Thinking	Operations and Algebraic Thinking

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

Table G9. Mathematics Grade 5 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.7	Number and Operations in Base Ten	Number and Operations in Base Ten
2	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.4a	Number and Operations—Fractions	Number and Operations—Fractions
3	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.3a	Measurement and Data	Measurement and Data
4	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.4	Number and Operations in Base Ten	Number and Operations in Base Ten
6	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
7	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.3	Number and Operations—Fractions	Number and Operations—Fractions
9	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.5c	Measurement and Data	Measurement and Data
10	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.7c	Number and Operations—Fractions	Number and Operations—Fractions
12	Multiple Choice	1	NGLS.Math.Content.NY-4.MD.2a	Measurement and Data	Measurement and Data
13	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.2	Number and Operations—Fractions	Number and Operations—Fractions
15	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
16	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.3b	Number and Operations in Base Ten	Number and Operations in Base Ten
17	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.6	Number and Operations—Fractions	Number and Operations—Fractions
19	Multiple Choice	1	NGLS.Math.Content.NY-5.G.3	Geometry	
20	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.7	Number and Operations in Base Ten	Number and Operations in Base Ten
21	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.7c	Number and Operations—Fractions	Number and Operations—Fractions
23	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.1	Measurement and Data	Measurement and Data
24	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.1	Number and Operations—Fractions	Number and Operations—Fractions
25	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.4	Measurement and Data	Measurement and Data
27	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.7a	Number and Operations—Fractions	Number and Operations—Fractions
28	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.2	Number and Operations—Fractions	Number and Operations—Fractions
29	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.2	Number and Operations in Base Ten	Number and Operations in Base Ten
30	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.4b	Number and Operations—Fractions	Number and Operations—Fractions
Session 2					
31	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.7	Number and Operations in Base Ten	Number and Operations in Base Ten
32	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.1	Number and Operations—Fractions	Number and Operations—Fractions
33	Multiple Choice	1	NGLS.Math.Content.NY-5.G.4	Geometry	
34	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.2	Number and Operations—Fractions	Number and Operations—Fractions
35	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.1	Measurement and Data	Measurement and Data

Item	Type	Points	Standard	Cluster	Subscore Category
36	Constructed Response	1	NGLS.Math.Content.NY-5.MD.5a	Measurement and Data	Measurement and Data
37	Constructed Response	1	NGLS.Math.Content.NY-5.NF.6	Number and Operations—Fractions	Number and Operations—Fractions
38	Constructed Response	1	NGLS.Math.Content.NY-5.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
39	Constructed Response	2	NGLS.Math.Content.NY-5.MD.2	Measurement and Data	Measurement and Data
40	Constructed Response	2	NGLS.Math.Content.NY-5.NF.4b	Number and Operations—Fractions	Number and Operations—Fractions
41	Constructed Response	2	NGLS.Math.Content.NY-5.NF.5a	Number and Operations—Fractions	Number and Operations—Fractions
42	Constructed Response	2	NGLS.Math.Content.NY-5.NBT.3a	Number and Operations in Base Ten	Number and Operations in Base Ten
43	Constructed Response	2	NGLS.Math.Content.NY-5.MD.5a	Measurement and Data	Measurement and Data
44	Constructed Response	3	NGLS.Math.Content.NY-5.NBT.7	Number and Operations in Base Ten	Number and Operations in Base Ten

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

Table G10. Mathematics Grade 6 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
2	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.2a	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
3	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.5	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
4	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.8	The Number System	The Number System
6	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.7c	The Number System	The Number System
7	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.7	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
8	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.3	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
9	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.1	The Number System	The Number System
11	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.8	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
12	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
14	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
15	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.1	The Number System	The Number System
16	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.8	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
18	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.6a	The Number System	The Number System
19	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.2b	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
20	Multiple Choice	1	NGLS.Math.Content.NY-6.G.2	Geometry	

Item	Type	Points	Standard	Cluster	Subscore Category
22	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.6	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
23	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
25	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.2c	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
26	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.6b	The Number System	The Number System
28	Multiple Choice	1	NGLS.Math.Content.NY-6.G.1	Geometry	
29	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.7a	The Number System	The Number System
30	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.5	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
Session 2					
31	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
32	Multiple Choice	1	NGLS.Math.Content.NY-6.G.4	Geometry	
33	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
34	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
35	Multiple Choice	1	NGLS.Math.Content.NY-6.G.2	Geometry	
36	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.7	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
37	Constructed Response	1	NGLS.Math.Content.NY-6.G.3	Geometry	
38	Constructed Response	1	NGLS.Math.Content.NY-6.EE.7	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
39	Constructed Response	1	NGLS.Math.Content.NY-6.RP.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
40	Constructed Response	2	NGLS.Math.Content.NY-6.G.1	Geometry	
41	Constructed Response	2	NGLS.Math.Content.NY-6.EE.4	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
42	Constructed Response	2	NGLS.Math.Content.NY-6.EE.9	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
43	Constructed Response	2	NGLS.Math.Content.NY-6.NS.4	The Number System	The Number System
44	Constructed Response	2	NGLS.Math.Content.NY-6.G.5	Geometry	
45	Constructed Response	2	NGLS.Math.Content.NY-6.NS.8	The Number System	The Number System
46	Constructed Response	3	NGLS.Math.Content.NY-6.RP.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships

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

Table G11. Mathematics Grade 7 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2b	Ratios and Proportional Relationships	Ratios and Proportional Relationships

Item	Type	Points	Standard	Cluster	Subscore Category
2	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System
3	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System
4	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.4a	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
5	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.1c	The Number System	The Number System
7	Multiple Choice	1	NGLS.Math.Content.NY-6.SP.2	Statistics and Probability	
8	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.3	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
10	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.4b	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
11	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.2	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
13	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
14	Multiple Choice	1	NGLS.Math.Content.NY-7.SP.1	Statistics and Probability	
15	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.1d	The Number System	The Number System
17	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.3	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
18	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
19	Multiple Choice	1	NGLS.Math.Content.NY-7.SP.8a	Statistics and Probability	
21	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
22	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.3	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
23	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
25	Multiple Choice	1	NGLS.Math.Content.NY-7.SP.4	Statistics and Probability	
26	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
27	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.1c	The Number System	The Number System
29	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
30	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.4b	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
31	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System
32	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
Session 2					
33	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.1a	The Number System	The Number System
34	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
35	Multiple Choice	1	NGLS.Math.Content.NY-7.SP.3	Statistics and Probability	
36	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System
37	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
38	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2b	Ratios and Proportional Relationships	Ratios and Proportional Relationships

Item	Type	Points	Standard	Cluster	Subscore Category
39	Constructed Response	1	NGLS.Math.Content.NY-7.RP.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
40	Constructed Response	1	NGLS.Math.Content.NY-7.NS.2c	The Number System	The Number System
41	Constructed Response	1	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
42	Constructed Response	1	NGLS.Math.Content.NY-7.NS.1d	The Number System	The Number System
43	Constructed Response	1	NGLS.Math.Content.NY-7.RP.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
44	Constructed Response	1	NGLS.Math.Content.NY-7.G.1	Geometry	
45	Constructed Response	1	NGLS.Math.Content.NY-7.EE.4a	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
46	Constructed Response	1	NGLS.Math.Content.NY-7.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
47	Constructed Response	1	NGLS.Math.Content.NY-7.SP.8a	Statistics and Probability	
48	Constructed Response	2	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships

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

Table G12. Mathematics Grade 8 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-8.F.4	Functions	Functions
2	Multiple Choice	1	NGLS.Math.Content.NY-8.SP.2	Statistics and Probability	
3	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
4	Multiple Choice	1	NGLS.Math.Content.NY-8.G.1b	Geometry	Geometry
5	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.7b	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
7	Multiple Choice	1	NGLS.Math.Content.NY-8.F.2	Functions	Functions
8	Multiple Choice	1	NGLS.Math.Content.NY-7.G.5	Geometry	Geometry
9	Multiple Choice	1	NGLS.Math.Content.NY-NY-8.EE.6	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
10	Multiple Choice	1	NGLS.Math.Content.NY-8.G.9	Geometry	Geometry
12	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.2	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
13	Multiple Choice	1	NGLS.Math.Content.NY-8.G.4	Geometry	Geometry
15	Multiple Choice	1	NGLS.Math.Content.NY-8.F.2	Functions	Functions
16	Multiple Choice	1	NGLS.Math.Content.NY-8.G.3	Geometry	Geometry
18	Multiple Choice	1	NGLS.Math.Content.NY-8.SP.3	Statistics and Probability	
19	Multiple Choice	1	NGLS.Math.Content.NY-8.G.2	Geometry	Geometry

Item	Type	Points	Standard	Cluster	Subscore Category
20	Multiple Choice	1	NGLS.Math.Content.NY-8.F.3	Functions	Functions
22	Multiple Choice	1	NGLS.Math.Content.NY-8.G.4	Geometry	Geometry
23	Multiple Choice	1	NGLS.Math.Content.NY-8.G.5	Geometry	Geometry
24	Multiple Choice	1	NGLS.Math.Content.NY-8.F.4	Functions	Functions
26	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.5	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
27	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.1	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
28	Multiple Choice	1	NGLS.Math.Content.NY-8.F.1	Functions	Functions
30	Multiple Choice	1	NGLS.Math.Content.NY-8.G.6	Geometry	Geometry
31	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.7a	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
32	Multiple Choice	1	NGLS.Math.Content.NY-7.G.4	Geometry	Geometry
Session 2					
33	Multiple Choice	1	NGLS.Math.Content.NY-8.F.3	Functions	Functions
34	Multiple Choice	1	NGLS.Math.Content.NY-8.G.1a	Geometry	Geometry
35	Multiple Choice	1	NGLS.Math.Content.NY-8.G.9	Geometry	Geometry
36	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.7a	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
37	Multiple Choice	1	NGLS.Math.Content.NY-8.NS.2	The Number System	
38	Multiple Choice	1	NGLS.Math.Content.NY-8.F.1	Functions	Functions
39	Constructed Response	1	NGLS.Math.Content.NY-8.G.9	Geometry	Geometry
40	Constructed Response	1	NGLS.Math.Content.NY-8.EE.2	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
41	Constructed Response	1	NGLS.Math.Content.NY-8.G.8	Geometry	Geometry
42	Constructed Response	2	NGLS.Math.Content.NY-8.G.5	Geometry	Geometry
43	Constructed Response	2	NGLS.Math.Content.NY-8.SP.3	Statistics and Probability	
44	Constructed Response	2	NGLS.Math.Content.NY-8.EE.5	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
45	Constructed Response	2	NGLS.Math.Content.NY-8.F.4	Functions	Functions
46	Constructed Response	2	NGLS.Math.Content.NY-8.EE.7b	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities
47	Constructed Response	2	NGLS.Math.Content.NY-8.NS.1	The Number System	
48	Constructed Response	3	NGLS.Math.Content.NY-8.EE.5	Expressions, Equations, and Inequalities	Expressions, Equations, and Inequalities

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

Appendix H: ELA Short-Response Rubric

2-Credit Rubric–Short Response

Score	Response Features
2 Credits	The features of a 2-credit 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 a 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 impact readability
1 Credit	The features of a 1-credit 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 a response according to the requirements of the prompt • Incomplete sentences or bullets
0 Credits	The features of a 0-credit 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

Note. If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 1. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item in that session completely blank (no response attempted).

Appendix I: ELA Extended-Response Rubrics

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 text(s)	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 text(s) 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

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:**
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

Note. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item 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	NGS	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 text(s)	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 text(s) 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

CRITERIA	NGS	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:**
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

Note. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item 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 Rubrics

1-Credit Constructed-Response Rubric

1 Credit	A 1-credit response is a correct answer to the question that indicates a thorough understanding of mathematical concepts and/or procedures.
0 Credits	A 0-credit response is incorrect, irrelevant, or incoherent.

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

2-Credit Constructed-Response Holistic Rubric

2 Credits	A 2-credit 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 Credit	A 1-credit 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 Credits	A 0-credit response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, they are not holistically sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.

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

Appendix K: Mathematics Extended-Response Rubric

3-Credit Constructed-Response Holistic Rubric

3 Credits	A 3-credit 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 Credits	A 2-credit 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 Credit	A 1-credit 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 the required work is limited
0 Credits	A 0-credit response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, they are not holistically sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.

Note. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item 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 2024 Grades 3–8 ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: English Language Learners (ELLs), students with disabilities (SWDs), and students using test accommodations (SUAs). Tables L1 through L12 contain the results of factor analysis on the subpopulation data for the Grades 3–8 ELA and Mathematics Tests, respectively.

Table L1. ELA Grade 3 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
			Variance Accounted for	
	N	Eigenvalue	%	Cumulative %
ELL	1	**4.99**	**17.82**	**17.82**
	2	1.50	5.37	23.19
	3	1.11	3.97	27.16
	4	1.06	3.79	30.95
	5	1.01	3.60	34.54
SWD	1	**5.96**	**21.27**	**21.27**
	2	1.37	4.88	26.15
	3	1.05	3.76	29.91
	4	1.03	3.66	33.58
SUA	1	**5.74**	**20.48**	**20.48**
	2	1.38	4.93	25.42
	3	1.07	3.83	29.24
	4	1.03	3.69	32.93

Table L2. ELA Grade 4 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
			Variance Accounted for	
	N	Eigenvalue	%	Cumulative %
ELL	1	**5.53**	**19.06**	**19.06**
	2	1.38	4.76	23.82
	3	1.19	4.11	27.93
	4	1.14	3.94	31.86
	5	1.04	3.59	35.45
	6	1.01	3.49	38.94
	7	1.00	3.45	42.40
SWD	1	**6.27**	**21.63**	**21.63**
	2	1.41	4.86	26.50
	3	1.17	4.04	30.54
	4	1.10	3.78	34.32
SUA	1	**6.22**	**21.45**	**21.45**

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
	2	1.41	4.87	26.33
	3	1.17	4.05	30.38
	4	1.08	3.73	34.11

Table L3. ELA Grade 5 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL	1	**4.19**	**13.10**	**13.10**
	2	1.31	4.09	17.19
	3	1.10	3.44	20.63
	4	1.07	3.33	23.96
	5	1.05	3.29	27.26
	6	1.04	3.26	30.52
	7	1.03	3.21	33.73
	8	1.02	3.17	36.90
SWD	1	**5.57**	**17.40**	**17.40**
	2	1.40	4.37	21.77
	3	1.08	3.37	25.14
	4	1.03	3.21	28.35
	5	1.01	3.16	31.51
	6	1.00	3.13	34.64
SUA	1	**5.68**	**17.75**	**17.75**
	2	1.41	4.40	22.14
	3	1.07	3.33	25.48
	4	1.03	3.22	28.70
	5	1.01	3.15	31.85

Table L4. ELA Grade 6 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL	1	**4.88**	**15.24**	**15.24**
	2	1.58	4.95	20.19
	3	1.15	3.60	23.79
	4	1.09	3.42	27.21
	5	1.06	3.30	30.51
	6	1.04	3.26	33.77
	7	1.03	3.23	37.00

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
	8	1.02	3.20	40.20
	9	1.01	3.15	43.35
SWD	1	**5.89**	**18.42**	**18.42**
	2	1.50	4.67	23.09
	3	1.15	3.60	26.69
	4	1.03	3.23	29.93
	5	1.00	3.14	33.06
SWD/SUA	1	**6.12**	**19.13**	**19.13**
	2	1.48	4.63	23.76
	3	1.14	3.56	27.32
	4	1.02	3.20	30.53
	5	1.00	3.13	33.66

Table L5. ELA Grade 7 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL	1	**5.94**	**15.22**	**15.22**
	2	1.58	4.05	19.27
	3	1.18	3.03	22.30
	4	1.14	2.91	25.22
	5	1.09	2.79	28.01
	6	1.08	2.76	30.77
	7	1.06	2.73	33.50
	8	1.03	2.65	36.15
	9	1.03	2.63	38.78
	10	1.00	2.57	41.35
SWD	1	**6.73**	**17.25**	**17.25**
	2	1.61	4.12	21.37
	3	1.15	2.95	24.31
	4	1.07	2.75	27.06
	5	1.04	2.68	29.74
	6	1.02	2.62	32.36
	7	1.00	2.57	34.93
SUA	1	**7.05**	**18.08**	**18.08**
	2	1.59	4.08	22.15
	3	1.14	2.93	25.08
	4	1.05	2.70	27.78

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	5	1.03	2.63	30.41
	6	1.01	2.58	32.99

Table L6. ELA Grade 8 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**5.45**	**13.97**	**13.97**
	2	1.54	3.95	17.92
	3	1.15	2.96	20.88
	4	1.10	2.82	23.7
	5	1.10	2.81	26.51
	6	1.04	2.67	29.18
	7	1.02	2.62	31.8
	8	1.01	2.6	34.4
SWD	1	**7.87**	**20.19**	**20.19**
	2	1.55	3.98	24.17
	3	1.18	3.02	27.19
	4	1.08	2.76	29.95
	5	1.02	2.61	32.56
SUA	1	**8.49**	**21.77**	**21.77**
	2	1.52	3.91	25.67
	3	1.16	2.99	28.66
	4	1.06	2.71	31.37
	5	1.01	2.58	33.96

Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**7.84**	**24.49**	**24.49**
	2	1.63	5.10	29.59
	3	1.15	3.59	33.18
	4	1.10	3.45	36.63
	5	1.00	3.14	39.76
SWD	1	**8.97**	**28.04**	**28.04**
	2	1.35	4.21	32.25
	3	1.08	3.39	35.64
	4	1.02	3.17	38.81

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
SUA	1	**8.32**	**26.01**	**26.01**
	2	1.32	4.13	30.14
	3	1.10	3.44	33.58
	4	1.03	3.21	36.79

Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**8.50**	**22.98**	**22.98**
	2	1.56	4.21	27.19
	3	1.22	3.30	30.49
	4	1.11	3.00	33.49
	5	1.05	2.84	36.33
SWD	1	**9.94**	**26.86**	**26.86**
	2	1.40	3.77	30.64
	3	1.17	3.17	33.80
	4	1.04	2.80	36.60
	5	1.00	2.72	39.32
SUA	1	**9.43**	**25.48**	**25.48**
	2	1.40	3.79	29.26
	3	1.18	3.20	32.46
	4	1.05	2.84	35.30
	5	1.01	2.73	38.04

Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**6.86**	**18.54**	**18.54**
	2	1.72	4.65	23.19
	3	1.30	3.52	26.71
	4	1.08	2.92	29.63
	5	1.05	2.84	32.47
	6	1.01	2.73	35.19
SWD	1	**8.17**	**22.08**	**22.08**
	2	1.70	4.59	26.67
	3	1.26	3.40	30.07
	4	1.03	2.79	32.86

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
SUA	1	**7.99**	**21.60**	**21.60**
	2	1.67	4.51	26.11
	3	1.26	3.42	29.52
	4	1.04	2.80	32.33
	5	1.01	2.74	35.06

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL	1	**6.73**	**17.26**	**17.26**
	2	1.39	3.57	20.83
	3	1.27	3.26	24.09
	4	1.18	3.02	27.11
	5	1.08	2.76	29.86
	6	1.06	2.72	32.59
	7	1.05	2.69	35.28
	8	1.01	2.6	37.88
SWD	1	**8.47**	**21.71**	**21.71**
	2	1.37	3.52	25.23
	3	1.11	2.84	28.06
	4	1.02	2.62	30.69
SUA	1	**8.87**	**22.75**	**22.75**
	2	1.38	3.55	26.30
	3	1.10	2.83	29.13
	4	1.01	2.58	31.71

Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL	1	**6.32**	**15.41**	**15.41**
	2	1.88	4.58	19.99
	3	1.41	3.43	23.42
	4	1.18	2.89	26.30
	5	1.09	2.66	28.96
	6	1.08	2.63	31.58
	7	1.06	2.58	34.16
	8	1.01	2.48	36.63

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
SWD	1	**7.79**	**18.99**	**18.99**
	2	1.71	4.17	23.16
	3	1.30	3.18	26.34
	4	1.13	2.75	29.09
	5	1.07	2.61	31.70
	6	1.03	2.52	34.22
	7	1.00	2.44	36.66
SUA	1	**8.31**	**20.27**	**20.27**
	2	1.70	4.14	24.41
	3	1.32	3.22	27.63
	4	1.11	2.71	30.34
	5	1.05	2.56	32.90
	6	1.02	2.48	35.38

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for %	Variance Accounted for Cumulative %
ELL	1	**7.01**	**17.10**	**17.10**
	2	1.27	3.10	20.20
	3	1.14	2.78	22.99
	4	1.08	2.63	25.62
	5	1.05	2.57	28.19
	6	1.03	2.51	30.70
	7	1.02	2.50	33.20
	8	1.01	2.47	35.66
	9	1.00	2.44	38.10
SWD	1	**7.69**	**18.75**	**18.75**
	2	1.31	3.19	21.94
	3	1.10	2.69	24.64
	4	1.08	2.64	27.27
	5	1.01	2.46	29.73
SUA	1	**8.02**	**19.55**	**19.55**
	2	1.33	3.24	22.8
	3	1.11	2.70	25.49
	4	1.08	2.64	28.13
	5	1.01	2.46	30.59

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, percent of omitted responses, *p* value, and the point-biserial correlations (PBis). 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	% Omit	*P* Value	PBis
1	MC	164,051	0	0.77	0.39
2	MC	164,051	0	0.71	0.51
3	MC	164,051	0	0.62	0.40
4	MC	164,051	0	0.83	0.37
5	MC	164,051	0	0.68	0.49
6	MC	164,051	0	0.52	0.36
13	MC	164,051	0	0.65	0.48
14	MC	164,051	0	0.60	0.37
15	MC	164,051	0	0.62	0.45
16	MC	164,051	0	0.40	0.41
17	MC	164,051	0	0.55	0.41
18	MC	164,051	0	0.51	0.33
19	MC	164,051	0	0.48	0.40
20	MC	164,051	0	0.42	0.32
21	MC	164,051	0	0.53	0.43
22	MC	164,051	0	0.54	0.40
23	MC	164,051	0	0.56	0.52
24	CR	164,051	–	0.48	0.61
25	CR	164,051	–	0.51	0.64
26	MC	164,051	0	0.38	0.42
27	MC	164,051	0	0.56	0.40
28	MC	164,051	0	0.65	0.44
29	MC	164,051	0	0.57	0.47
30	MC	164,051	0	0.43	0.28
31	MC	164,051	0	0.56	0.48
32	CR	164,051	–	0.62	0.57
33	CR	164,051	–	0.50	0.57
34	CR	164,051	–	0.51	0.59

Table M2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	168,586	0	0.57	0.38
2	MC	168,586	0	0.85	0.35
3	MC	168,586	0	0.72	0.44
4	MC	168,586	0	0.60	0.28

Item	Type	N-Count	% Omit	*P* Value	PBis
5	MC	168,586	0	0.77	0.51
6	MC	168,586	0	0.73	0.46
13	MC	168,586	0	0.71	0.40
14	MC	168,586	0	0.76	0.47
15	MC	168,586	0	0.72	0.37
16	MC	168,586	0	0.44	0.36
17	MC	168,586	0	0.76	0.46
18	MC	168,586	0	0.45	0.32
19	MC	168,586	0	0.46	0.36
20	MC	168,586	0	0.49	0.37
21	MC	168,586	0	0.52	0.35
22	MC	168,586	0	0.67	0.44
23	MC	168,586	0	0.53	0.46
24	CR	168,586	–	0.62	0.60
25	CR	168,586	–	0.57	0.64
26	MC	168,586	0	0.70	0.47
27	MC	168,586	0	0.48	0.29
28	MC	168,586	0	0.84	0.38
29	MC	168,586	0	0.61	0.39
30	MC	168,586	0	0.59	0.35
31	MC	168,586	0	0.71	0.36
32	CR	168,586	–	0.62	0.59
33	CR	168,586	–	0.68	0.65
34	CR	168,586	–	0.58	0.58
35	CR	168,586	–	0.45	0.67

Table M3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	158,764	0	0.43	0.20
2	MC	158,764	0	0.55	0.39
3	MC	158,764	0	0.59	0.34
4	MC	158,764	0	0.52	0.25
5	MC	158,764	0	0.63	0.31
6	MC	158,764	0	0.54	0.33
7	MC	158,764	0	0.67	0.39
15	MC	158,764	0	0.40	0.28
16	MC	158,764	0	0.51	0.38
17	MC	158,764	0	0.35	0.24
18	MC	158,764	0	0.81	0.46
19	MC	158,764	0	0.44	0.44
20	MC	158,764	0	0.64	0.31
21	MC	158,764	0	0.61	0.48

Item	Type	N-Count	% Omit	*P* Value	PBis
22	MC	158,764	0	0.43	0.17
23	MC	158,764	0	0.61	0.37
24	MC	158,764	0	0.51	0.29
25	MC	158,764	0	0.38	0.23
26	MC	158,764	0	0.34	0.14
27	CR	158,764	–	0.64	0.60
28	CR	158,764	–	0.61	0.54
29	MC	158,764	0	0.68	0.27
30	MC	158,764	0	0.66	0.39
31	MC	158,764	0	0.71	0.44
32	MC	158,764	0	0.59	0.41
33	MC	158,764	0	0.61	0.40
34	MC	158,764	0	0.63	0.30
35	MC	158,764	0	0.57	0.33
36	CR	158,764	–	0.64	0.56
37	CR	158,764	–	0.65	0.59
38	CR	158,764	–	0.59	0.59
39	CR	158,764	–	0.44	0.64

Table M4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	164,803	0	0.45	0.41
2	MC	164,803	0	0.61	0.33
3	MC	164,803	0	0.73	0.41
4	MC	164,803	0	0.60	0.34
5	MC	164,803	0	0.57	0.38
6	MC	164,803	0	0.58	0.46
7	MC	164,803	0	0.68	0.36
15	MC	164,803	0	0.81	0.43
16	MC	164,803	0	0.37	0.24
17	MC	164,803	0	0.40	0.27
18	MC	164,803	0	0.58	0.43
19	MC	164,803	0	0.59	0.40
20	MC	164,803	0	0.76	0.47
21	MC	164,803	0	0.38	0.23
22	MC	164,803	0	0.43	0.38
23	MC	164,803	0	0.40	0.20
24	MC	164,803	0	0.53	0.31
25	MC	164,803	0	0.62	0.43
26	MC	164,803	0	0.43	0.25
27	CR	164,803	–	0.74	0.60
28	CR	164,803	–	0.77	0.61

Item	Type	N-Count	% Omit	*P* Value	PBis
29	MC	164,803	0	0.75	0.35
30	MC	164,803	0	0.71	0.45
31	MC	164,803	0	0.42	0.23
32	MC	164,803	0	0.54	0.29
33	MC	164,803	0	0.72	0.39
34	MC	164,803	0	0.67	0.38
35	MC	164,803	0	0.74	0.39
36	CR	164,803	–	0.72	0.61
37	CR	164,803	–	0.64	0.58
38	CR	164,803	–	0.72	0.64
39	CR	164,803	–	0.44	0.66

Table M5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	159,817	0	0.46	0.39
2	MC	159,817	0	0.62	0.32
3	MC	159,817	0	0.71	0.36
4	MC	159,817	0	0.70	0.36
5	MC	159,817	0	0.91	0.39
6	MC	159,817	0	0.64	0.26
7	MC	159,817	0	0.74	0.44
15	MC	159,817	0	0.63	0.31
16	MC	159,817	0	0.68	0.50
17	MC	159,817	0	0.77	0.52
18	MC	159,817	0	0.59	0.31
19	MC	159,817	0	0.49	0.32
20	MC	159,817	0	0.46	0.34
21	MC	159,817	0	0.79	0.50
22	MC	159,817	0	0.59	0.19
23	MC	159,817	0	0.46	0.39
24	MC	159,817	0	0.54	0.23
25	MC	159,817	0	0.32	0.14
26	MC	159,817	0	0.69	0.48
27	CR	159,817	–	0.67	0.63
28	CR	159,817	–	0.68	0.58
29	MC	159,817	0	0.83	0.44
30	MC	159,817	0	0.73	0.48
31	MC	159,817	0	0.73	0.35
32	MC	159,817	0	0.64	0.37
33	MC	159,817	0	0.60	0.29
34	MC	159,817	0	0.73	0.25
35	MC	159,817	0	0.56	0.32

Item	Type	N-Count	% Omit	*P* Value	PBis
36	MC	159,817	0	0.59	0.28
37	MC	159,817	0	0.30	0.27
38	MC	159,817	0	0.53	0.39
39	MC	159,817	0	0.59	0.41
40	MC	159,817	0	0.67	0.38
41	MC	159,817	0	0.52	0.23
42	MC	159,817	0	0.50	0.35
43	CR	159,817	–	0.82	0.58
44	CR	159,817	–	0.75	0.61
45	CR	159,817	–	0.77	0.65
46	CR	159,817	–	0.56	0.67

Table M6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	147,257	0	0.70	0.48
2	MC	147,257	0	0.71	0.36
3	MC	147,257	0	0.70	0.37
4	MC	147,257	0	0.77	0.44
5	MC	147,257	0	0.50	0.34
6	MC	147,257	0	0.73	0.26
7	MC	147,257	0	0.48	0.32
15	MC	147,257	0	0.72	0.47
16	MC	147,257	0	0.55	0.33
17	MC	147,257	0	0.65	0.40
18	MC	147,257	0	0.61	0.30
19	MC	147,257	0	0.58	0.49
20	MC	147,257	0	0.61	0.46
21	MC	147,257	0	0.80	0.58
22	MC	147,257	0	0.76	0.50
23	MC	147,257	0	0.55	0.39
24	MC	147,257	0	0.65	0.46
25	MC	147,257	0	0.78	0.50
26	MC	147,257	0	0.72	0.41
27	CR	147,257	–	0.76	0.63
28	CR	147,257	–	0.75	0.62
29	MC	147,257	0	0.57	0.33
30	MC	147,257	0	0.49	0.39
31	MC	147,257	0	0.80	0.48
32	MC	147,257	0	0.84	0.45
33	MC	147,257	0	0.68	0.49
34	MC	147,257	0	0.55	0.47
35	MC	147,257	0	0.62	0.40

Item	Type	N-Count	% Omit	*P* Value	PBis
36	MC	147,257	0	0.49	0.33
37	MC	147,257	0	0.81	0.53
38	MC	147,257	0	0.73	0.36
39	MC	147,257	0	0.70	0.52
40	MC	147,257	0	0.72	0.45
41	MC	147,257	0	0.68	0.42
42	MC	147,257	0	0.70	0.56
43	CR	147,257	–	0.71	0.60
44	CR	147,257	–	0.67	0.55
45	CR	147,257	–	0.76	0.64
46	CR	147,257	–	0.48	0.66

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	156,673	0	0.73	0.44
2	MC	156,673	0	0.70	0.42
4	MC	156,673	0	0.80	0.42
5	MC	156,673	0	0.76	0.49
7	MC	156,673	0	0.66	0.48
8	MC	156,673	0	0.82	0.46
9	MC	156,673	0	0.57	0.51
11	MC	156,673	0	0.58	0.56
12	MC	156,673	0	0.34	0.38
13	MC	156,673	0	0.62	0.33
14	MC	156,673	0	0.75	0.43
16	MC	156,673	0	0.63	0.54
17	MC	156,673	0	0.77	0.57
18	MC	156,673	0	0.70	0.48
20	MC	156,673	0	0.68	0.28
21	MC	156,673	0	0.85	0.39
23	MC	156,673	0	0.59	0.57
24	MC	156,673	0	0.38	0.38
25	MC	156,673	0.01	0.75	0.52
26	MC	156,673	0	0.86	0.36
27	MC	156,673	0	0.72	0.52
28	MC	156,673	0	0.79	0.31
29	MC	156,673	0	0.41	0.29
30	MC	156,673	0	0.52	0.43
31	CR	156,673	–	0.73	0.51
32	CR	156,673	–	0.43	0.40
33	CR	156,673	–	0.66	0.59
34	CR	156,673	–	0.52	0.67

Item	Type	N-Count	% Omit	*P* Value	PBis
35	CR	156,673	–	0.39	0.63
36	CR	156,673	–	0.47	0.69
37	CR	156,673	–	0.42	0.65
38	CR	156,673	–	0.24	0.52

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	157,499	0	0.91	0.42
2	MC	157,499	0	0.77	0.50
3	MC	157,499	0	0.71	0.48
4	MC	157,499	0	0.88	0.40
5	MC	157,499	0	0.65	0.34
7	MC	157,499	0	0.67	0.49
8	MC	157,499	0	0.72	0.55
10	MC	157,499	0	0.74	0.46
11	MC	157,499	0	0.77	0.55
12	MC	157,499	0	0.82	0.48
14	MC	157,499	0	0.32	0.38
15	MC	157,499	0	0.69	0.44
16	MC	157,499	0	0.73	0.56
18	MC	157,499	0	0.68	0.36
19	MC	157,499	0	0.46	0.48
21	MC	157,499	0	0.72	0.40
22	MC	157,499	0	0.46	0.52
23	MC	157,499	0	0.60	0.46
24	MC	157,499	0	0.58	0.43
26	MC	157,499	0	0.60	0.36
27	MC	157,499	0	0.69	0.55
29	MC	157,499	0	0.69	0.56
30	MC	157,499	0	0.54	0.54
31	MC	157,499	0	0.58	0.06
32	MC	157,499	0	0.58	0.47
33	MC	157,499	0	0.74	0.40
34	MC	157,499	0	0.67	0.46
35	MC	157,499	0	0.63	0.58
36	CR	157,499	–	0.63	0.48
37	CR	157,499	–	0.58	0.57
38	CR	157,499	–	0.52	0.53
39	CR	157,499	–	0.40	0.66
40	CR	157,499	–	0.53	0.60
41	CR	157,499	–	0.43	0.62
42	CR	157,499	–	0.46	0.63

Item	Type	N-Count	% Omit	*P* Value	PBis
43	CR	157,499	–	0.41	0.56
44	CR	157,499	–	0.38	0.72

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	152,137	0	0.89	0.31
2	MC	152,137	0	0.75	0.42
3	MC	152,137	0	0.70	0.42
4	MC	152,137	0	0.62	0.44
6	MC	152,137	0	0.80	0.48
7	MC	152,137	0	0.43	0.51
9	MC	152,137	0	0.51	0.48
10	MC	152,137	0	0.71	0.47
12	MC	152,137	0	0.54	0.48
13	MC	152,137	0	0.36	0.47
15	MC	152,137	0	0.42	0.40
16	MC	152,137	0	0.83	0.38
17	MC	152,137	0	0.39	0.46
19	MC	152,137	0	0.48	0.35
20	MC	152,137	0	0.71	0.39
21	MC	152,137	0	0.64	0.43
23	MC	152,137	0	0.45	0.48
24	MC	152,137	0	0.63	0.51
25	MC	152,137	0	0.65	0.37
27	MC	152,137	0	0.72	0.46
28	MC	152,137	0	0.46	0.50
29	MC	152,137	0	0.45	0.35
30	MC	152,137	0	0.27	0.39
31	MC	152,137	0	0.73	0.54
32	MC	152,137	0	0.61	0.53
33	MC	152,137	0	0.38	0.38
34	MC	152,137	0	0.71	0.57
35	MC	152,137	0	0.49	0.27
36	CR	152,137	–	0.57	0.54
37	CR	152,137	–	0.12	0.24
38	CR	152,137	–	0.62	0.57
39	CR	152,137	–	0.49	0.70
40	CR	152,137	–	0.27	0.58
41	CR	152,137	–	0.22	0.50
42	CR	152,137	–	0.24	0.45
43	CR	152,137	–	0.52	0.64
44	CR	152,137	–	0.18	0.62

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	154,141	0	0.91	0.28
2	MC	154,141	0	0.71	0.46
3	MC	154,141	0	0.71	0.54
4	MC	154,141	0	0.62	0.48
6	MC	154,141	0	0.42	0.34
7	MC	154,141	0	0.59	0.38
8	MC	154,141	0	0.57	0.53
9	MC	154,141	0	0.52	0.60
11	MC	154,141	0	0.55	0.37
12	MC	154,141	0	0.79	0.48
14	MC	154,141	0	0.58	0.55
15	MC	154,141	0	0.54	0.63
16	MC	154,141	0	0.55	0.43
18	MC	154,141	0	0.28	0.31
19	MC	154,141	0	0.53	0.36
20	MC	154,141	0	0.44	0.43
22	MC	154,141	0	0.58	0.49
23	MC	154,141	0	0.54	0.56
25	MC	154,141	0	0.50	0.30
26	MC	154,141	0	0.38	0.30
28	MC	154,141	0	0.35	0.35
29	MC	154,141	0	0.47	0.32
30	MC	154,141	0	0.70	0.52
31	MC	154,141	0	0.89	0.39
32	MC	154,141	0	0.53	0.35
33	MC	154,141	0	0.50	0.49
34	MC	154,141	0	0.73	0.47
35	MC	154,141	0	0.54	0.57
36	MC	154,141	0	0.71	0.55
37	CR	154,141	–	0.20	0.46
38	CR	154,141	–	0.58	0.62
39	CR	154,141	–	0.37	0.61
40	CR	154,141	–	0.34	0.70
41	CR	154,141	–	0.37	0.61
42	CR	154,141	–	0.28	0.56
43	CR	154,141	–	0.42	0.58
44	CR	154,141	–	0.30	0.56
45	CR	154,141	–	0.45	0.66
46	CR	154,141	–	0.39	0.70

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	148,736	0	0.89	0.33
2	MC	148,736	0	0.85	0.35
3	MC	148,736	0	0.81	0.45
4	MC	148,736	0	0.68	0.56
5	MC	148,736	0	0.52	0.40
7	MC	148,736	0	0.59	0.44
8	MC	148,736	0	0.57	0.48
10	MC	148,736	0	0.49	0.40
11	MC	148,736	0	0.30	0.39
13	MC	148,736	0	0.68	0.35
14	MC	148,736	0	0.58	0.38
15	MC	148,736	0	0.69	0.38
17	MC	148,736	0	0.40	0.51
18	MC	148,736	0	0.53	0.43
19	MC	148,736	0	0.68	0.42
21	MC	148,736	0	0.52	0.40
22	MC	148,736	0	0.44	0.55
23	MC	148,736	0	0.55	0.45
25	MC	148,736	0	0.50	0.52
26	MC	148,736	0	0.47	0.41
27	MC	148,736	0	0.38	0.33
29	MC	148,736	0	0.67	0.39
30	MC	148,736	0	0.28	0.15
31	MC	148,736	0	0.70	0.29
32	MC	148,736	0	0.57	0.44
33	MC	148,736	0	0.78	0.39
34	MC	148,736	0	0.75	0.45
35	MC	148,736	0	0.63	0.38
36	MC	148,736	0	0.55	0.35
37	MC	148,736	0	0.60	0.38
38	MC	148,736	0	0.65	0.44
39	CR	148,736	–	0.31	0.56
40	CR	148,736	–	0.81	0.33
41	CR	148,736	–	0.33	0.58
42	CR	148,736	–	0.60	0.61
43	CR	148,736	–	0.60	0.60
44	CR	148,736	–	0.72	0.57
45	CR	148,736	–	0.67	0.67
46	CR	148,736	–	0.38	0.64
47	CR	148,736	–	0.36	0.63
48	CR	148,736	–	0.25	0.68

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	90,895	0	0.75	0.42
2	MC	90,895	0	0.68	0.41
3	MC	90,895	0	0.66	0.44
4	MC	90,895	0	0.47	0.51
5	MC	90,895	0	0.50	0.41
7	MC	90,895	0	0.47	0.24
8	MC	90,895	0	0.56	0.52
9	MC	90,895	0	0.31	0.42
10	MC	90,895	0	0.51	0.49
12	MC	90,895	0	0.79	0.47
13	MC	90,895	0	0.44	0.50
15	MC	90,895	0	0.57	0.44
16	MC	90,895	0	0.48	0.34
18	MC	90,895	0	0.59	0.39
19	MC	90,895	0	0.39	0.31
20	MC	90,895	0	0.63	0.49
22	MC	90,895	0	0.57	0.32
23	MC	90,895	0	0.47	0.39
24	MC	90,895	0	0.47	0.53
26	MC	90,895	0	0.53	0.43
27	MC	90,895	0	0.65	0.39
28	MC	90,895	0	0.60	0.33
30	MC	90,895	0	0.46	0.33
31	MC	90,895	0	0.38	0.31
32	MC	90,895	0	0.46	0.49
33	MC	90,895	0	0.71	0.39
34	MC	90,895	0	0.68	0.43
35	MC	90,895	0	0.51	0.43
36	MC	90,895	0	0.50	0.43
37	MC	90,895	0	0.65	0.40
38	MC	90,895	0	0.55	0.39
39	CR	90,895	–	0.36	0.56
40	CR	90,895	–	0.22	0.49
41	CR	90,895	–	0.19	0.40
42	CR	90,895	–	0.22	0.55
43	CR	90,895	–	0.11	0.43
44	CR	90,895	–	0.39	0.66
45	CR	90,895	–	0.27	0.66
46	CR	90,895	–	0.34	0.65
47	CR	90,895	–	0.23	0.61
48	CR	90,895	–	0.46	0.74

Appendix N: IRT Statistics

Tables N1 through N12 show item-calibration results for the operational items.

Table N1. ELA Grade 3 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.651	0.552				
2	1	0.773	-0.270				
3	1	0.601	-0.283				
4	1	0.735	-0.318				
5	1	0.699	-0.660				
6	1	0.376	0.468				
7	2	0.971	0.079	0.896	-0.896		
8	2	1.035	-0.068	0.824	-0.824		
9	1	0.583	0.103				
10	1	0.640	-0.124				
11	1	0.441	0.481				
12	1	0.585	-0.187				
13	1	0.893	-0.256				
14	1	0.460	-0.071				
15	1	0.618	0.466				
16	1	0.818	-0.598				
17	1	0.704	-0.521				
18	1	0.604	-0.233				
19	1	0.540	-0.511				
20	1	0.713	-1.293				
21	1	0.877	-0.724				
22	1	0.752	-1.567				
23	1	0.617	-0.575				
24	1	0.508	-0.109				
25	1	0.991	-0.793				
26	2	0.857	-0.568	0.956	-0.956		
27	2	0.858	-0.033	0.838	-0.838		
28	2	0.743	-0.006	0.730	-0.730		

Table N2. ELA Grade 4 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.792	-1.010				
2	1	0.965	-1.078				
3	1	0.535	-0.359				
4	1	0.664	-1.902				
5	1	0.368	-0.679				
6	1	0.709	-1.010				
7	1	0.770	-0.869				

Item	MaxPts	a	b	d1	d2	d3	d4
8	1	0.392	0.162				
9	2	0.851	-0.386	0.965	-0.965		
10	2	1.071	-0.689	0.663	-0.663		
11	2	0.849	-0.505	0.775	-0.775		
12	4	0.774	0.271	1.362	0.683	-0.517	-1.527
13	1	0.723	-1.725				
14	1	0.567	-0.577				
15	1	0.488	-0.495				
16	1	0.537	-1.132				
17	1	0.559	-1.173				
18	1	0.622	-1.034				
19	1	0.802	-1.135				
20	1	0.542	0.336				
21	1	0.446	0.301				
22	1	0.835	-1.103				
23	2	0.886	-0.522	0.855	-0.855		
24	2	1.006	-0.269	0.760	-0.760		
25	1	0.533	0.082				
26	1	0.489	-0.130				
27	1	0.520	0.211				
28	1	0.692	-0.764				
29	1	0.728	-0.105				

Table N3. ELA Grade 5 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.434	-0.865				
2	1	0.347	1.104				
3	1	0.946	-1.305				
4	1	0.388	0.677				
5	1	0.548	-0.031				
6	1	0.706	0.275				
7	1	0.774	-0.455				
8	1	0.596	-0.539				
9	1	0.424	-0.806				
10	1	0.378	-1.268				
11	1	0.599	-0.445				
12	1	0.584	-0.792				
13	1	0.465	-0.403				
14	1	0.743	-0.921				
15	2	0.902	-0.615	0.773	-0.773		
16	2	0.707	-0.529	0.750	-0.750		
17	1	0.215	0.838				

Item	MaxPts	a	b	d1	d2	d3	d4
18	1	0.534	-0.558				
19	1	0.399	-0.081				
20	1	0.183	2.239				
21	1	0.327	0.986				
22	2	0.882	-0.368	0.811	-0.811		
23	2	0.848	-0.628	0.687	-0.687		
24	2	0.755	-0.631	0.708	-0.708		
25	1	0.469	-0.226				
26	1	0.336	-0.130				
27	1	0.481	-0.525				
28	1	0.571	-0.260				
29	1	0.439	-0.822				
30	1	0.594	-0.831				
31	1	0.262	0.620				
32	4	0.713	0.316	1.530	0.770	-0.613	-1.687

Table N4. ELA Grade 6 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	2	0.855	-0.620	0.773	-0.773		
2	2	1.148	-0.839	0.639	-0.639		
3	1	0.742	-0.936				
4	1	0.396	-0.272				
5	1	0.588	-1.108				
6	1	0.313	0.664				
7	1	0.563	-0.873				
8	1	0.619	-1.198				
9	1	0.547	-1.412				
10	1	0.554	0.365				
11	1	0.646	-0.575				
12	1	0.416	-0.194				
13	1	0.346	0.490				
14	1	0.262	0.915				
15	2	1.030	-0.915	0.721	-0.721		
16	1	0.531	-0.991				
17	1	0.471	-0.565				
18	1	0.601	0.223				
19	1	0.466	-0.624				
20	1	0.648	-1.115				
21	1	0.705	-0.327				
22	1	0.550	-0.374				
23	4	0.838	0.306	1.550	0.720	-0.588	-1.682
24	1	0.583	-0.449				

Item	MaxPts	a	b	d1	d2	d3	d4
25	1	0.629	-0.352				
26	1	0.821	-1.410				
27	1	0.333	1.027				
28	1	0.843	-1.124				
29	1	0.311	0.981				
30	1	0.359	0.706				
31	2	1.119	-1.111	0.658	-0.658		
32	2	1.082	-1.068	0.856	-0.856		

Table N5. ELA Grade 7 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	2	1.011	-1.406	0.677	-0.677		
2	2	1.025	-1.051	0.667	-0.667		
3	2	1.197	-1.050	0.583	-0.583		
4	1	0.598	0.185				
5	1	0.813	-0.776				
6	1	0.226	-1.012				
7	1	0.292	-0.322				
8	1	0.187	2.369				
9	1	0.456	-0.377				
10	1	0.841	-1.497				
11	1	0.540	-0.716				
12	1	0.354	-1.776				
13	1	0.815	-0.968				
14	1	0.541	-1.239				
15	1	0.396	-0.646				
16	4	0.809	-0.245	1.256	0.951	-0.625	-1.583
17	1	0.528	-1.150				
18	1	0.718	-1.075				
19	1	0.595	0.186				
20	1	0.435	-0.731				
21	1	1.004	-1.970				
22	1	0.350	-1.074				
23	1	0.545	-1.091				
24	1	0.598	-0.437				
25	1	0.572	-0.869				
26	1	0.498	-0.017				
27	1	0.576	-0.127				
28	1	0.384	-0.587				
29	1	0.298	-0.182				
30	1	0.430	1.300				
31	2	0.966	-0.706	0.653	-0.653		

Item	MaxPts	a	b	d1	d2	d3	d4
32	2	0.824	-0.778	0.708	-0.708		
33	1	0.941	-1.177				
34	1	0.424	-0.847				
35	1	0.986	-1.081				
36	1	0.839	-0.746				
37	1	0.498	0.208				
38	1	0.447	0.084				
39	1	0.406	-0.613				

Table N6. ELA Grade 8 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.918	-1.474				
2	1	0.465	-0.380				
3	1	0.841	-0.728				
4	1	0.755	-0.222				
5	1	0.572	0.036				
6	1	0.580	-0.624				
7	1	0.917	-1.264				
8	1	0.469	0.057				
9	1	0.557	-1.269				
10	1	0.654	-0.872				
11	1	1.054	-0.733				
12	1	0.921	-0.797				
13	1	0.747	-1.001				
14	1	1.138	-1.198				
15	1	0.368	-1.743				
16	1	0.752	-1.261				
17	1	0.797	-0.855				
18	1	0.491	-0.019				
19	1	0.555	-1.092				
20	1	0.538	-1.173				
21	1	0.458	0.139				
22	2	1.007	-1.111	0.625	-0.625		
23	2	0.938	-1.032	0.586	-0.586		
24	1	0.646	-1.079				
25	1	0.915	-1.059				
26	1	0.590	-0.255				
27	1	0.747	-0.653				
28	1	0.968	-1.129				
29	1	0.603	-0.753				
30	1	1.327	-1.107				
31	1	0.806	-0.318				

Item	MaxPts	a	b	d1	d2	d3	d4
32	1	0.798	-0.962				
33	1	0.410	-0.696				
34	1	0.465	-0.275				
35	1	0.716	-0.466				
36	2	1.016	-1.035	0.483	-0.483		
37	2	0.852	-0.969	0.844	-0.844		
38	2	0.721	-0.844	0.857	-0.857		
39	4	0.740	0.143	1.542	0.811	-0.649	-1.704

Table N7. Mathematics Grade 3 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.406	-1.213				
2	1	0.974	-1.039				
3	1	1.027	-0.316				
4	1	1.230	-0.571				
5	1	0.989	-0.841				
6	1	1.049	-1.307				
7	1	0.852	-0.821				
8	1	0.773	-1.031				
9	1	0.984	-0.912				
10	2	0.931	0.279	0.355	-0.355		
11	1	0.982	-0.487				
12	1	0.589	0.355				
13	1	0.800	-1.781				
14	1	0.774	-1.129				
15	1	0.652	-0.109				
16	1	0.418	0.544				
17	1	0.798	-0.650				
18	1	0.579	0.600				
19	1	0.809	-1.344				
20	1	1.401	-0.920				
21	1	1.101	-0.938				
22	1	0.874	-1.594				
23	1	0.694	-0.946				
24	1	0.854	-0.296				
25	1	1.009	-0.283				
26	1	0.620	0.782				
27	1	0.464	-0.714				
28	1	0.539	-1.696				
29	2	0.965	0.085	-0.161	0.161		
30	3	0.540	1.181	0.573	-0.376	-0.198	

Item	MaxPts	a	b	d1	d2	d3	d4
31	2	0.830	-0.071	-0.267	0.267		
32	2	0.760	0.377	-0.360	0.360		

Table N8. Mathematics Grade 4 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.921	-0.162				
2	1	1.098	-0.711				
3	2	0.719	-0.124	0.553	-0.553		
4	1	1.017	-0.315				
5	1	0.107	-1.748				
6	1	1.262	-0.970				
7	1	1.644	-1.599				
8	1	1.077	-1.254				
9	1	0.782	-0.570				
10	1	1.045	-0.722				
11	2	0.699	0.269	-0.148	0.148		
12	1	0.830	-0.691				
13	1	1.092	-1.648				
14	1	0.719	-0.871				
15	1	1.186	-0.827				
16	1	0.762	0.144				
17	1	0.514	-0.537				
18	1	0.545	-0.966				
19	1	1.080	-0.458				
20	1	0.867	-0.074				
21	1	0.721	-0.451				
22	1	0.719	-0.353				
23	1	0.482	-0.844				
24	1	0.743	-0.726				
25	1	0.873	0.146				
26	1	0.813	-1.025				
27	1	0.645	-1.065				
28	1	1.042	-1.059				
29	1	1.108	-0.833				
30	1	0.863	-0.886				
31	1	0.663	-1.174				
32	1	0.596	0.922				
33	1	0.649	-0.394				
34	2	0.886	0.355	0.097	-0.097		
35	2	0.617	0.386	0.362	-0.362		
36	2	0.774	0.139	0.234	-0.234		
37	3	0.935	0.459	0.390	0.264	-0.654	

Table N9. Mathematics Grade 5 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.896	-0.904				
2	1	0.745	0.164				
3	2	0.863	-0.116	0.426	-0.426		
4	1	0.775	-0.203				
5	2	0.763	0.992	0.522	-0.522		
6	1	0.471	0.079				
7	1	0.984	-0.288				
8	1	0.757	-0.906				
9	1	0.823	-1.095				
10	1	0.869	-1.454				
11	1	0.472	2.729				
12	1	0.904	-0.832				
13	2	0.961	-0.005	-0.204	0.204		
14	1	1.433	-0.716				
15	1	0.570	-0.794				
16	1	1.200	-0.450				
17	1	0.954	-0.501				
18	1	0.717	-0.625				
19	1	0.884	-1.873				
20	1	0.820	0.270				
21	1	0.983	-0.430				
22	1	0.734	0.565				
23	1	0.786	0.139				
24	1	1.375	-1.085				
25	1	0.666	-1.017				
26	1	0.354	0.070				
27	1	0.583	0.401				
28	1	0.534	0.606				
29	1	0.694	0.463				
30	1	1.262	-0.821				
31	1	0.476	0.276				
32	1	0.610	1.161				
33	1	0.709	-0.527				
34	1	0.748	-0.059				
35	2	0.571	1.311	-0.061	0.061		
36	2	0.500	1.377	0.347	-0.347		
37	3	0.781	1.196	-0.152	0.17	-0.018	

Table N10. Mathematics Grade 6 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.990	-0.324				
2	1	1.302	-0.300				
3	1	0.796	-0.370				
4	1	1.257	-1.045				
5	2	0.654	0.465	-0.313	0.313		
6	1	0.518	-0.266				
7	1	1.149	0.423				
8	1	1.285	-0.747				
9	1	1.446	-1.473				
10	1	0.904	-0.311				
11	1	0.854	1.285				
12	1	1.223	-0.771				
13	1	0.637	-0.230				
14	1	1.030	-0.184				
15	1	0.495	0.861				
16	1	1.124	-0.117				
17	2	0.667	0.960	0.412	-0.412		
18	2	0.580	0.269	-0.401	0.401		
19	1	0.872	-2.020				
20	1	0.780	-0.034				
21	2	0.967	0.504	-0.134	0.134		
22	2	0.832	0.157	0.339	-0.339		
23	1	1.281	-0.171				
24	1	0.631	0.264				
25	1	0.488	-0.178				
26	1	0.557	-0.470				
27	2	0.635	0.885	0.264	-0.264		
28	1	0.462	1.361				
29	1	0.510	-0.191				
30	1	1.084	-0.736				
31	1	0.461	0.436				
32	1	0.401	0.816				
33	1	0.437	0.149				
34	1	0.849	-0.862				
35	1	0.407	-0.033				
36	1	0.802	-0.526				
37	1	0.992	-0.182				
38	1	0.940	-0.928				
39	3	0.639	0.324	-0.507	-0.210	0.717	

Table N11. Mathematics Grade 7 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.661	-1.610				
2	1	0.903	-1.004				
3	1	1.093	-1.192				
4	1	0.564	-0.412				
5	1	1.064	0.664				
6	1	0.583	-0.119				
7	1	1.124	0.547				
8	1	0.860	-0.051				
9	1	0.780	-0.332				
10	1	0.575	-0.501				
11	1	0.655	-0.158				
12	1	0.795	-1.654				
13	1	0.692	-0.421				
14	2	0.787	0.400	0.042	-0.042		
15	1	0.583	-0.647				
16	1	0.588	0.012				
17	1	0.717	-0.826				
18	1	0.605	0.142				
19	1	0.501	-0.289				
20	1	0.628	-0.832				
21	1	1.216	-0.664				
22	1	0.538	-0.958				
23	1	0.607	-0.965				
24	1	0.432	-1.260				
25	1	0.681	-0.333				
26	1	0.588	-0.130				
27	1	0.692	-0.230				
28	1	0.225	2.569				
29	1	0.774	-1.246				
30	1	0.472	0.666				
31	1	0.937	0.205				
32	1	0.952	-1.766				
33	1	0.640	0.945				
34	1	0.846	0.342				
35	1	0.733	-0.669				
36	2	0.721	-0.408	-0.394	0.394		
37	2	0.669	-0.389	-0.900	0.900		
38	2	0.758	0.440	-0.309	0.309		
39	2	1.116	-0.570	-0.186	0.186		
40	2	0.760	-0.767	-0.943	0.943		
41	3	0.964	0.962	0.541	0.175	-0.715	

Table N12. Mathematics Grade 8 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.864	1.156				
2	2	0.859	0.737	-0.484	0.484		
3	1	0.412	0.727				
4	1	0.733	-0.831				
5	1	0.920	-0.259				
6	1	0.651	-0.046				
7	1	0.969	0.493				
8	1	0.447	0.195				
9	1	0.651	-0.697				
10	2	0.636	2.234	0.430	-0.430		
11	2	0.821	1.073	0.268	-0.268		
12	1	0.590	-0.046				
13	1	0.708	-0.315				
14	2	0.789	0.518	-0.107	0.107		
15	1	0.893	0.054				
16	1	0.680	1.580				
17	1	0.763	0.112				
18	1	0.688	-0.680				
19	1	0.785	0.204				
20	3	0.937	0.087	0.206	0.035	-0.242	
21	1	0.587	-0.454				
22	1	0.655	-0.149				
23	1	1.478	-1.007				
24	1	0.922	-0.530				
25	1	0.803	-0.679				
26	1	0.415	0.735				
27	1	0.890	-1.026				
28	1	0.849	0.061				
29	1	0.581	-0.256				
30	1	0.564	0.136				
31	1	0.797	-0.066				
32	1	0.644	-0.090				
33	1	0.635	0.884				
34	1	0.776	-0.774				
35	1	0.479	-0.569				
36	1	0.452	-0.407				
37	1	0.306	0.206				
38	1	0.713	-0.944				
39	1	0.456	0.118				
40	2	0.822	0.341	-0.116	0.116		
41	2	0.599	1.143	-0.725	0.725		

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

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

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, $C_1, C_2, \ldots, C_{H-1}$. 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:

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

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

Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each operational (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 for the two administrations is

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

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

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

The agreement index (classification consistency) can be computed as

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

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

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

Then, kappa (Cohen, 1960) is

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

Classification Accuracy

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

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

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

Lee (2010) 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 3 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-Score-to-Scale-Score and Scale-Score Frequency Tables

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

Table P1. ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	383	14
1	386	13
2	390	12
3	393	11
4	399	9
5	404	9
6	408	8
7	412	7
8	415	7
9	418	7
10	421	7
11	423	7
12	426	6
13	428	6
14	432	6
15	433	6
16	436	6
17	438	6
18	441	6
19	443	6
20	446	7
21	450	7
22	451	7
23	454	7
24	457	7
25	460	7
26	463	8
27	467	8
28	474	9
29	477	10
30	483	11
31	487	12
32	490	13
33	494	15

Table P2. ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	371	15
1	375	14
2	378	13
3	382	12
4	388	10
5	394	9
6	398	8
7	402	8
8	405	7
9	408	7
10	411	7
11	414	7
12	416	7
13	419	6
14	421	6
15	424	6
16	426	6
17	428	6
18	431	6
19	433	6
20	436	6
21	438	6
22	440	7
23	443	7
24	446	7
25	450	7
26	451	7
27	454	7
28	457	8
29	461	8
30	465	8
31	471	9
32	474	10
33	479	11
34	483	12
35	486	12
36	490	13
37	493	14

Table P3. ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	373	16
1	376	15
2	380	13
3	383	13
4	387	12
5	390	11
6	395	10
7	400	9
8	404	8
9	407	8
10	410	8
11	413	7
12	416	7
13	419	7
14	421	7
15	424	7
16	426	7
17	429	7
18	432	7
19	434	7
20	436	7
21	439	7
22	441	7
23	444	7
24	446	7
25	450	7
26	452	8
27	455	8
28	458	8
29	461	8
30	465	9
31	468	9
32	474	10
33	477	10
34	483	11
35	486	12
36	490	13
37	493	14
38	497	15
39	500	16
40	504	17

Table P4. ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	368	16
1	371	14
2	375	13
3	378	12
4	385	10
5	390	9
6	394	8
7	398	8
8	401	7
9	404	7
10	407	7
11	410	6
12	412	6
13	415	6
14	417	6
15	420	6
16	422	6
17	424	6
18	426	6
19	429	6
20	431	6
21	433	6
22	435	6
23	438	6
24	440	6
25	443	7
26	445	7
27	450	7
28	451	7
29	454	8
30	457	8
31	461	8
32	465	9
33	470	10
34	475	11
35	481	12
36	484	12
37	488	13
38	491	14
39	495	15
40	498	16

Table P5. ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	363	17
1	366	16
2	370	14
3	373	13
4	377	12
5	383	10
6	388	9
7	392	8
8	396	8
9	399	7
10	402	7
11	404	7
12	407	6
13	409	6
14	412	6
15	414	6
16	416	6
17	418	6
18	420	6
19	422	6
20	424	6
21	426	6
22	428	6
23	430	6
24	433	6
25	434	6
26	436	6
27	438	6
28	440	6
29	442	6
30	444	6
31	447	7
32	450	7
33	452	7
34	454	7
35	457	8
36	460	8
37	464	8
38	467	9
39	472	10
40	476	10

Raw Score	Scale Score	Standard Error
41	481	11
42	485	12
43	488	13
44	492	14
45	495	15
46	499	16
47	502	17

Table P6. ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	363	16
1	367	15
2	371	14
3	374	12
4	381	11
5	386	9
6	391	9
7	394	8
8	398	7
9	400	7
10	403	7
11	406	6
12	408	6
13	410	6
14	412	6
15	414	6
16	416	6
17	418	6
18	420	6
19	422	6
20	424	6
21	425	6
22	427	6
23	430	6
24	431	6
25	432	6
26	434	6
27	436	6
28	438	6
29	440	6
30	442	6
31	444	6

Raw Score	Scale Score	Standard Error
32	446	6
33	450	7
34	451	7
35	454	7
36	456	7
37	459	8
38	462	8
39	466	9
40	472	9
41	474	10
42	480	11
43	486	12
44	490	13
45	493	14
46	497	15
47	500	16

Table P7. Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	381	14
1	384	13
2	388	13
3	391	12
4	394	11
5	400	10
6	405	9
7	409	8
8	413	8
9	416	7
10	419	7
11	424	7
12	425	7
13	428	7
14	430	7
15	433	7
16	435	6
17	437	6
18	440	6
19	442	6
20	445	6
21	447	6
22	450	7

Raw Score	Scale Score	Standard Error
23	452	7
24	454	7
25	457	7
26	459	7
27	462	7
28	465	8
29	468	8
30	472	8
31	475	9
32	480	10
33	487	11
34	491	12
35	495	13
36	498	14
37	501	15
38	505	16

Table P8. Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	376	16
1	380	15
2	383	13
3	386	13
4	389	12
5	396	10
6	401	9
7	405	9
8	409	8
9	412	8
10	415	8
11	418	7
12	421	7
13	424	7
14	427	7
15	431	7
16	432	7
17	434	7
18	436	7
19	439	7
20	441	7
21	443	7
22	446	7

Raw Score	Scale Score	Standard Error
23	448	7
24	450	7
25	453	7
26	455	7
27	457	7
28	460	7
29	462	7
30	465	7
31	467	7
32	470	8
33	473	8
34	476	8
35	479	9
36	483	9
37	486	10
38	491	10
39	496	11
40	499	12
41	503	13
42	506	14
43	509	14
44	512	15

Table P9. Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	384	14
1	387	13
2	390	13
3	394	12
4	397	11
5	403	10
6	407	9
7	411	8
8	415	8
9	419	8
10	422	7
11	425	7
12	427	7
13	430	7
14	432	7
15	435	7
16	437	7

Raw Score	Scale Score	Standard Error
17	440	7
18	442	7
19	444	7
20	447	7
21	450	7
22	452	7
23	454	7
24	456	7
25	459	7
26	461	7
27	464	7
28	467	8
29	470	8
30	472	8
31	475	8
32	478	8
33	483	8
34	485	9
35	488	9
36	492	10
37	496	10
38	501	11
39	505	12
40	508	12
41	511	13
42	514	14
43	518	15
44	521	16

Table P10. Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	387	14
1	390	13
2	393	12
3	396	11
4	399	10
5	405	9
6	409	8
7	413	8
8	417	8
9	420	7
10	423	7

Raw Score	Scale Score	Standard Error
11	426	7
12	428	7
13	431	6
14	433	6
15	435	6
16	437	6
17	439	6
18	441	6
19	443	6
20	445	6
21	447	6
22	450	6
23	451	6
24	453	6
25	455	6
26	457	6
27	459	6
28	460	6
29	462	6
30	464	6
31	466	6
32	468	6
33	471	7
34	473	7
35	475	7
36	478	7
37	481	8
38	485	8
39	487	9
40	491	9
41	495	10
42	501	12
43	504	12
44	507	13
45	510	14
46	514	15
47	517	16

Table P11. Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	378	16
1	381	15
2	385	14
3	388	13
4	391	13
5	395	12
6	400	11
7	405	10
8	409	9
9	413	9
10	416	9
11	419	8
12	422	8
13	425	8
14	428	7
15	430	7
16	432	7
17	435	7
18	437	7
19	439	7
20	441	7
21	443	7
22	445	7
23	447	7
24	450	7
25	451	7
26	453	7
27	455	7
28	458	7
29	460	7
30	462	7
31	464	7
32	467	7
33	469	7
34	471	7
35	474	7
36	477	8
37	479	8
38	482	8
39	485	8
40	489	9

Raw Score	Scale Score	Standard Error
41	492	9
42	496	10
43	501	11
44	506	12
45	510	12
46	513	13
47	516	14
48	520	15
49	523	16

Table P12. Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	382	15
1	385	14
2	388	13
3	391	13
4	394	12
5	398	11
6	403	10
7	407	9
8	411	9
9	415	8
10	418	8
11	421	8
12	424	8
13	427	7
14	430	7
15	432	7
16	436	7
17	437	7
18	439	7
19	441	7
20	444	7
21	446	7
22	448	7
23	450	6
24	452	6
25	454	6
26	456	6
27	458	6
28	460	7
29	462	7

Raw Score	Scale Score	Standard Error
30	464	7
31	466	7
32	468	7
33	470	7
34	473	7
35	475	7
36	478	8
37	482	8
38	483	8
39	486	8
40	490	9
41	493	9
42	497	10
43	502	11
44	505	12
45	509	12
46	512	13
47	515	14
48	518	14
49	521	15

Table P13. ELA Grade 3 Scale Score Frequency Distribution

Scale			Cumulative	
Score	Freq.	%	Freq.	%
383	34	0.02	34	0.0
386	96	0.06	130	0.1
390	340	0.21	470	0.3
393	894	0.54	1,364	0.8
399	1,739	1.06	3,103	1.9
404	2,660	1.62	5,763	3.5
408	3,736	2.27	9,499	5.8
412	4,433	2.70	13,932	8.5
415	5,255	3.20	19,187	11.7
418	5,493	3.34	24,680	15.0
421	5,780	3.51	30,460	18.5
423	6,000	3.65	36,460	22.2
426	6,307	3.83	42,767	26.0
428	6,378	3.88	49,145	29.9
432	6,491	3.95	55,636	33.8
433	6,480	3.94	62,116	37.8
436	6,349	3.86	68,465	41.6
438	6,442	3.92	74,907	45.5

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
441	6,755	4.11	81,662	49.7
443	6,608	4.02	88,270	53.7
446	6,625	4.03	94,895	57.7
450	6,790	4.13	101,685	61.8
451	6,943	4.22	108,628	66.1
454	6,941	4.22	115,569	70.3
457	6,889	4.19	122,458	74.5
460	6,936	4.22	129,394	78.7
463	6,796	4.13	136,190	82.8
467	6,383	3.88	142,573	86.7
474	6,138	3.73	148,711	90.4
477	5,390	3.28	154,101	93.7
483	4,540	2.76	158,641	96.5
487	3,234	1.97	161,875	98.4
490	1,887	1.15	163,762	99.6
494	699	0.43	164,461	100.0

Table P14. ELA Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
371	6	0.00	6	0.0
375	28	0.02	34	0.0
378	127	0.08	161	0.1
382	357	0.21	518	0.3
388	742	0.44	1,260	0.7
394	1,272	0.75	2,532	1.5
398	1,749	1.04	4,281	2.5
402	2,190	1.30	6,471	3.8
405	2,639	1.56	9,110	5.4
408	3,049	1.81	12,159	7.2
411	3,423	2.03	15,582	9.2
414	3,641	2.16	19,223	11.4
416	3,887	2.30	23,110	13.7
419	4,030	2.39	27,140	16.1
421	4,402	2.61	31,542	18.7
424	4,553	2.70	36,095	21.4
426	4,707	2.79	40,802	24.2
428	5,095	3.02	45,897	27.2
431	5,287	3.13	51,184	30.3
433	5,468	3.24	56,652	33.6
436	5,907	3.50	62,559	37.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
438	6,282	3.72	68,841	40.8
440	6,703	3.97	75,544	44.8
443	7,071	4.19	82,615	48.9
446	7,390	4.38	90,005	53.3
450	7,617	4.51	97,622	57.8
451	7,921	4.69	105,543	62.5
454	8,107	4.80	113,650	67.3
457	8,491	5.03	122,141	72.4
461	8,339	4.94	130,480	77.3
465	8,286	4.91	138,766	82.2
471	8,050	4.77	146,816	87.0
474	7,032	4.17	153,848	91.1
479	6,031	3.57	159,879	94.7
483	4,422	2.62	164,301	97.3
486	2,761	1.64	167,062	99.0
490	1,358	0.80	168,420	99.8
493	391	0.23	168,811	100.0

Table P15. ELA Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
373	2	0.00	2	0.0
376	16	0.01	18	0.0
380	56	0.04	74	0.0
383	187	0.12	261	0.2
387	444	0.28	705	0.4
390	842	0.53	1,547	1.0
395	1,480	0.93	3,027	1.9
400	2,002	1.26	5,029	3.2
404	2,453	1.55	7,482	4.7
407	3,066	1.93	10,548	6.6
410	3,413	2.15	13,961	8.8
413	3,569	2.25	17,530	11.0
416	3,882	2.45	21,412	13.5
419	4,255	2.68	25,667	16.2
421	4,446	2.80	30,113	19.0
424	4,607	2.90	34,720	21.9
426	4,822	3.04	39,542	24.9
429	5,135	3.23	44,677	28.1
432	5,409	3.41	50,086	31.5
434	5,678	3.58	55,764	35.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
436	5,876	3.70	61,640	38.8
439	6,312	3.98	67,952	42.8
441	6,574	4.14	74,526	46.9
444	6,675	4.20	81,201	51.1
446	6,959	4.38	88,160	55.5
450	7,156	4.51	95,316	60.0
452	7,421	4.67	102,737	64.7
455	7,360	4.64	110,097	69.3
458	7,246	4.56	117,343	73.9
461	7,105	4.48	124,448	78.4
465	6,703	4.22	131,151	82.6
468	6,447	4.06	137,598	86.7
474	5,901	3.72	143,499	90.4
477	4,987	3.14	148,486	93.5
483	4,051	2.55	152,537	96.1
486	2,881	1.81	155,418	97.9
490	1,921	1.21	157,339	99.1
493	920	0.58	158,259	99.7
497	393	0.25	158,652	99.9
500	98	0.06	158,750	100.0
504	15	0.01	158,765	100.0

Table P16. ELA Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
368	2	0.00	2	0.0
371	12	0.01	14	0.0
375	62	0.04	76	0.0
378	142	0.09	218	0.1
385	395	0.24	613	0.4
390	653	0.40	1,266	0.8
394	1,021	0.62	2,287	1.4
398	1,478	0.90	3,765	2.3
401	1,893	1.15	5,658	3.4
404	2,199	1.33	7,857	4.8
407	2,658	1.61	10,515	6.4
410	3,100	1.88	13,615	8.3
412	3,433	2.08	17,048	10.3
415	3,648	2.21	20,696	12.5
417	3,855	2.34	24,551	14.9
420	4,075	2.47	28,626	17.4

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
422	4,364	2.65	32,990	20.0
424	4,703	2.85	37,693	22.9
426	4,829	2.93	42,522	25.8
429	5,126	3.11	47,648	28.9
431	5,490	3.33	53,138	32.2
433	5,813	3.52	58,951	35.7
435	5,949	3.61	64,900	39.4
438	6,329	3.84	71,229	43.2
440	6,760	4.10	77,989	47.3
443	7,194	4.36	85,183	51.7
445	7,322	4.44	92,505	56.1
450	7,620	4.62	100,125	60.7
451	7,804	4.73	107,929	65.4
454	7,972	4.83	115,901	70.3
457	8,021	4.86	123,922	75.1
461	8,045	4.88	131,967	80.0
465	7,467	4.53	139,434	84.6
470	6,925	4.20	146,359	88.7
475	6,237	3.78	152,596	92.5
481	4,975	3.02	157,571	95.5
484	3,496	2.12	161,067	97.7
488	2,243	1.36	163,310	99.0
491	1,121	0.68	164,431	99.7
495	411	0.25	164,842	100.0
498	71	0.04	164,913	100.0

Table P17. ELA Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
363	2	0.00	2	0.0
366	8	0.01	10	0.0
370	13	0.01	23	0.0
373	39	0.02	62	0.0
377	116	0.07	178	0.1
383	227	0.14	405	0.3
388	347	0.22	752	0.5
392	577	0.36	1,329	0.8
396	796	0.50	2,125	1.3
399	1,012	0.63	3,137	2.0
402	1,208	0.76	4,345	2.7
404	1,449	0.91	5,794	3.6

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
407	1,656	1.04	7,450	4.7
409	1,872	1.17	9,322	5.8
412	2,092	1.31	11,414	7.1
414	2,317	1.45	13,731	8.6
416	2,344	1.47	16,075	10.1
418	2,558	1.60	18,633	11.7
420	2,718	1.70	21,351	13.4
422	2,901	1.81	24,252	15.2
424	3,082	1.93	27,334	17.1
426	3,260	2.04	30,594	19.1
428	3,601	2.25	34,195	21.4
430	3,673	2.30	37,868	23.7
433	4,073	2.55	41,941	26.2
434	4,492	2.81	46,433	29.0
436	4,683	2.93	51,116	32.0
438	4,961	3.10	56,077	35.1
440	5,488	3.43	61,565	38.5
442	5,757	3.60	67,322	42.1
444	6,083	3.80	73,405	45.9
447	6,331	3.96	79,736	49.9
450	6,580	4.12	86,316	54.0
452	6,953	4.35	93,269	58.3
454	7,530	4.71	100,799	63.0
457	7,572	4.74	108,371	67.8
460	7,577	4.74	115,948	72.5
464	7,511	4.70	123,459	77.2
467	7,321	4.58	130,780	81.8
472	6,883	4.30	137,663	86.1
476	6,414	4.01	144,077	90.1
481	5,301	3.32	149,378	93.4
485	4,283	2.68	153,661	96.1
488	3,046	1.90	156,707	98.0
492	1,845	1.15	158,552	99.2
495	933	0.58	159,485	99.7
499	343	0.21	159,828	100.0
502	74	0.05	159,902	100.0

Table P18. ELA Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
363	4	0.00	4	0.0
367	3	0.00	7	0.0
371	12	0.01	19	0.0
374	26	0.02	45	0.0
381	84	0.06	129	0.1
386	168	0.11	297	0.2
391	286	0.19	583	0.4
394	521	0.35	1,104	0.7
398	790	0.54	1,894	1.3
400	1,132	0.77	3,026	2.1
403	1,368	0.93	4,394	3.0
406	1,609	1.09	6,003	4.1
408	1,824	1.24	7,827	5.3
410	1,938	1.32	9,765	6.6
412	2,141	1.45	11,906	8.1
414	2,212	1.50	14,118	9.6
416	2,359	1.60	16,477	11.2
418	2,385	1.62	18,862	12.8
420	2,337	1.59	21,199	14.4
422	2,594	1.76	23,793	16.2
424	2,543	1.73	26,336	17.9
425	2,748	1.87	29,084	19.8
427	2,805	1.90	31,889	21.7
430	2,946	2.00	34,835	23.7
431	2,995	2.03	37,830	25.7
432	3,188	2.16	41,018	27.9
434	3,405	2.31	44,423	30.2
436	3,538	2.40	47,961	32.6
438	3,744	2.54	51,705	35.1
440	3,928	2.67	55,633	37.8
442	4,197	2.85	59,830	40.6
444	4,409	2.99	64,239	43.6
446	4,684	3.18	68,923	46.8
450	5,061	3.44	73,984	50.2
451	5,413	3.68	79,397	53.9
454	5,856	3.98	85,253	57.9
456	6,171	4.19	91,424	62.1
459	6,476	4.40	97,900	66.5
462	6,667	4.53	104,567	71.0
466	7,025	4.77	111,592	75.8

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
472	7,092	4.82	118,684	80.6
474	6,900	4.69	125,584	85.3
480	6,473	4.40	132,057	89.7
486	5,745	3.90	137,802	93.6
490	4,487	3.05	142,289	96.6
493	2,983	2.03	145,272	98.7
497	1,506	1.02	146,778	99.7
500	479	0.33	147,257	100.0

Table P19. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
381	10	0.01	10	0.0
384	75	0.04	85	0.1
388	261	0.16	346	0.2
391	630	0.38	976	0.6
394	1,205	0.72	2,181	1.3
400	2,044	1.22	4,225	2.5
405	2,711	1.62	6,936	4.2
409	3,539	2.12	10,475	6.3
413	3,912	2.34	14,387	8.6
416	4,199	2.51	18,586	11.1
419	4,451	2.66	23,037	13.8
424	4,570	2.73	27,607	16.5
425	4,549	2.72	32,156	19.2
428	4,834	2.89	36,990	22.1
430	4,754	2.84	41,744	25.0
433	4,830	2.89	46,574	27.9
435	4,914	2.94	51,488	30.8
437	4,928	2.95	56,416	33.8
440	5,089	3.05	61,505	36.8
442	5,243	3.14	66,748	39.9
445	5,298	3.17	72,046	43.1
447	5,498	3.29	77,544	46.4
450	5,598	3.35	83,142	49.7
452	5,713	3.42	88,855	53.2
454	5,798	3.47	94,653	56.6
457	5,786	3.46	100,439	60.1
459	6,013	3.60	106,452	63.7
462	6,183	3.70	112,635	67.4
465	6,096	3.65	118,731	71.0

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
468	6,260	3.75	124,991	74.8
472	6,279	3.76	131,270	78.5
475	6,346	3.80	137,616	82.3
480	5,944	3.56	143,560	85.9
487	5,913	3.54	149,473	89.4
491	5,428	3.25	154,901	92.7
495	4,692	2.81	159,593	95.5
498	3,741	2.24	163,334	97.7
501	2,529	1.51	165,863	99.2
505	1,260	0.75	167,123	100.0

Table P20. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
376	10	0.01	10	0.0
380	21	0.01	31	0.0
383	88	0.05	119	0.1
386	264	0.16	383	0.2
389	631	0.38	1,014	0.6
396	1,177	0.70	2,191	1.3
401	1,899	1.13	4,090	2.4
405	2,512	1.50	6,602	3.9
409	3,091	1.84	9,693	5.8
412	3,539	2.11	13,232	7.9
415	3,782	2.25	17,014	10.1
418	3,888	2.31	20,902	12.4
421	3,993	2.38	24,895	14.8
424	4,004	2.38	28,899	17.2
427	3,934	2.34	32,833	19.5
431	4,088	2.43	36,921	22.0
432	4,162	2.48	41,083	24.5
434	4,141	2.46	45,224	26.9
436	4,103	2.44	49,327	29.4
439	4,112	2.45	53,439	31.8
441	4,232	2.52	57,671	34.3
443	4,407	2.62	62,078	37.0
446	4,294	2.56	66,372	39.5
448	4,315	2.57	70,687	42.1
450	4,470	2.66	75,157	44.7
453	4,481	2.67	79,638	47.4
455	4,599	2.74	84,237	50.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
457	4,608	2.74	88,845	52.9
460	4,677	2.78	93,522	55.7
462	4,820	2.87	98,342	58.5
465	4,953	2.95	103,295	61.5
467	4,959	2.95	108,254	64.4
470	5,100	3.04	113,354	67.5
473	5,199	3.09	118,553	70.6
476	5,420	3.23	123,973	73.8
479	5,418	3.23	129,391	77.0
483	5,587	3.33	134,978	80.3
486	5,435	3.24	140,413	83.6
491	5,463	3.25	145,876	86.8
496	5,201	3.10	151,077	89.9
499	4,883	2.91	155,960	92.8
503	4,390	2.61	160,350	95.4
506	3,680	2.19	164,030	97.6
509	2,614	1.56	166,644	99.2
512	1,351	0.80	167,995	100.0

Table P21. Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
384	3	0.00	3	0.0
387	20	0.01	23	0.0
390	107	0.07	130	0.1
394	344	0.22	474	0.3
397	897	0.58	1,371	0.9
403	1,673	1.07	3,044	2.0
407	2,786	1.79	5,830	3.7
411	3,782	2.43	9,612	6.2
415	4,616	2.96	14,228	9.1
419	5,167	3.32	19,395	12.4
422	5,304	3.40	24,699	15.9
425	5,409	3.47	30,108	19.3
427	5,251	3.37	35,359	22.7
430	5,316	3.41	40,675	26.1
432	5,321	3.42	45,996	29.5
435	5,359	3.44	51,355	33.0
437	5,244	3.37	56,599	36.3
440	5,194	3.33	61,793	39.7
442	4,994	3.21	66,787	42.9

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
444	5,052	3.24	71,839	46.1
447	4,944	3.17	76,783	49.3
450	4,958	3.18	81,741	52.5
452	4,774	3.06	86,515	55.5
454	4,696	3.01	91,211	58.5
456	4,578	2.94	95,789	61.5
459	4,495	2.89	100,284	64.4
461	4,364	2.80	104,648	67.2
464	4,313	2.77	108,961	69.9
467	4,197	2.69	113,158	72.6
470	4,125	2.65	117,283	75.3
472	4,062	2.61	121,345	77.9
475	3,882	2.49	125,227	80.4
478	3,779	2.43	129,006	82.8
483	3,615	2.32	132,621	85.1
485	3,488	2.24	136,109	87.4
488	3,346	2.15	139,455	89.5
492	3,068	1.97	142,523	91.5
496	2,808	1.80	145,331	93.3
501	2,638	1.69	147,969	95.0
505	2,294	1.47	150,263	96.5
508	1,882	1.21	152,145	97.7
511	1,511	0.97	153,656	98.6
514	1,144	0.73	154,800	99.4
518	676	0.43	155,476	99.8
521	311	0.20	155,787	100.0

Table P22. Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
387	6	0.00	6	0.0
390	32	0.02	38	0.0
393	112	0.07	150	0.1
396	424	0.26	574	0.3
399	962	0.58	1,536	0.9
405	1,923	1.17	3,459	2.1
409	3,235	1.97	6,694	4.1
413	4,335	2.63	11,029	6.7
417	5,301	3.22	16,330	9.9
420	5,843	3.55	22,173	13.5
423	6,002	3.65	28,175	17.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
426	5,837	3.55	34,012	20.7
428	5,528	3.36	39,540	24.0
431	5,177	3.15	44,717	27.2
433	5,227	3.18	49,944	30.3
435	4,911	2.98	54,855	33.3
437	4,664	2.83	59,519	36.2
439	4,421	2.69	63,940	38.8
441	4,414	2.68	68,354	41.5
443	4,271	2.59	72,625	44.1
445	4,182	2.54	76,807	46.7
447	4,121	2.50	80,928	49.2
450	4,139	2.51	85,067	51.7
451	4,021	2.44	89,088	54.1
453	4,040	2.45	93,128	56.6
455	3,918	2.38	97,046	59.0
457	3,752	2.28	100,798	61.2
459	3,777	2.29	104,575	63.5
460	3,754	2.28	108,329	65.8
462	3,898	2.37	112,227	68.2
464	3,648	2.22	115,875	70.4
466	3,687	2.24	119,562	72.6
468	3,524	2.14	123,086	74.8
471	3,629	2.20	126,715	77.0
473	3,610	2.19	130,325	79.2
475	3,457	2.10	133,782	81.3
478	3,481	2.11	137,263	83.4
481	3,403	2.07	140,666	85.5
485	3,363	2.04	144,029	87.5
487	3,330	2.02	147,359	89.5
491	3,172	1.93	150,531	91.5
495	3,080	1.87	153,611	93.3
501	2,832	1.72	156,443	95.0
504	2,497	1.52	158,940	96.6
507	2,164	1.31	161,104	97.9
510	1,713	1.04	162,817	98.9
514	1,217	0.74	164,034	99.7
517	567	0.34	164,601	100.0

Table P23. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
378	4	0.00	4	0.0
381	11	0.01	15	0.0
385	32	0.02	47	0.0
388	88	0.06	135	0.1
391	221	0.14	356	0.2
395	470	0.30	826	0.5
400	948	0.60	1,774	1.1
405	1,452	0.92	3,226	2.0
409	2,058	1.30	5,284	3.3
413	2,769	1.76	8,053	5.1
416	3,208	2.03	11,261	7.1
419	3,820	2.42	15,081	9.6
422	3,950	2.50	19,031	12.1
425	4,229	2.68	23,260	14.7
428	4,342	2.75	27,602	17.5
430	4,376	2.77	31,978	20.3
432	4,343	2.75	36,321	23.0
435	4,395	2.79	40,716	25.8
437	4,390	2.78	45,106	28.6
439	4,389	2.78	49,495	31.4
441	4,441	2.82	53,936	34.2
443	4,337	2.75	58,273	36.9
445	4,326	2.74	62,599	39.7
447	4,370	2.77	66,969	42.5
450	4,236	2.69	71,205	45.1
451	4,310	2.73	75,515	47.9
453	4,122	2.61	79,637	50.5
455	4,195	2.66	83,832	53.1
458	4,129	2.62	87,961	55.8
460	4,123	2.61	92,084	58.4
462	4,099	2.60	96,183	61.0
464	4,145	2.63	100,328	63.6
467	3,964	2.51	104,292	66.1
469	3,967	2.51	108,259	68.6
471	3,922	2.49	112,181	71.1
474	3,758	2.38	115,939	73.5
477	3,805	2.41	119,744	75.9
479	3,711	2.35	123,455	78.3
482	3,635	2.30	127,090	80.6
485	3,484	2.21	130,574	82.8

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
489	3,544	2.25	134,118	85.0
492	3,447	2.19	137,565	87.2
496	3,323	2.11	140,888	89.3
501	3,290	2.09	144,178	91.4
506	3,024	1.92	147,202	93.3
510	2,821	1.79	150,023	95.1
513	2,595	1.64	152,618	96.7
516	2,291	1.45	154,909	98.2
520	1,840	1.17	156,749	99.4
523	1,002	0.64	157,751	100.0

Table P24. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
382	11	0.01	11	0.0
385	30	0.03	41	0.0
388	59	0.06	100	0.1
391	199	0.21	299	0.3
394	424	0.45	723	0.8
398	902	0.96	1,625	1.7
403	1,615	1.72	3,240	3.4
407	2,345	2.49	5,585	5.9
411	2,911	3.09	8,496	9.0
415	3,468	3.69	11,964	12.7
418	3,800	4.04	15,764	16.8
421	3,853	4.10	19,617	20.9
424	3,716	3.95	23,333	24.8
427	3,613	3.84	26,946	28.6
430	3,448	3.67	30,394	32.3
432	3,278	3.48	33,672	35.8
436	3,091	3.29	36,763	39.1
437	3,015	3.21	39,778	42.3
439	2,863	3.04	42,641	45.3
441	2,744	2.92	45,385	48.2
444	2,696	2.87	48,081	51.1
446	2,627	2.79	50,708	53.9
448	2,438	2.59	53,146	56.5
450	2,449	2.60	55,595	59.1
452	2,406	2.56	58,001	61.7
454	2,237	2.38	60,238	64.0
456	2,229	2.37	62,467	66.4

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
458	2,119	2.25	64,586	68.7
460	2,081	2.21	66,667	70.9
462	2,007	2.13	68,674	73.0
464	1,967	2.09	70,641	75.1
466	1,819	1.93	72,460	77.0
468	1,839	1.96	74,299	79.0
470	1,846	1.96	76,145	81.0
473	1,692	1.80	77,837	82.7
475	1,576	1.68	79,413	84.4
478	1,492	1.59	80,905	86.0
482	1,496	1.59	82,401	87.6
483	1,412	1.50	83,813	89.1
486	1,341	1.43	85,154	90.5
490	1,262	1.34	86,416	91.9
493	1,311	1.39	87,727	93.3
497	1,222	1.30	88,949	94.6
502	1,107	1.18	90,056	95.7
505	965	1.03	91,021	96.8
509	924	0.98	91,945	97.7
512	774	0.82	92,719	98.6
515	638	0.68	93,357	99.2
518	491	0.52	93,848	99.8
521	215	0.23	94,063	100.0

Appendix Q: Test Characteristic Curves

Figure Q.1. ELA Grade 3 TCC

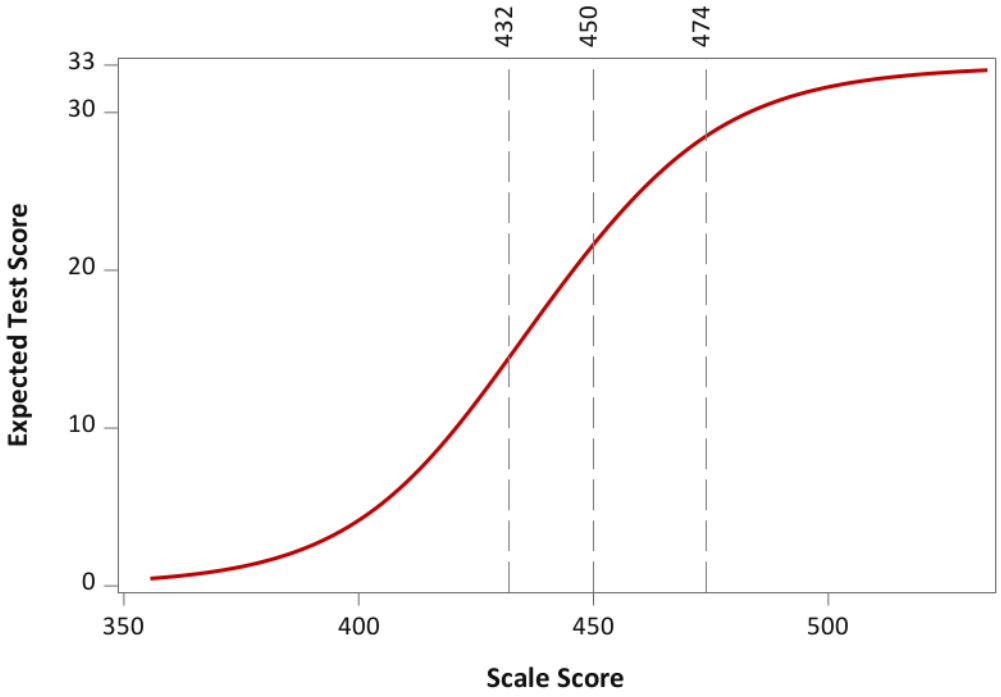


Figure Q.2. ELA Grade 4 TCC

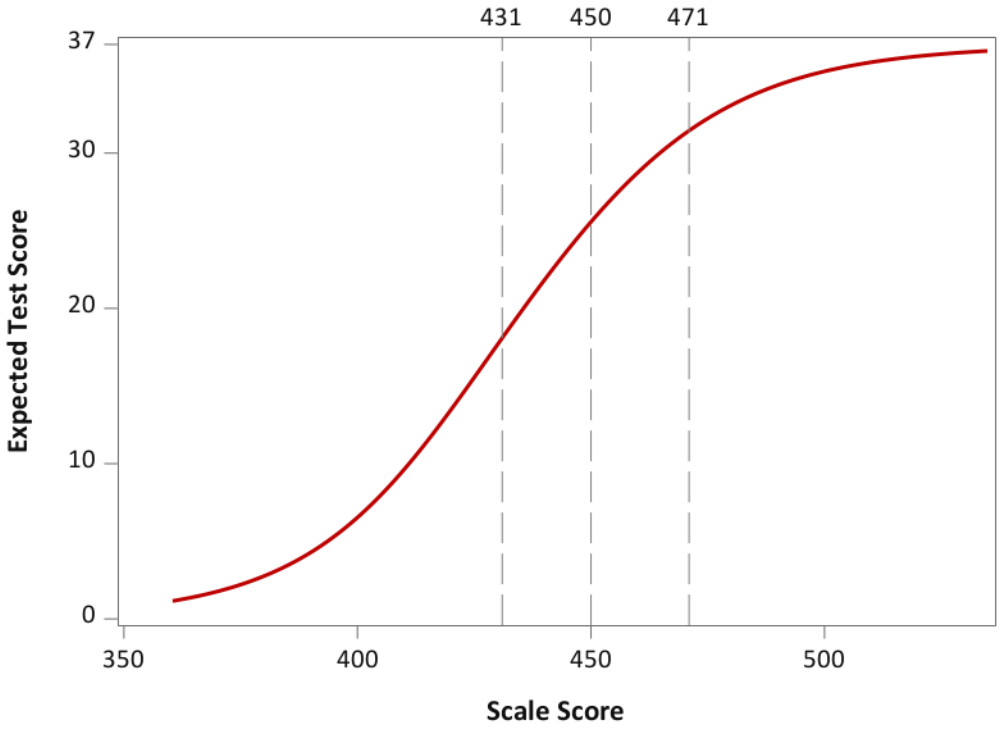


Figure Q.3. ELA Grade 5 TCC

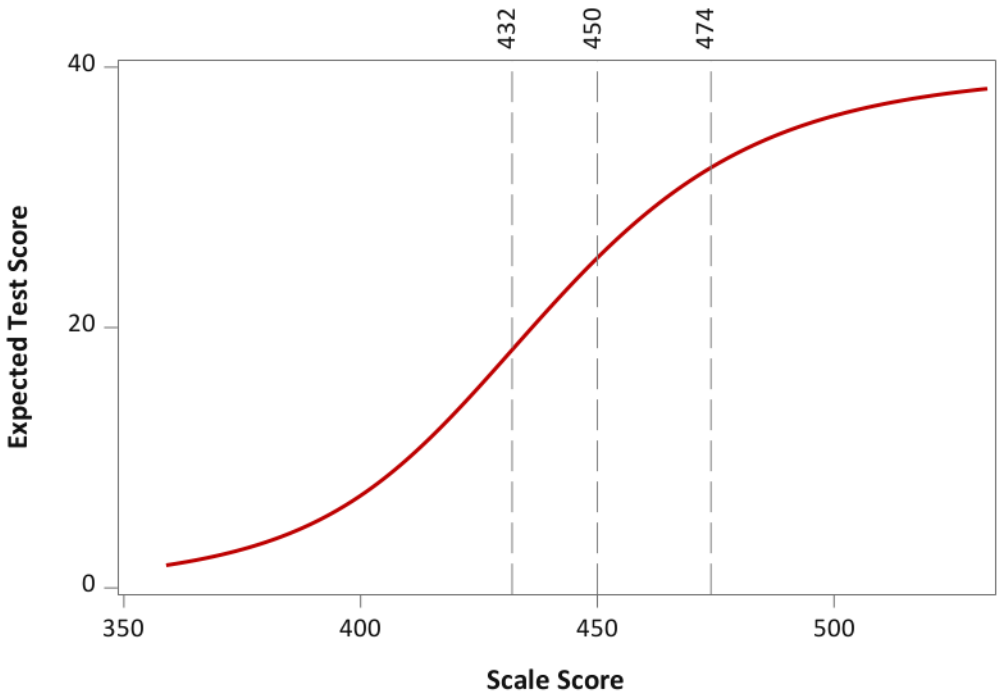


Figure Q.4. ELA Grade 6 TCC

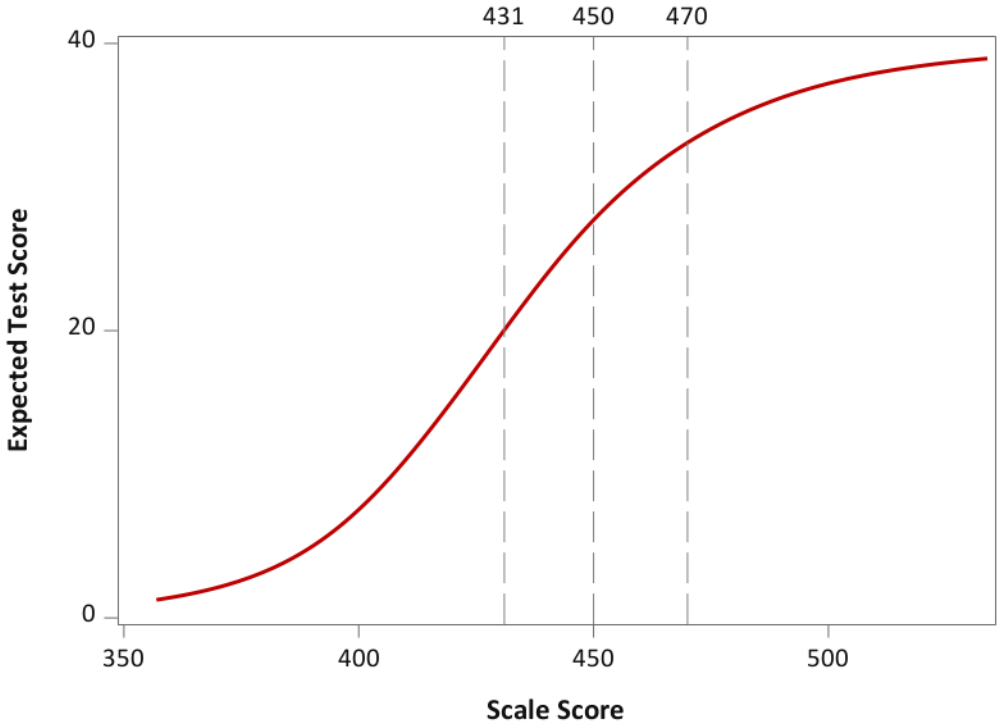


Figure Q.5. ELA Grade 7 TCC

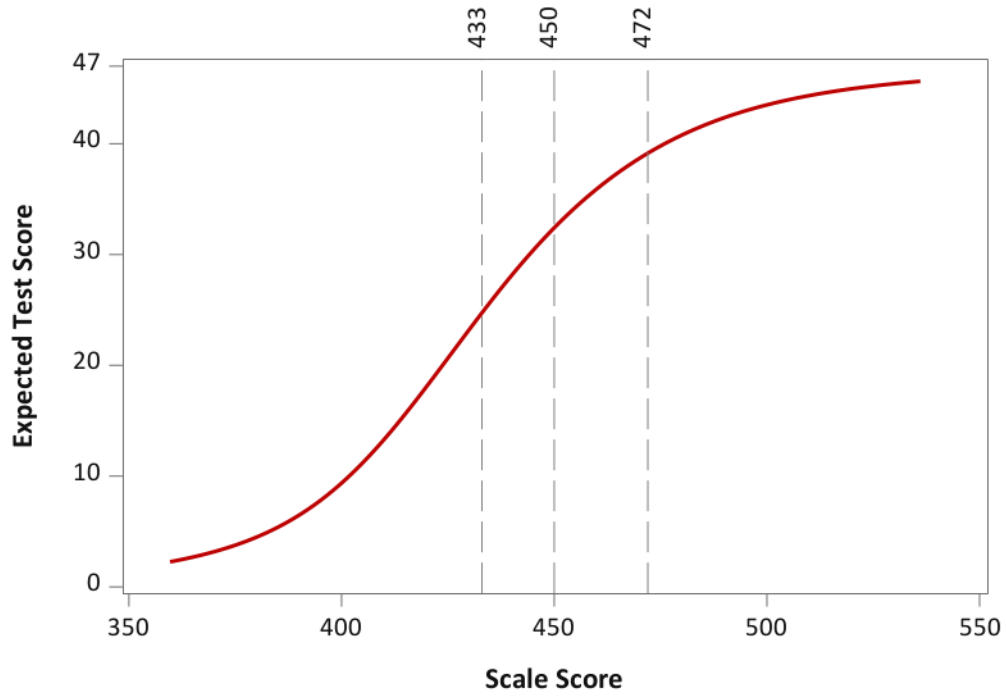


Figure Q.6. ELA Grade 8 TCC

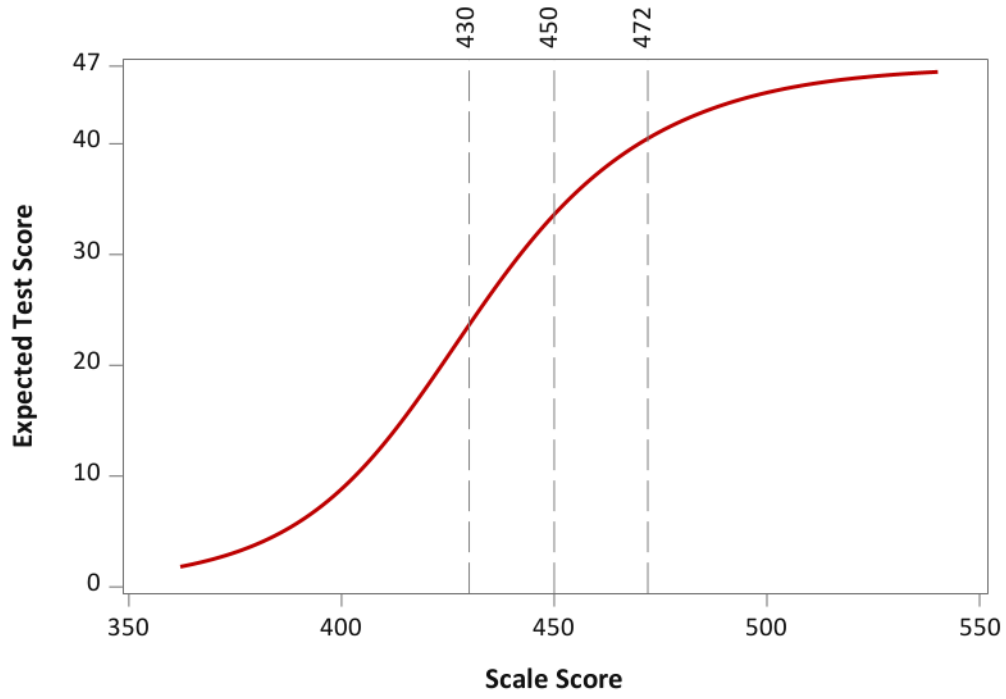


Figure Q.7. Mathematics Grade 3 TCC

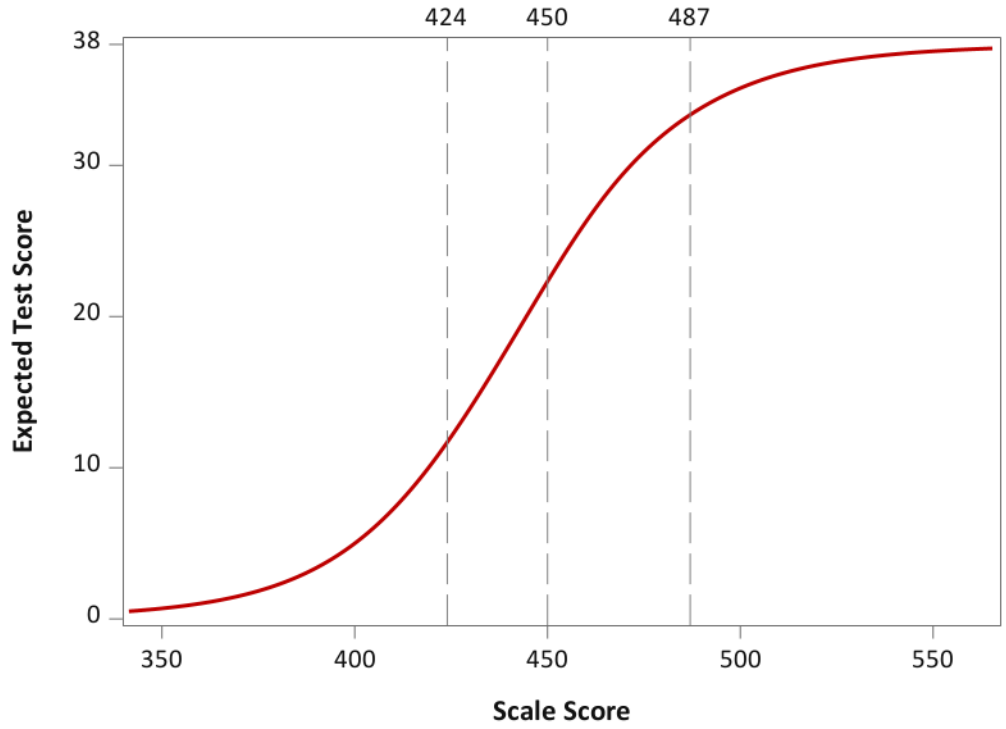


Figure Q.8. Mathematics Grade 4 TCC

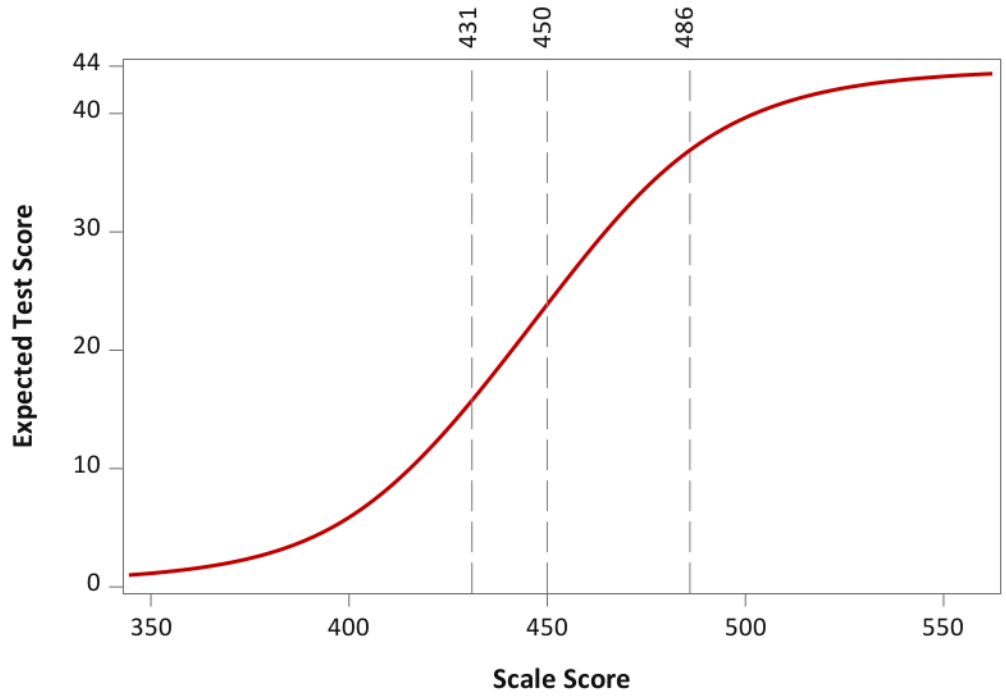


Figure Q.9. Mathematics Grade 5 TCC

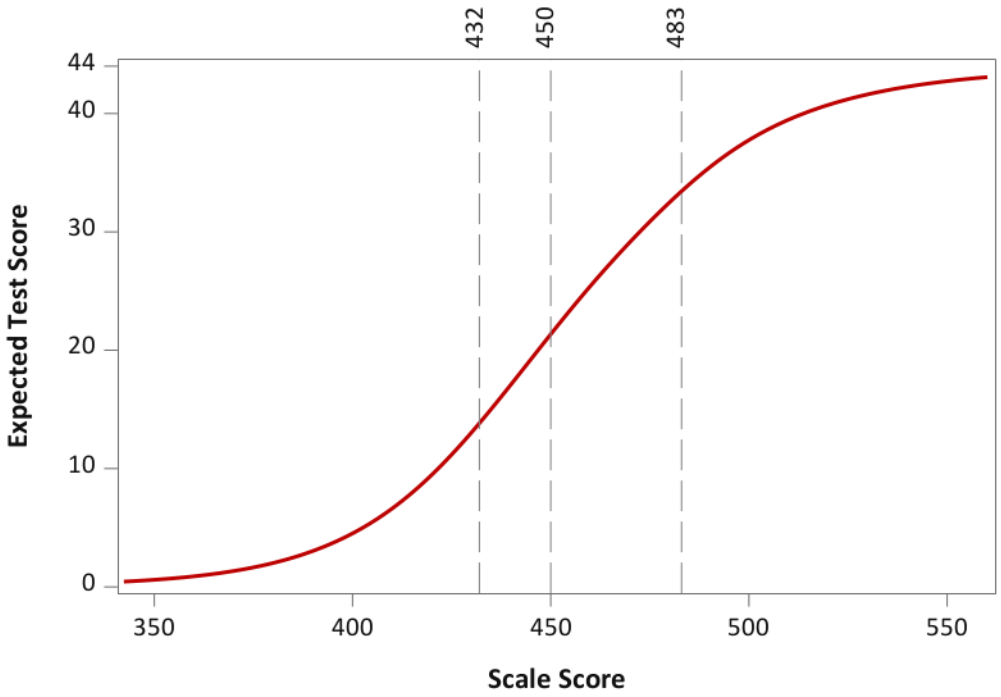


Figure Q.10. Mathematics Grade 6 TCC

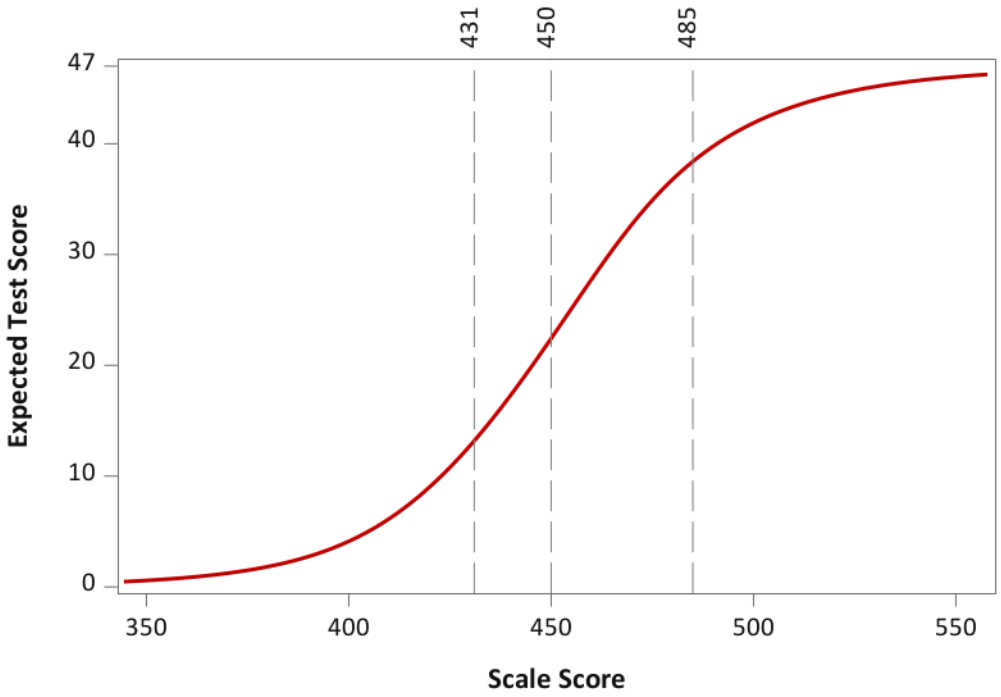


Figure Q.11. Mathematics Grade 7 TCC

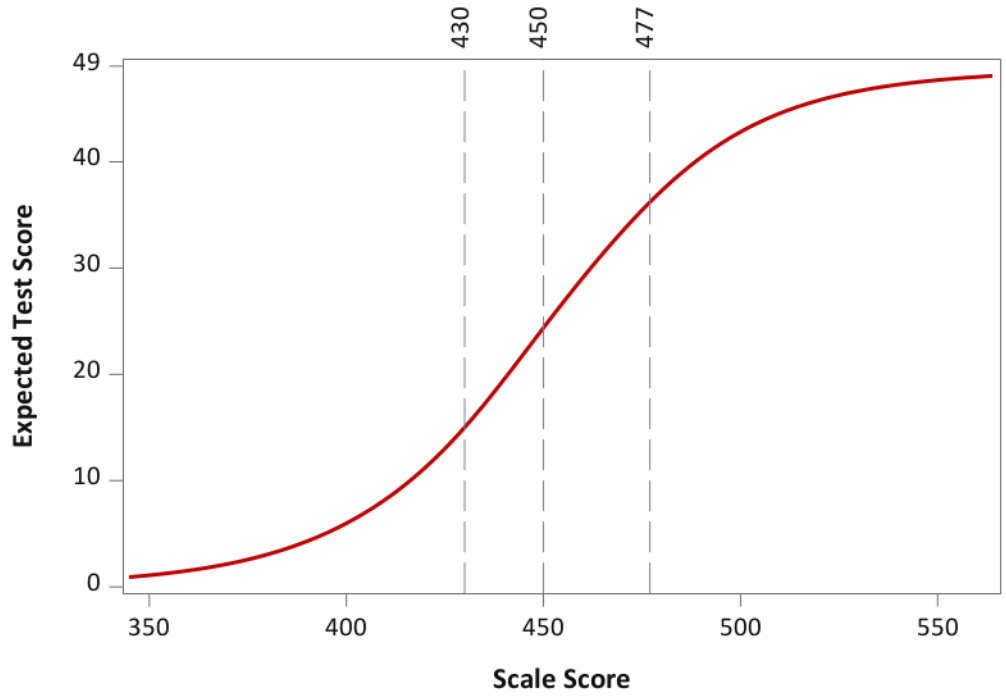


Figure Q.12. Mathematics Grade 8 TCC