New York State Testing Program 2015: English Language Arts and Mathematics Grades 3–8

Technical Report

Pearson
2015

Table of Contents

List of Tables

List of Figures

Section 1: Introduction and Overview

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 Common Core English Language Arts (ELA) and Mathematics 2015 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, differential item functioning studies, test administration, scoring, equating, scaling, and student performance.

Test Purpose

The 2015 Grades 3–8 Common Core ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade-level New York State Common Core Learning Standards (CCLS) in ELA and Mathematics. The tests are designed to allow the classification of student proficiency into four performance levels. Likewise, the test provides students at each of these performance levels opportunities to demonstrate their knowledge and skills in the CCLS. Details about the content standards for ELA and mathematics are described in Section 2, subsections "Development and Review Process," and "Test Blueprints."

Target Population

Students in New York State public school Grades 3, 4, 5, 6, 7, and 8 (and ungraded students of equivalent chronological ages) are the target population for the Grades 3–8 NYSTP. Non-public schools may participate in the testing program, but their participation is not mandatory. In 2015, some non-public schools participated in the testing program across all grade levels. These schools were included in the data analyses. Public school students were required to take all State assessments administered at their grade level, except for a very small percentage of students with severe cognitive disabilities, who took the New York State Alternate Assessment (NYSAA). For more detail on this exemption, please refer to the *NYSTP Grades 3–8 Common Core English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available online at:

http://www.p12.nysed.gov/assessment/sam/ei/eisam15rev.pdf.

Test Use and Decisions Based on Assessment

The NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State CCLS in ELA and Mathematics, respectively, in order to determine whether or not schools, districts, and the State meet the required progress targets specified in the New York State accountability system. There are several types of scores available from the Grades 3–8 ELA and Mathematics Tests, and they are discussed in this section.

Scale Scores

The scale scores are a quantification of the proficiency measured by the Grades 3–8 Common Core ELA and Mathematics Tests at each grade level. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades or across subjects. The scale scores are reported at the individual student level, and can be aggregated. Detailed information on the derivation and properties of the scale scores is provided in Section 6, "IRT Calibration and Equating." The Grades 3–8 ELA and Mathematics Tests' scale scores are the basis for placing students into proficiency levels, which are used to determine student progress within schools and districts, support registration of schools and districts, determine eligibility of students for additional educational services, and provide teachers with indicators of a student's need, or lack of need, for remediation in specific content-area knowledge.

Statewide Percentile Ranks

Students' scale scores were also presented as percentile ranks in order to indicate students' performance relative to the entire testing population on a scale that may be more familiar than the operational test's scale. Such statistics were estimated based on the how often each student earned a given scale score and thus present similar information as the scale score itself, but on an alternate scale.

Proficiency Level Cut Scores and Classification

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

> **NYS Level I:** Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered insufficient for the expectations at this grade.
>
> **NYS Level II:** Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered partial but insufficient for the expectations at this grade.
>
> **NYS Level III:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered sufficient for the expectations at this grade.
>
> **NYS Level IV:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for English Language Arts/Literacy that are considered more than sufficient for the expectations at this grade.

The definitions of performance levels for the Grades 3–8 Common Core Mathematics Tests are as follows:

> **NYS Level I:** Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered insufficient for the expectations at this grade.
>
> **NYS Level II:** Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered partial, but insufficient for the expectations at this grade.
>
> **NYS Level III:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered sufficient for the expectations at this grade.
>
> **NYS Level IV:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Common Core Learning Standards for Mathematics that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels I, II, III, and IV were established during the process of standard-setting in Summer 2013. The process is described in detail in Section 8 and Appendix P in the 2013 technical report (NYSED, 2013).

Subscores

The Grades 3–8 Common Core 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 Common Core mathematics tests have three subscores. The mathematics subscores are the domain level scores for questions measuring the *Major Clusters* in each grade. The CCLS are divided into *Major*, *Supporting*, and *Additional Clusters*. Standards within *Major Clusters* are the intended focus of instruction and assessment and account for the majority of the mathematics test items. The *Supporting* and *Additional Clusters* are mathematics standards that serve to both introduce and reinforce *Major Clusters*. Table 1 and Table 2 below present the reporting subscore categories and the point values that correspond to each on the 2015 tests. In 2015, subscores were reported in two ways: (1) a raw score (i.e., number of points earned) out of the total score on the test; and (2) the average score at the state level for each subscore category.

Table 1. ELA Subscore Categories and Total Possible Score Points

Grade	Total Subscore Points	
	Reading	Writing to Sources
3	31	24
4	31	24
5	42	24
6	42	24
7	42	24
8	42	24

Table 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 27	Number and Operations—Fractions 12	Measurement and Data 12
4	Operations and Algebraic Thinking 11	Number and Operations in Base Ten 17	Number and Operations—Fractions 18
5	Number and Operations in Base Ten 18	Number and Operations—Fractions 25	Measurement and Data 10
6	Ratios and Proportional Relationships 18	The Number System 12	Expressions and Equations 28
7	Ratios and Proportional Relationships 20	The Number System 14	Expressions and Equations 22
8	Expressions and Equations 30	Functions 19	Geometry 12

Testing Accommodations

In accordance with federal law under the Americans with Disabilities Act and the section, Fairness in Testing, as outlined by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. The allowance is in accordance with a student's Individualized Education Program (IEP) or Section 504 Accommodation Plan (504 Plan). School principals are responsible for ensuring that proper accommodations are provided when necessary, and that staff providing accommodations are

properly trained. Details on testing accommodations can be found in the 2015 School Administrator's Manual (SAM).

Test Transcriptions

For visually impaired students, large-type and Braille editions of the test books are provided. In most cases, the students dictate and/or record their responses, the teachers transcribe student responses to the multiple-choice (MC) items onto scannable answer sheets, and the teachers transcribe the responses to the constructed-response (CR) items onto the regular test books. Some of the students who use large-type editions will fill in the answer sheets by themselves. The large-type editions are created by Pearson and printed by the New York State Education Department (NYSED), and the Braille editions are produced by gh, LLC. gh employs certified Library of Congress Braille transcribers and delivers Braille in accordance to the Braille Authority of North America (BANA) standard. Camera-copy versions of the regular test books are provided to the Braille vendor, which then produces the Braille editions. Proofs of the Braille editions are submitted to NYSED for review and approval prior to production.

Test Translations

The NYSTP Grades 3–8 Common Core Mathematics Tests are translated into five languages: Chinese, Haitian-Creole, Korean, Russian, and Spanish. These tests are translated to provide students the opportunity to demonstrate mathematical proficiency independent of their command of the English language. Sample tests are available in each translated language at the following location:

http://www.p12.nysed.gov/assessment/math/samplers/

English language learners taking the Grades 3–8 Common Core 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. Additionally, the following testing accommodations were made available to English language learners: time extension, 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 Common Core ELA Tests are not translated into any other language because they are assessments of proficiency in English language arts.

Section 2: Test Design and Development

Test Descriptions

The 2015 Grades 3–8 Common Core ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) test items based on the New York State P–12 Common Core Learning Standards (CCLS). The tests were administered in New York State classrooms during April 2015 over a three-day period. Details on the administration and scoring of these tests can be found in Section 4, "Test Administration and Scoring." Additional information can be found in the *NYSTP Grades 3–8 Common Core English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available at:

http://www.p12.nysed.gov/assessment/sam/ei/eisam15rev.pdf.

ELA Tests

The 2015 Grade 3–8 Common Core English Language Arts Tests were designed to measure student literacy as defined by the CCLS. The tests assessed Reading, Writing, and Language standards by using multiple-choice, short-response, and extended-response questions. All questions were based on close readings of informational, literary, or paired texts. All texts were drawn from authentic, grade-level works.

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

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

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

Mathematics Tests

The 2015 Grade 3–8 Common Core Mathematics Tests were designed to measure student mathematic understanding as defined by the CCLS. The tests required that students understand mathematics conceptually, use prerequisite skills with grade-level mathematical facts, decide which formulas and tools (e.g., protractors and rulers) to use, and solve mathematic problems rooted in the real world. The tests contained multiple-choice, short-response (2-point), and extended-response (3-point) questions. For multiple-choice questions, students selected the correct response from four answer choices. For short- and extended-response questions, students wrote an answer to an open-ended question. Some questions required students to show their work or to explain, in words, how they arrived at their answers.

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

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

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

Test Configuration

Test Book Design and Testing Times

The 2015 Grades 3–8 Common Core ELA Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book; Book 1 and Book 2 contained literary and informational reading passages and MC items based on the passages. Book 2 also contained reading passages with short-response items and an extended-response item based on those passages. Book 3 contained only reading passages with short-response items and an extended-response item based on those passages.

The 2015 Grades 3–8 Common Core Mathematics Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book. Book 1 and Book 2 contained MC items. Book 3 contained short- and extended-response items. The tables in Appendix A provide information on the numbers and types of items in each book for the Grades 3–8 Common Core ELA and Mathematics Tests and the testing times.

Embedded Field-Test Items

In 2010, the Department 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 2015 administrations of the Common Core assessments. Embedding field-test items allows for a better representation of student responses and provides more reliable field-test data on which to build future operational tests. In other words, since the specific locations of the embedded field-test items were not disclosed and they look the same as operational test items, students were unable to differentiate field-test items from operational test items. Therefore, field-test data derived from embedded items are free of the effects of differential student motivation that may characterize stand-alone field-test designs. Embedding field-test items also reduced the number of stand-alone field-tests during the Spring of 2015, but did not eliminate the need for them.

New York State Educators' Involvement in Test Development

New York State educators are actively involved in Common Core ELA and Mathematics test development. New York State educators provide critical input throughout all stages of the test development process, which include Educator Item Review, Rangefinding, Final Eyes Meeting (a final review of the test books prior to printing) and Standard-Settings.

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; and
- Type of school (urban, suburban, or rural).

The selected participants must be certified and have both teaching and testing experience. The majority of the participants are classroom teachers. In addition, specialists such as reading coaches, literacy coaches, and special education and bilingual instructors participate, as well. Some participants are also recommended by principals, professional organizations, Big Five Cities, and/or the Staff and Curriculum Development Network (SCDN). A file of participants is maintained and is routinely updated with current participant information and the addition of possible future participants, as recruitment forms are received. This gives many educators the opportunity to participate in the test development process. Every effort is made to have diverse groups of educators participate in each testing event.

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

Development and Review Process

During the process of transitioning from legacy assessments to the new CCSS-aligned assessments, NYSED and Pearson sought consultation with the following:

- New York State Educators
- Student Achievement Partners
- College Board
- HumRRO (Human Resources Research Organization)

Test Blueprints

After careful consideration of administration constraints (e.g., feasibility of paper-based tests versus online tests, number and length of test forms, and location of multiple-choice and constructed-response items within test books) and timing constraints, the representation and distribution of content was determined.

The CCLS for ELA are organized into four strands: Reading, Writing, Language, and Speaking/Listening. Due to administration constraints, Speaking/Listening was determined to best be assessed in the classroom only; therefore, the Common Core ELA Tests assess three of the four strands: Reading, Writing, and Language. Content experts reviewed the Reading, Writing, and Language standards and recommended content coverage by standard and item-type based on the depth and breadth of each standard.

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

Tables B1 and B2 in Appendix B show the test blueprint and actual number of score points in the Grades 3–8 Common Core ELA and Mathematics Tests, respectively. Included in the tables are the ranges of allowable points for each ELA Strand and mathematics Domain and the actual number of points on the 2015 operational tests.

Passage Selection and Item Criteria Documents

To guide test item development and to help ensure that NYS tests were measuring the CCLS for ELA and mathematics with fidelity, NYSED and Pearson established criteria for selecting passages and writing test items based on the consultation with the groups listed above.

The *Passage Selection Guidelines for Assessing CCSS ELA* were created to provide a framework that allows for the consistent selection of passages that are appropriately complex for the given grade; and contain the specific characteristics necessary to measure different standards (see Appendix C). The guidelines describe the quantitative methods used to determine the grade appropriateness of a given text. They also describe the grade-specific text characteristics needed

to develop questions that measure any particular reading standard. The complete guidelines can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/passage_selection_guidelines_for_assessing_ccss_ela.pdf.

Passage Review Criteria documents were created based on the passage selection guidelines and were used to evaluate each potential passage and determine whether or not it could be used to measure the CCSS for ELA. The criteria documents were used to determine whether or not each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas. The full passage review criteria can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/new_york_state_passage_review_criteria_protocol_document.doc.

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

http://www.engageny.org/sites/default/files/resource/attachments/ela_item_review_criteria_grades_3-5.doc

and here:

http://www.engageny.org/sites/default/files/resource/attachments/ela_item_review_criteria_grades_6-8.doc.

Item Review Criteria for Grade 3–8 Mathematics Tests were used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to CCSS, and conformity to the expectations for specific item types and formats for each test question. Each

section of the criteria includes pertinent questions that determine whether or not an item is of sufficient quality. The first two categories, Clarity and Graphical Appropriateness and Fairness, identify the basic components of quality assessment items. The criteria for Clarity and Graphical Appropriateness are used to help ensure that students understand what is asked in each question and that the language in the question 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 question containing art is reasonable, and that interpreting a graphic does not confuse the underlying construct. Likewise, the Fairness criteria are used to evaluate whether or not questions are un-biased, 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 criterion addresses the specific demands for different item types and formats. For example, the criteria for a three-point constructed-response item include making sure that the items involve a multi-step process and require students to show work. The complete math criteria can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/math_item_review_criteria.doc.

The Multiple Representations for NYS Grade 3–8 Common Core Mathematics Tests document was developed to ensure that the tests measured the deep conceptual understanding that CCSS demands, rather than focusing on predictable mathematics questions that require only algorithmic strategies to be solved correctly. Multiple Representations are a broad set of specifications that describe, refer, and symbolize the various, but not all, ways that mathematics standards could be measured within the constraints of the NYSTP. The document specifies three overarching families: procedural skills, conceptual understanding, and application. It also includes information about how to identify standards that might be measured through the use of a particular representation, and identifies types of mathematics skills (e.g. application of process; explanation of a principle, etc.) that are appropriate for assessing different representations. The full document can be found here:

http://www.engageny.org/sites/default/files/resource/attachments/mathematics_multiple_representations.pdf.

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

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased 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 checklist (Universal Design Item Checklist) in Appendix D was developed for use during item development.

Passage Finding

The goal of passage finding is to obtain high-quality texts from which to generate CCSS-aligned test questions. To do so, Pearson recruited independent passage finders and trained them, using passage selection resources including the passage selection criteria. Passage finders were given assignments based on the test blueprint requirements. Passage finders submitted passages along with completed criteria documents and source information to Pearson ELA content specialists, who reviewed the passages against the agreed-upon criteria. Passages that did not meet the criteria were rejected, and passages that did meet the criteria were moved forward in the process, where the text from scanned copies of the original sources was entered into templates. Once in the templates, readability metrics were determined for each text, and it was then proofread by Pearson copyeditors, fact checked by Pearson research librarians, reviewed for subject-specific content issues by science and social studies content specialists, and reviewed for universal-design issues by specifically trained Pearson reviewers. After the passages went through these review steps, Pearson ELA content specialists posted the passages and completed criteria documents for NYSED's review and approval for moving forward in the process.

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

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

Item Development

The goal of item development is to develop a sufficient number of high-quality, CCSS-aligned items to populate the test forms. Using the criteria documents for both subjects and the multiple-perspective document for mathematics, Pearson content leads trained item writers. The item writers had teaching or assessment experience in the subject area for which they were writing items, experience in writing for large-scale, high-stakes assessments, and, at least, a bachelor's degree in either education and/or the subject area for which they were assigned. The item writers were given specific assignments based on the test blueprint. For ELA, the item writers were also provided with the completed passage criteria documents.

The item writers provided items and completed criteria documents to Pearson content specialists for review. Two content specialists reviewed each item and its corresponding criteria document, and any items that did not meet the criteria were sent back to the writers with specific feedback

for revision. Items that did not meet the criteria after an attempted revision were rejected and replaced by Pearson content specialists. After the Pearson content specialists were satisfied that all of the items met the criteria, the items were reviewed by Pearson copyeditors. The mathematics items were also reviewed by subject-specific content specialists in science and social studies and by research librarians. The Pearson ELA and mathematics content specialists evaluated the feedback from the different internal groups and edited the items accordingly. Then, the items and criteria documents were posted for NYSED's review and approval for moving forward in the process.

NYSED content experts retrieved the items and criteria documents and reviewed both the items and criteria documents. If NYSED staff determined that an item did not meet the criteria, the item was rejected, and the NYSED staff provided Pearson with an explanation for the reason for rejection. Pearson then replaced the item and completed criteria documents, which were resubmitted to NYSED. If NYSED staff determined that an item met the criteria, but could be improved with editing, the staff member recorded notes for the edits. Those notes were reviewed at face-to-face meetings at which Pearson content staff and NYSED staff reviewed and edited all of the items to ensure that they met the criteria. All passages and items accepted at that meeting were moved forward for the Educator Item Review.

Educator Item Review

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

1. Does the item align to the designated standard(s)?
    - The item measures the content standard(s) that it was designed to measure.

2. Does the item meet quality standards?
    - The item is worded clearly.
    - The reading level of the item is grade appropriate.
    - The item has one correct answer.
    - The item has plausible, unambiguous distractors.
    - All of the distractors are mutually exclusive.

3. Is the item fair?
    - The item is free from bias on the basis of students' personal characteristics, such as gender or ethnicity.

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

Field-Testing

Once the items have been developed and thoroughly reviewed by a variety of stakeholders, they must then be field-tested. The process of field-testing items is a critically important step in the test development process, as it is only through the gathering of actual student response data that a variety of psychometric characteristics may be evaluated. In particular, a summary of the unique items that passed the scrutiny of NYSED and Pearson content specialists, as well as that of New York State educators, and were field-tested is included in Table 3. Note that more items were field tested than were needed on the operational forms because that enabled tests to be constructed with the best possible characteristics from both a content and psychometric perspective.

Table 3. Summary of Unique 2014 Field-Test Items

Grade	Unique ELA Items by Type		Unique Mathematics Items by Type	
	Multiple-Choice	Constructed-Response	Multiple-Choice	Constructed-Response
3	126	48	96	22
4	125	48	120	25
5	138	48	120	25
6	137	48	132	25
7	137	48	132	25
8	137	48	135	25

Note. All constructed-response items were field-tested under stand-alone conditions, while multiple-choice items were administered under both embedded and stand-alone conditions.

The first set of field test items were administered in the Spring of 2014 as embedded field test items within the 2014 operational test forms. As was noted above, the use of embedded field test items both yielded more reliable field-test data and reduced—but did not eliminate—the need for multiple-choice stand-alone field testing. One additional round of field testing was administered separately from the 2014 operational forms (i.e., as stand-alone tests)—later in the Spring of 2014.

In order to better understand how 2014 field test items may perform on future operational forms, a variety of analyses were conducted. All of the field test data underwent a series of representativeness checks. Because only a small sample of schools participate for any given subject and grade for stand-alone field testing, it was necessary to ensure that the stand-alone field test samples were representative of the entire State population in terms of student achievement on prior years' tests, student gender, student ethnicity, and school Needs/Resource Capacity Category (NRC). Finally, a variety of psychometric analyses were conducted, including classical item analysis, inter-rater reliability for constructed-response items, differential item functioning (DIF), item response theory (IRT) item calibration, equating, scaling, and fit evaluation. Many of these analyses are described at length below; however, inter-rater reliability analyses were not possible for the operational test, as only a single rater scored each constructed-response.

Rangefinding

After constructed-response items have been field-tested, rangefinding occurs. The purpose of rangefinding is to have New York State teachers review student constructed-responses and arrive at consensus scores that are 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 teachers review, discuss, and rate student responses to the field test questions. This process is overseen by NYSED content experts and Pearson Scoring Directors. The first step in the rangefinding process was to have the teacher committees review rubrics and a NYSED-approved grounding guide set, previously used for the 2014 field-test rangefinding sessions, to familiarize teachers with the application of NYSED standards and rubrics. The grounding guide sets contain student responses that illustrate the full range of scores on the rubric. The grounding guide sets are composed of student responses that had previously gone through the rangefinding process and been approved by NYSED to guide the scoring of field test and operational student responses. Referencing the previously approved guide set papers during the rangefinding sessions ensures consistency in the application of NYSED standards and rubrics from year-to-year.

After the committee reviewed the preapproved anchor set, the committee members familiarized themselves with each item type by group, scoring a small number of responses that were representative of each of the different score points. Upon completion of the group-scoring exercise, committee members independently scored other student responses. After the independent scoring was completed, the committee reviewed and discussed their results and determined consensus scores for the responses. The rangefinding results were then used to build training materials for Pearson scorers, who scored the field-test responses to constructed-response items.

Item Selection and Test Creation (*Criteria and Process*)

The NYSTP Grades 3–8 Common Core ELA and Mathematics Tests were administered in April 2015. The test items were selected from the pools of available ELA and mathematics items. These items were field-tested either in embedded field-testing or stand-alone field-testing in 2013 or 2014.

The test construction process involved several iterative steps. Three criteria governed the item selection process. The first of these was to meet the ELA and mathematics content specifications provided by NYSED; the second was for content experts to select items with the best psychometric characteristics from the ELA and mathematics item pools, respectively; the third criteria required that the combined psychometric characteristics of all selected items combined with the intended psychometric goals for each entire form. Pearson content specialists were provided with the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines were based on the classical and item response theory (IRT) statistics associated with the test items. Using the pool of field-tested items, content specialists made preliminary selections for each grade and subject. The selections were then reviewed by the content leads for each subject to make sure that the items conformed to the different criteria. If the content criteria were not met, new items were selected. After the content leads' review, the item selections were reviewed by Pearson psychometricians. If items with undesirable statistics

were selected, the psychometricians proposed items with more desirable statistics and those items were reviewed by the content specialists and their leads. Once both the Pearson content teams and the psychometric teams were satisfied that the content and statistics of the selected items, and the proposed whole forms met the requirements, the items were given to NYSED staff (including content and assessment experts) to review. Pearson content specialists and psychometricians traveled to Albany, New York, in October 2014 to finalize item selection and test creation with NYSED staff (including content and assessment experts). NYSED discussed the content and data of the proposed selections, explored alternate selections for consideration, approved the item selections, and assigned item positions to those items in the operational test books.

Test Form Production and Reviews

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

Final Eyes Committee

After NYSED and Pearson staff reviewed copies of the test forms, the test forms were reviewed by the Final Eyes Committees. For each subject the committee consisted of approximately 12 Grade 3–8 New York State educators from around the state. During that review, the educators were charged with taking the test to make sure that each multiple-choice item had a single correct answer and to look for spelling, capitalization, punctuation, grammar, and formatting errors. Upon completion of the Final Eyes review and after NYSED approved edits made as a result of the review, the tests were then considered final. The test files were then produced for the April 2015 administration.

Proficiency and Performance Standards

In summer 2013, after the operational administration of the 2013 tests, a standard-setting meeting occurred in Albany where at which 95 New York State educators went through a rigorous process guided by the best practices indicated by this intensely studied process to recommend performance standards for the new tests measuring the Common Core Learning Standards. These recommendations were presented to the Commissioner and the Board of Regents, who, in turn, adopted the recommended standards set forth by the committees. For additional details, see Section 8 and Appendix P in the 2013 technical report (NYSED, 2013).

For each grade level, there are four proficiency levels. Three cut points demarcate the performance standards needed to demonstrate each ascending level of proficiency. Detailed information related to performance standards can be found in Section 6, subsection, “Raw Score-to-Scale Score and SEM Conversion Tables.”

Section 3: Validity

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

The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) *Standards for Educational and Psychological Testing* (2014) addressed the concept of validity in testing. Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. 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.

Content Validity

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

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

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

An external evaluation was conducted to review the development process and to investigate the degree to which NYSTP measures the CCLS. Human Resources Research Organization (HumRRO) found that NYSTP development process meets or exceeds the industry standards indicated in the *Standards for Educational and Psychological Testing* (2014). HumRRO also evaluated the degree to which the Grades 3–8 Common Core ELA and Mathematics Tests measured the CCLS. HumRRO concluded that the Grades 3–8 Common Core ELA and Mathematics Tests do assess the content described by the CCLS. Additionally, HumRRO found that NYSTP measured the CCLS at the intended Depth of Knowledge (DOK).

Construct (Internal Structure) Validity

Construct validity—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 are supported by several types of evidence that can be obtained from the ELA and mathematics test data, respectively.

Internal Consistency

Empirical studies of the internal structure of the test provide one type of evidence of construct validity. For example, high internal consistency constitutes evidence of validity. This is because high coefficients imply that the test items are measuring the same domain of skill and are reliable and consistent. Reliability coefficients of the tests for total populations and subgroups of students are presented in Section 7 subsection, "Test Reliability." For the total population, the ELA reliability coefficients (Cronbach's alpha) ranged from .89 to .92, and for all subgroups, the reliability coefficients were greater than or equal to .81. For the total population, the mathematics reliability coefficients (Cronbach's alpha) ranged from .93 to .95, and for all subgroups, the reliability coefficients were greater than or equal to .80. Overall, high internal consistency of the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests provided sound evidence of construct validity.

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 equate the tests, as well as to generate student scores. The models, among other things, require that the items fit the model well (item fit) and that the items in a test measure a single domain of skill (unidimensionality).

The first step is to assess the degree to which the items fit the IRT model. The item-model fit for the mathematics and ELA tests was assessed using Q_1 statistics (Yen, 1981), and the results are described in detail in Section 6, "IRT Calibration and Equating." That the majority of the items demonstrated sound fit across grades and subjects, and only a few items were deemed to have deviate fit, 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 New York State Tests are related to each other, within the respective subject areas. This relationship

of the items within the ELA or mathematics tests is, simply stated, 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 among responses to test items. A large first component in factor analysis would provide evidence of the latent proficiency that students have in common with respect to the particular items asked. A large main factor found from a factor analysis of an achievement test would suggest a primary construct that may be related to what the items were designed to have in common (i.e., mathematics proficiency or ELA proficiency).

To demonstrate the common factor underlying student responses to ELA and mathematics test items, principal component factor analyses were conducted on a correlation matrix of individual items for the ELA and mathematics tests, respectively. Factoring a correlation (i.e., tetrachoric correlation) matrix rather than actual item response data is preferable when dichotomous variables are in the analyzed data set. Because the New York State ELA and mathematics tests contain both MC and CR items, the matrices of *polychoric* correlations were used as input for the factor analyses, as polychoric correlations are appropriate with both MC and CR data. The study was conducted on the New York State public, charter, and non-public school students for whom data were available during the equating process. 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 further investigate the factor analysis results to determine the number of "meaningful" factors. Specifically, more than one factor with an eigenvalue greater than 1.0 present in each data set would suggest the presence of small additional factors. Additionally, the magnitude of the ratio of the variance accounted for by the first factor compared to the remaining factors also provides evidence as to the number of meaningful factors. In addition, the total amount of variance accounted for by the main factor was evaluated. According to M. Reckase (1979),

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

Factor analyses related to the Grades 3–8 Common Core ELA and Mathematics Tests indicated that the ratio of the variance accounted for by the first factor to the remaining factors was sufficiently large to support the claim that the ELA and mathematics tests were essentially

unidimensional; the ELA-related ratios and the mathematics-related ratios showed that the first eigenvalues were at least five times as large as the second eigenvalues for all of the grades.

It was found that all of the New York State Grades 3–8 Common Core ELA and Mathematics Tests exhibited first principal component accounting for more than 19% and 30% of the test variance, respectively. The results of factor analyses, including eigenvalues greater than 1.0 and proportions of variance explained by the extracted factors, are presented for ELA (see Table 4) and mathematics (see Table 5).

Table 4. ELA Tests Factor Analysis

Grade	Extracted Factor			
	#	Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
3	**1**	**8.21**	**20.02**	**20.02**
	2	1.65	4.02	24.04
	3	1.32	3.21	27.25
	4	1.03	2.52	29.77
	5	1.00	2.44	32.21
4	**1**	**8.22**	**20.05**	**20.05**
	2	1.35	3.29	23.34
	3	1.14	2.77	26.11
	4	1.03	2.51	28.62
	5	1.01	2.46	31.08
5	**1**	**10.39**	**19.98**	**19.98**
	2	1.62	3.11	23.09
	3	1.31	2.51	25.60
	4	1.08	2.08	27.68
	5	1.03	1.99	29.67
6	**1**	**10.17**	**19.55**	**19.55**
	2	1.58	3.03	22.58
	3	1.23	2.37	24.95
	4	1.18	2.26	27.21
	5	1.00	1.92	29.13
7	**1**	**9.72**	**18.68**	**18.68**
	2	1.94	3.73	22.41
	3	1.23	2.36	24.77
	4	1.08	2.07	26.84
	5	1.05	2.01	28.85
8	**1**	**10.82**	**20.80**	**20.80**
	2	1.89	3.64	24.44
	3	1.18	2.27	26.71
	4	1.09	2.10	28.81
	5	1.04	2.00	30.81

This evidence supports the claim that there is one single construct underlying 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.

Table 5. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
	#	Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
3	**1**	**12.51**	**25.53**	**25.53**
	2	1.75	3.58	29.11
	3	1.24	2.52	31.63
	4	1.11	2.26	33.89
4	**1**	**15.38**	**29.57**	**29.57**
	2	1.49	2.87	32.44
	3	1.32	2.53	34.97
	4	1.15	2.21	37.18
	5	1.03	1.98	39.16
5	**1**	**12.90**	**24.80**	**24.80**
	2	2.13	4.10	28.90
	3	1.14	2.20	31.10
	4	1.02	1.96	33.06
6	**1**	**15.32**	**26.42**	**26.42**
	2	1.73	2.98	29.40
	3	1.34	2.31	31.71
	4	1.05	1.81	33.52
	5	1.02	1.76	35.28
7	**1**	**14.49**	**24.98**	**24.98**
	2	2.09	3.61	28.59
	3	1.21	2.09	30.68
	4	1.07	1.85	32.53
	5	1.01	1.73	34.26
8	**1**	**12.98**	**22.38**	**22.38**
	2	1.74	3.00	25.38
	3	1.26	2.17	27.55
	4	1.06	1.83	29.38
	5	1.05	1.81	31.19

This evidence supports the claim that there is a common construct underlying 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.

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

Detection of Bias

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

The developers of the NYSTP tests gave careful attention to items of possible ethnic, gender, socioeconomic status (SES), and—only for the mathematics tests—translation bias. All materials were written and reviewed to conform to Pearson's editorial policies and guidelines for equitable assessment, as well as NYSED's guidelines for item development. At the same time, all materials were written to NYSED's specifications and carefully checked by groups of trained New York State educators during the item review process. These steps are essential in keeping bias to a minimum. However, current evidence suggests that expertise in this area is no substitute for data; reviewers are sometimes wrong about which items work to the disadvantage of a group, apparently because some of their ideas about how students will react to items may be faulty (Sandoval and Mille, 1979; Jensen, 1980). Thus, empirical studies were conducted.

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

In each grade, for both the ELA and mathematics tests, few items were flagged for DIF. Moreover, the magnitude of DIF for the flagged items was typically small (for more detail, see

Appendix N). In addition, very few items were flagged by multiple methods. Items that were flagged for statistically significant DIF were carefully reviewed by multiple reviewers during the operational test item selection. All such items were deemed by the reviewers to be free of bias (i.e., judged not to adversely affect any demographic subgroup studied) and remained in the tests.

Section 4: Test Administration and Scoring

Listed in this section are brief summaries of New York State test administration and scoring procedures. For further information, refer to the aforementioned School Administrator's Manual and the *New York State Scoring Leader Handbook (2015)* located here:

http://www.p12.nysed.gov/assessment/ei/2015/scoringleaderhandbook15.pdf.

Test Administration

NYSTP Grades 3–8 Common Core ELA and Mathematics Tests were administered to students during April 2015. The testing window was Tuesday, April 14–Thursday, April 16 for the Grades 3–8 Common Core ELA Tests and Wednesday, April 22–Friday, April 24 for the Grades 3–8 Common Core Mathematics Tests. The makeup test administration window was Friday, April 17–Tuesday, April 21 for the Grades 3–8 Common Core ELA Tests and Monday, April 27–Wednesday, April 29 for the Grades 3–8 Common Core Mathematics Tests. The makeup test administration windows allowed students who were ill or otherwise unable to test during the assigned window to take the tests.

Scoring Procedures of Operational Tests

The scoring of the Common Core operational tests was performed at designated sites by qualified teachers and administrators. The number of personnel at a given site varied, as districts have the option of regional, district-wide, or school-wide scoring (please refer to the next subsection, "Scoring Models," for more details). Administrators were responsible for the oversight of scoring operations, including the preparation of the test site, the security of test books, and the supervision of the scoring process. At each site, designated trainers taught scoring committee members the basic criteria for scoring each item and monitored the scoring sessions in the room. The trainers were assisted by facilitators or leaders, who also helped in monitoring the sessions and enforced scoring accuracy.

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

Scoring Models

For the 2014–2015 school year, schools and school districts were able to score Grades 3–8 Common Core ELA and/or Mathematics Tests regionally, multi-district, district-wide, or school-wide based on local need. Schools were required to enter one of the following scoring model codes on student answer sheets:

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

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

Scoring of Constructed-Response Items

The key resource for both the training of scoring committee members and the scoring of CR items was the scoring guides. These documents were created by Pearson from sets of actual field test student responses that were consensus scored by NYSED and New York State teachers during Rangefinding sessions. Trainers used these materials to train scoring-committee members on the criteria for scoring CR items. Additionally, scoring leader handbooks were also distributed to outline the responsibilities of the scoring roles. Pearson and NYSED staff also conducted turnkey training sessions across the state to better equip the teachers and administrators with enhanced knowledge of scoring principles and criteria.

Upon completion of the training of scoring committee members, scoring was conducted with pen-and-pencil scoring as opposed to electronic scoring, and each scoring-committee member evaluated actual student papers instead of electronically scanned papers. All scoring-committee members were trained by previously trained and approved trainers along with guidance from scoring guides. Each constructed-response test book was scored by three separate scoring committee members, who scored three distinct sections of the test book. After test books were

completed, the table facilitator or subject (ELA or mathematics) leader conducted a "read behind" of approximately 12 sets of test books per hour to verify the accuracy of scoring. If an item arose that was not covered in the training materials, facilitators or trainers were to call the New York State ELA Helpline or the New York State Mathematics Helpline for assistance with the ELA or mathematics scoring, respectively (see the subsection "Quality Control Process").

Scorer Qualifications and Training

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

Regardless of the scoring model used, a minimum of three scorers is necessary to score each student's test. However, to comply with a State requirement, none of the scorers assigned to score a student's test responses may be that student's teacher. This policy is detailed in the *School Administrator's Manual* section "Assigning Scorer Numbers and Questions to Scoring Committee Members" on page 21, found online at:

http://www.p12.nysed.gov/assessment/ei/2015/scoringleaderhandbook15.pdf.

Quality Control Process

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

Section 5: Operational Test Data Collection and Classical Analysis

Data Collection

Test data were collected in two phases. During Phase 1, a sample of approximately 95% of the student test records were received from the data warehouse and delivered to Pearson, beginning at the end of May 2015. During Phase 2, "straggler files" were submitted to Pearson in June 2015.

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

Data Processing

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

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

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

Sampling Down for Representativeness

Historically, after data cleaning, the sample is reviewed for representativeness of the prior year's operational population (i.e., all students testing in Spring 2014) in terms of key variables such as student gender, racial / ethnic identity, student disability status, English Language Learner (ELL) status, presence of test accommodation(s), and school Needs/Resource Capacity Category (NRC). In Spring 2015, a sampling down approach was adopted to make the sample used for equating as similar as possible to the previous year's testing population.

The numbers of cases dropped because of sampling down varied across grades and subjects, but the process for all grades was consistent. The cleaned data file for a given subject and grade was the starting point. Across the subjects and grades, the two areas where the 2015 cleaned data samples differed from the 2014 operational population were school NRC and student racial / ethnic identity. As such, based on the prior year's operational sample, sample weights were assigned to each combination of school NRC and student racial / ethnic identity for the 2015 cleaned data files and 20 students were removed from the sample. The current sample was then re-evaluated for its match to the 2014 operational sample in terms of percentages of students by school NRC and student racial / ethnic identity. If the revised sample still deviated by more than 5% on any school NRC or student racial / ethnic identity, an additional 20 students were dropped and the process was repeated until no school NRC or student racial / ethnic identity differed by more than 5% in the revised 2015 sample, when compared to the 2014 operational population.

The data cleaning procedures, including sampling down, and accompanying case counts are represented for ELA and mathematics in Tables 6A–6F and Tables 7A–7F, respectively. Comparison results between the final 2015 sample and 2014 operational population are further described in Section 6, "IRT Calibration and Equating."

Table 6A. ELA Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	169,613
Wrong Subject	0	169,613
No Grade	0	169,613
Wrong Grade	25	169,588
Language Mismatched Form	62	169,526
School Type	419	169,107
Missing Entire Book	1,043	168,064
Invalid Score	0	168,064
Out-of-Range CR Scores	0	168,064
Duplicated Record	6	168,058
Sampled Down	11,578	156,480

Table 6B. ELA Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	168,859
Wrong Subject	0	168,859
No Grade	0	168,859
Wrong Grade	22	168,837
Language Mismatched Form	34	168,803
School Type	451	168,352
Missing Entire Book	716	167,636
Invalid Score	0	167,636
Out-of-Range CR Scores	0	167,636
Duplicated Record	12	167,624
Sampled Down	15,744	151,880

Table 6C. ELA Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	162,686
Wrong Subject	0	162,686
No Grade	1	162,685
Wrong Grade	20	162,665
Language Mismatched Form	66	162,599
School Type	491	162,108
Missing Entire Book	570	161,538
Invalid Score	0	161,538
Out-of-Range CR Scores	0	161,538
Duplicated Record	10	161,528
Sampled Down	12,048	149,480

Table 6D. ELA Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	162,820
Wrong Subject	0	162,820
No Grade	0	162,820
Wrong Grade	24	162,796
Language Mismatched Form	65	162,731
School Type	546	162,185
Missing Entire Book	798	161,387
Invalid Score	0	161,387
Out-of-Range CR Scores	0	161,387
Duplicated Record	4	161,383
Sampled Down	12,583	148,800

Table 6E. ELA Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	155,862
Wrong Subject	0	155,862
No Grade	1	155,861
Wrong Grade	36	155,825
Language Mismatched Form	47	155,778
School Type	744	155,034
Missing Entire Book	1,250	153,784
Invalid Score	0	153,784
Out-of-Range CR Scores	0	153,784
Duplicated Record	4	153,780
Sampled Down	16,120	137,660

Table 6F. ELA Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	153,944
Wrong Subject	0	153,944
No Grade	1	153,943
Wrong Grade	33	153,910
Language Mismatched Form	57	153,853
School Type	948	152,905
Missing Entire Book	1,367	151,538
Invalid Score	0	151,538
Out-of-Range CR Scores	0	151,538
Duplicated Record	6	151,532
Sampled Down	18,352	133,180

Table 7A. Mathematics Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	169,056
Wrong Subject	0	169,056
No Grade	0	169,056
Wrong Grade	23	169,033
Language Mismatched Form	24	169,009
School Type	408	168,601
Missing Entire Book	259	168,342
Invalid Score	0	168,342
Out-of-Range CR Scores	0	168,342
Duplicated Record	6	168,336
Sampled Down	13,136	155,200

Table 7B. Mathematics Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	167,998
Wrong Subject	0	167,998
No Grade	0	167,998
Wrong Grade	19	167,979
Language Mismatched Form	36	167,943
School Type	423	167,520
Missing Entire Book	212	167,308
Invalid Score	0	167,308
Out-of-Range CR Scores	0	167,308
Duplicated Record	18	167,290
Sampled Down	18,490	148,800

Table 7C. Mathematics Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	159,943
Wrong Subject	0	159,943
No Grade	0	159,943
Wrong Grade	21	159,922
Language Mismatched Form	23	159,899
School Type	467	159,432
Missing Entire Book	297	159,135
Invalid Score	0	159,135
Out-of-Range CR Scores	0	159,135
Duplicated Record	10	159,125
Sampled Down	15,205	143,920

Table 7D. Mathematics Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	158,160
Wrong Subject	0	158,160
No Grade	0	158,160
Wrong Grade	23	158,137
Language Mismatched Form	79	158,058
School Type	499	157,559
Missing Entire Book	315	157,244
Invalid Score	0	157,244
Out-of-Range CR Scores	0	157,244
Duplicated Record	6	157,238
Sampled Down	15,358	141,880

Table 7E. Mathematics Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	147,893
Wrong Subject	0	147,893
No Grade	0	147,893
Wrong Grade	30	147,863
Language Mismatched Form	153	147,710
School Type	633	147,077
Missing Entire Book	577	146,500
Invalid Score	0	146,500
Out-of-Range CR Scores	0	146,500
Duplicated Record	4	146,496
Sampled Down	20,236	126,260

Table 7F. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	116,469
Wrong Subject	0	116,469
No Grade	0	116,469
Wrong Grade	30	116,439
Language Mismatched Form	51	116,388
School Type	782	115,606
Missing Entire Book	658	114,948
Invalid Score	0	114,948
Out-of-Range CR Scores	0	114,948
Duplicated Record	6	114,942
Sampled Down	19,662	95,280

Classical Analysis and Calibration Sample Characteristics

The cleaned and sampled-down data sets included over 96% of New York State students, and were used for classical analyses as well as calibration and equating. The demographic characteristics of students in these data sets are presented in the following tables, with the ELA tables provided first (Tables 8A–8F), followed by the mathematics tables (Tables 9A–9F). The Needs/Resource Capacity Category (NRC) is assigned at the district level, and is an indicator of district and school socioeconomic status. The ethnicity and gender designations are based on student-level information.

Table 8A. ELA Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,902	49.14
	Male	79,578	50.86
Ethnicity	Asian	14,868	9.50
	Black	30,246	19.33
	Hispanic	43,405	27.74
	American Indian	959	0.61
	Multiracial	3,098	1.98
	Pacific Islander	421	0.27
	White	63,483	40.57
NRC	New York	62,276	39.84
	Big 4 Cities	6,785	4.34
	Urban/Suburban	12,765	8.17
	Rural	8,581	5.49
	Average Needs	38,197	24.44
	Low Needs	17,416	11.14
	Charter	7,298	4.67
	Non-Public	2,997	1.92
SWD	No	132,970	84.98
	Yes	23,510	15.02
SUA	No	138,262	88.36
	Yes	18,218	11.64
ELL	No	142,967	91.36
	Yes	13,513	8.64

Note. The total n-count was 156,480.

Table 8B. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	74,667	49.16
	Male	77,213	50.84
Ethnicity	Asian	14,713	9.69
	Black	29,000	19.09
	Hispanic	40,129	26.42
	American Indian	883	0.58
	Multiracial	2,438	1.61
	Pacific Islander	346	0.23
	White	64,371	42.38
NRC	New York	57,551	37.93
	Big 4 Cities	6,432	4.24
	Urban/Suburban	11,799	7.78
	Rural	8,119	5.35
	Average Needs	37,550	24.74
	Low Needs	17,506	11.54
	Charter	5,876	3.87
	Non-Public	6,916	4.56
SWD	No	128,124	84.36
	Yes	23,756	15.64
SUA	No	133,887	88.15
	Yes	17,993	11.85
ELL	No	140,317	92.39
	Yes	11,563	7.61

Note. The total n-count was 151,880.

Table 8C. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	73,442	49.13
	Male	76,038	50.87
Ethnicity	Asian	14,376	9.62
	Black	28,940	19.36
	Hispanic	39,110	26.16
	American Indian	867	0.58
	Multiracial	2,110	1.41
	Pacific Islander	331	0.22
	White	63,746	42.65
NRC	New York	56,927	38.12
	Big 4 Cities	6,425	4.30
	Urban/Suburban	11,895	7.97
	Rural	7,954	5.33
	Average Needs	37,665	25.22
	Low Needs	18,730	12.54
	Charter	6,866	4.60
	Non-Public	2,870	1.92
SWD	No	124,470	83.27
	Yes	25,010	16.73
SUA	No	130,831	87.52
	Yes	18,649	12.48
ELL	No	138,613	92.73
	Yes	10,867	7.27

Note. The total n-count was 149,480.

Table 8D. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	72,832	48.95
	Male	75,968	51.05
Ethnicity	Asian	14,117	9.49
	Black	29,822	20.04
	Hispanic	38,772	26.06
	American Indian	798	0.54
	Multiracial	1,952	1.31
	Pacific Islander	327	0.22
	White	63,012	42.35
NRC	New York	55,019	37.01
	Big 4 Cities	6,193	4.17
	Urban/Suburban	11,437	7.69
	Rural	7,747	5.21
	Average Needs	36,332	24.44
	Low Needs	17,982	12.10
	Charter	6,856	4.61
	Non-Public	7,075	4.76
SWD	No	124,959	83.98
	Yes	23,841	16.02
SUA	No	131,588	88.43
	Yes	17,212	11.57
ELL	No	139,308	93.62
	Yes	9,492	6.38

Note. The total n-count was 148,800.

Table 8E. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	67,182	48.80
	Male	70,478	51.20
Ethnicity	Asian	12,881	9.36
	Black	28,005	20.34
	Hispanic	35,180	25.56
	American Indian	799	0.58
	Multiracial	1,524	1.11
	Pacific Islander	294	0.21
	White	58,977	42.84
NRC	New York	52,766	38.36
	Big 4 Cities	5,880	4.28
	Urban/Suburban	10,852	7.89
	Rural	7,760	5.64
	Average Needs	33,707	24.51
	Low Needs	18,665	13.57
	Charter	5,104	3.71
	Non-Public	2,808	2.04
SWD	No	115,384	83.82
	Yes	22,276	16.18
SUA	No	122,226	88.79
	Yes	15,434	11.21
ELL	No	129,168	93.83
	Yes	8,492	6.17

Note. The total n-count was 137,660.

Table 8F. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	64,436	48.38
	Male	68,744	51.62
Ethnicity	Asian	13,096	9.83
	Black	27,114	20.36
	Hispanic	33,146	24.89
	American Indian	735	0.55
	Multiracial	1,404	1.05
	Pacific Islander	303	0.23
	White	57,382	43.09
NRC	New York	51,028	38.36
	Big 4 Cities	5,332	4.01
	Urban/Suburban	10,112	7.60
	Rural	7,583	5.70
	Average Needs	31,762	23.87
	Low Needs	17,672	13.28
	Charter	3,752	2.82
	Non-Public	5,794	4.36
SWD	No	112,605	84.55
	Yes	20,575	15.45
SUA	No	118,700	89.13
	Yes	14,480	10.87
ELL	No	125,760	94.43
	Yes	7,420	5.57

Note. The total n-count was 133,180.

Table 9A. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,023	48.98
	Male	79,177	51.02
Ethnicity	Asian	15,093	9.72
	Black	29,844	19.23
	Hispanic	43,592	28.09
	American Indian	944	0.61
	Multiracial	3,045	1.96
	Pacific Islander	421	0.27
	White	62,261	40.12
NRC	New York	62,433	40.27
	Big 4 Cities	6,822	4.40
	Urban/Suburban	12,659	8.16
	Rural	8,330	5.37
	Average Needs	37,473	24.17
	Low Needs	17,124	11.04
	Charter	7,180	4.63
	Non-Public	3,022	1.95
SWD	No	131,860	84.96
	Yes	23,340	15.04
SUA	No	136,606	88.02
	Yes	18,594	11.98
ELL	No	140,104	90.27
	Yes	15,096	9.73

Note. The total n-count was 155,200.

Table 9B. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	72,708	48.86
	Male	76,092	51.14
Ethnicity	Asian	14,737	9.90
	Black	28,296	19.02
	Hispanic	39,740	26.71
	American Indian	855	0.57
	Multiracial	2,375	1.60
	Pacific Islander	344	0.23
	White	62,453	41.97
NRC	New York	56,989	38.33
	Big 4 Cities	6,386	4.30
	Urban/Suburban	11,669	7.85
	Rural	7,756	5.22
	Average Needs	36,159	24.32
	Low Needs	17,170	11.55
	Charter	5,706	3.84
	Non-Public	6,846	4.60
SWD	No	125,654	84.44
	Yes	23,146	15.56
SUA	No	131,243	88.20
	Yes	17,557	11.80
ELL	No	135,849	91.30
	Yes	12,951	8.70

Note. The total n-count was 148,800.

Table 9C. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	70,206	48.78
	Male	73,714	51.22
Ethnicity	Asian	14,161	9.84
	Black	27,779	19.30
	Hispanic	38,145	26.50
	American Indian	804	0.56
	Multiracial	2,039	1.42
	Pacific Islander	329	0.23
	White	60,663	42.15
NRC	New York	55,601	38.67
	Big 4 Cities	6,267	4.36
	Urban/Suburban	11,415	7.94
	Rural	7,413	5.16
	Average Needs	35,645	24.79
	Low Needs	18,006	12.52
	Charter	6,616	4.60
	Non-Public	2,830	1.97
SWD	No	120,112	83.46
	Yes	23,808	16.54
SUA	No	126,124	87.63
	Yes	17,796	12.37
ELL	No	131,871	91.63
	Yes	12,049	8.37

Note. The total n-count was 143,920.

Table 9D. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	69,067	48.68
	Male	72,813	51.32
Ethnicity	Asian	13,859	9.77
	Black	28,525	20.11
	Hispanic	37,601	26.50
	American Indian	746	0.53
	Multiracial	1,819	1.28
	Pacific Islander	333	0.23
	White	58,997	41.58
NRC	New York	53,695	37.88
	Big 4 Cities	6,004	4.24
	Urban/Suburban	10,883	7.68
	Rural	7,297	5.15
	Average Needs	33,517	23.65
	Low Needs	16,919	11.94
	Charter	6,552	4.62
	Non-Public	6,875	4.85
SWD	No	119,650	84.33
	Yes	22,230	15.67
SUA	No	124,917	88.04
	Yes	16,963	11.96
ELL	No	131,015	92.34
	Yes	10,865	7.66

Note. The total n-count was 141,880.

Table 9E. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	61,061	48.36
	Male	65,199	51.64
Ethnicity	Asian	12,210	9.67
	Black	25,734	20.38
	Hispanic	33,018	26.15
	American Indian	696	0.55
	Multiracial	1,328	1.05
	Pacific Islander	279	0.22
	White	52,995	41.97
NRC	New York	50,074	39.69
	Big 4 Cities	5,443	4.31
	Urban/Suburban	9,932	7.87
	Rural	6,733	5.34
	Average Needs	29,819	23.64
	Low Needs	16,696	13.23
	Charter	4,753	3.77
	Non-Public	2,710	2.15
SWD	No	106,346	84.23
	Yes	19,914	15.77
SUA	No	111,703	88.47
	Yes	14,557	11.53
ELL	No	116,800	92.51
	Yes	9,460	7.49

Note. The total n-count was 126,260.

Table 9F. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	45,138	47.37
	Male	50,142	52.63
Ethnicity	Asian	8,010	8.41
	Black	21,956	23.04
	Hispanic	27,105	28.45
	American Indian	537	0.56
	Multiracial	955	1.00
	Pacific Islander	206	0.22
	White	36,511	38.32
NRC	New York	39,652	41.65
	Big 4 Cities	4,561	4.79
	Urban/Suburban	7,935	8.34
	Rural	5,586	5.87
	Average Needs	19,669	20.66
	Low Needs	9,257	9.72
	Charter	3,253	3.42
	Non-Public	5,284	5.55
SWD	No	78,018	81.88
	Yes	17,262	18.12
SUA	No	83,420	87.55
	Yes	11,860	12.45
ELL	No	87,326	91.65
	Yes	7,954	8.35

Note. The total n-count was 95,280.

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 (e.g., a printing error or two correct answers to one item), item analysis is the stage at which errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution). Analyses of test-level data comprise the second element of classical data analysis. These include examination of the raw score (RS) statistics (mean and standard deviation or "SD") and test reliability measures Cronbach's alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). Assessment of test speededness is another important element of classical analysis. Additionally, classical DIF analysis is conducted at this stage. DIF analysis includes computation of standardized mean differences and Mantel-Haenszel statistics for New York State items to identify potential item bias. All classical data analysis results contribute information on the validity and reliability of the tests (also see Section 3, "Validity," and Section 7, "Reliability and Standard Error of Measurement").

Item Difficulty and Point Biserial Correlation Coefficients

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

Item difficulties (p-values) on the ELA tests ranged from .26 to .89. For Grade 3, the item p-values ranged from .40 to .83, with a mean of .55. For Grade 4, the item p-values ranged from .26 to .81, with a mean of .57. For Grade 5, the item p-values ranged from .28 to .86, with a mean of .62. For Grade 6, the item p-values ranged from .36 to .89, with a mean of .59. For Grade 7, the item p-values ranged from .32 to .87, with a mean of .58. For Grade 8, the item p-values ranged from .29 to .89, with a mean of .61. These p-value statistics are in Appendix M Tables M1–M6, along with pbis statistics of the keys.

Item difficulties (p-values) on the Mathematics tests ranged from .19 to .96. For Grade 3, the item p-values ranged from .33 to .92, with a mean of .66. For Grade 4, the item p-values ranged from .36 to .89, with a mean of .63. For Grade 5, the item p-values ranged from .26 to .92, with a mean of .62. For Grade 6, the item p-values ranged from .27 to .89, with a mean of .57. For Grade 7, the item p-values ranged from .19 to .96, with a mean of .54. For Grade 8, the item p-values ranged from .28 to .91, with a mean of .52. These statistics are provided in Appendix M Tables M7–M12, along with other classical test summary statistics.

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

Point biserial correlations are presented in Appendix M Tables M1–M12. The column labeled "Pbis Key" contains the point biserial correlation associated with the correct response. The guideline for building the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests was that the pbis correlation for the key for MC items should be equal to or greater than .20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. There were very few exceptions to this guideline, due to content considerations, which required the inclusion of particular items. Decisions to use such items were made very carefully, and no item with a negative point-biserial correlation was allowed on the test.

Point biserials for correct answer options on the ELA tests ranged from .12 to .56, as shown in Appendix M in Tables M1–M6. For Grade 3, the item pbis values were between .14 and .50, with a mean of .37. For Grade 4, the item pbis values were between .12 and .47, with a mean of

.36. For Grade 5, the item pbis values were between .21 and .54, with a mean of .38. For Grade 6, the item pbis values were between .15 and .53, with a mean of .36. For Grade 7, the item pbis values were between .20 and .54, with a mean of .36. For Grade 8, the item pbis values were between .13 and .56, with a mean of .39.

Point biserials for correct answer options on the Mathematics tests ranged from .19 to .64, as shown in Appendix M in Tables M7–M12. For Grade 3, the item pbis values were between .34 and .62, with a mean of .46. For Grade 4, the item pbis values were between .33 and .64, with a mean of .50. For Grade 5, the item pbis values were between .29 and .59, with a mean of .44. For Grade 6, the item pbis values were between .29 and .60, with a mean of .47. For Grade 7, the item pbis values were between .19 and .62, with a mean of .44. For Grade 8, the item pbis values were between .23 and .54, with a mean of .42.

Speededness

Speededness refers to interference in test scores due to insufficient testing time. It is NYSED policy that ample testing time should be given for students to complete the entire test. Furthermore, both the validity (i.e., accuracy) and reliability (i.e., precision) of test scores are adversely affected when tests are speeded. For these reasons, sufficient administration time limits were set for the NYSTP tests.

Speededness is routinely checked, based on test data, after each administration. One method of analyzing data to determine if speededness has occurred is to review the proportion of students not answering (i.e., omitting) items, especially those items that appeared towards the end of the test form. Tables M1–M12 in Appendix M show the omit rates for items on the Grades 3–8 Common Core ELA and Mathematics Tests, respectively. The industry standard general rule of thumb is that omit rates for multiple-choice items should be less than 5.0%. Omit rates across multiple-choice and constructed-response items on the Grades 3–8 Common Core ELA and Mathematics Tests typically ranged from 0% to 3%. As may be expected, omit rates tended to increase for items at the end of the test booklets. That is, these omit rates remained within the acceptable range for large-scale achievement tests. In summary, the low omit rates observed across entire forms are consistent with tests that are not speeded.

More detailed approaches to check for speededness include examining the relationships of test scores between test books that measure similar content and student performance on individual test books. Beyond omit rates, a test that is not speeded should show empirical relationships between and across all test books. In other words, students performing well on multiple-choice items in Book 1 would also be expected to perform well on multiple-choice items in Book 2. In the presence of speededness, scores on books measuring similar content would exhibit low correlations.

Correlation analysis was conducted to compare the relationship between student performance on each book with student performance on the remaining books. The results are presented in Tables 10A–10B and it can be seen that the correlation coefficients were sufficiently high and consistent across books for both ELA and Mathematics Tests. The patterns in both ELA and mathematics reflect what would be expected for tests that do not exhibit speededness.

Table 10A. ELA Correlations (across Books)

Grade	Correlation with Other Books		
	Book 1	Book 2	Book 3
3	.72	.77	.72
4	.74	.78	.76
5	.77	.80	.76
6	.76	.81	.76
7	.73	.80	.73
8	.75	.81	.74

Table 10B. Mathematics Correlations (across Books)

Grade	Correlation with Other Books		
	Book 1	Book 2	Book 3
3	.85	.84	.84
4	.87	.89	.88
5	.86	.87	.87
6	.89	.89	.89
7	.85	.87	.87
8	.85	.86	.86

Next, correlation analysis of student performance was performed on items from each of the books (i.e., book by book) containing similar item types. The results of this analysis set, as presented in Tables 11A–11B, were similar to those from the previous analysis: the correlations between individual books were positive and moderately strong. Also, no strong evidence of speededness was observed for ELA Books 2 and 3.

Table 11A. ELA Correlations (Book by Book)

Grade	Book	Book 1	Book 2
3	2	.69	
	3	.63	.72
4	2	.70	
	3	.68	.73
5	2	.74	
	3	.69	.75
6	2	.74	
	3	.69	.77
7	2	.72	
	3	.63	.76
8	2	.75	
	3	.66	.76

Table 11B. Mathematics Correlations (Book by Book)

Grade	Book	Book 1	Book 2
3	2	.81	
	3	.80	.80
4	2	.84	
	3	.84	.86
5	2	.83	
	3	.82	.84
6	2	.85	
	3	.86	.86
7	2	.82	
	3	.81	.84
8	2	.81	
	3	.81	.82

Differential Item Functioning

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 which may later be judged by content experts to exhibit bias, rather than directly detecting bias. First, the psychometric phenomenon of DIF was extensively investigated and experts' judgments of bias collected when items were field-tested, which reduced the likelihood of including any differentially functioning items on the operational forms for 2015. Turning to the analysis of the 2015 operational data, as discussed in the "Detection of Bias" subsection of Section 3 above, items flagged for DIF do not necessarily indicate item bias. For example, DIF may be attributed to true group differences on the content measured by the item or Type I error, which refers to statistically flagging items that have no true DIF. Operational items flagged for DIF are given additional scrutiny by content specialists, above and beyond the existing rounds of reviews by New York State educators, and those content specialists make the final judgment of whether or not an item is biased for or against the focal group.

DIF was evaluated using two methods, both of which involve checks on statistical and practical significance. First, the Mantel-Haenszel (MH) method is employed for MC items. This non-parametric DIF method partitions the sample of examinees into categories based on total raw test scores. It then compares the log-odds ratio of keyed responses for the focal and reference groups. In terms of statistical significance, the Mantel-Haenszel method has a critical value of 6.63 (degrees of freedom = 1 for MC items; alpha = .01) and as far as practical significance is concerned, it is compared to its corresponding delta-value. Delta-values are a commonly used metric in testing that indicates the magnitude of DIF. Typically, delta-values above 1.50 are considered indicative of moderate DIF that should be examined more closely (Zwick, Donoghue, and Grima, 1993). Second, the standardized mean difference (SMD) was computed for CR items. The SMD statistic (Dorans, Schmitt, and Bleistein, 1992) compares the mean scores of reference and focal groups, after adjusting for proficiency differences. The SMD was also

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

Classical DIF analyses were conducted on subgroups of the Needs/Resource Capacity Category (focal group: High Needs; reference group: Low Needs), gender (focal group: Female; reference group: Male), ethnicity (focal groups: Black, Hispanic, and Asian; reference group: White), and English language learners (focal group: English language learners; reference group: Non-English language learners). The DIF analyses were conducted using all cases from the clean data sets. Table 12 and Table 13 show the numbers of cases for the subgroups for ELA and mathematics, respectively.

Table 12. ELA Classical DIF Sample N-Counts

Grade	Ethnicity			Gender		Needs/Resource Capacity Category	
	Black/African American	Hispanic/ Latino	White	Female	Male	High	Low
3	30,246	43,405	63,483	76,902	79,578	90,407	55,613
4	29,000	40,129	64,371	74,667	77,213	83,901	55,056
5	28,940	39,110	63,746	73,442	76,038	83,201	56,395
6	29,822	38,772	63,012	72,832	75,968	80,396	54,314
7	28,005	35,180	58,977	67,182	70,478	77,258	52,372
8	27,114	33,146	57,382	64,436	68,744	74,055	49,434

Table 13. Mathematics Classical DIF Sample N-Counts

Grade	Ethnicity			Gender		Needs/Resource Capacity Category	
	Black/African American	Hispanic/ Latino	White	Female	Male	High	Low
3	29,844	43,592	62,261	76,023	79,177	90,244	54,597
4	28,296	39,740	62,453	72,708	76,092	82,800	53,329
5	27,779	38,145	60,663	70,206	73,714	80,696	53,651
6	28,525	37,601	58,997	69,067	72,813	77,879	50,436
7	25,734	33,018	52,995	61,061	65,199	72,182	46,515
8	21,956	27,105	36,511	45,138	50,142	57,734	28,926

Table 14 (ELA) and Table 15 (mathematics) present the number of items flagged for DIF by either of the classical methods described earlier. A detailed list of items flagged by either one or both of these classical DIF methods, including DIF direction and associated DIF statistics, is presented in Appendix N.

Table 14. ELA Items Flagged for DIF

Grade	Flagged Items
3	1
4	2
5	4
6	0
7	8
8	2

Table 15. Mathematics Items Flagged for DIF

Grade	Flagged Items
3	2
4	4
5	1
6	2
7	3
8	2

As discussed previously (Section 3, Validity), items showing statistically significant DIF (flagged as described above for MH statistics on MC items and SMD statistics for CR items) do not necessarily pose bias. The items flagged with DIF were examined by the content experts again, and no sign of potential bias was found. In other words, based on combinations of statistical and content evaluations, none of the items on the 3–8 tests showed bias.

Section 6: IRT Calibration and Equating

IRT Models and Rationale for Use

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

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

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

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

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

where

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

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

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

where

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

and

k is the item response category (k = 1, 2, …. m_j).

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

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

and

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

where

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

and

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

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

Calibration Sample

The cleaned and sampled-down data were used for calibration and equating of the NYSTP 2015 Grades 3–8 Common Core ELA and Mathematics Tests. It should be noted that the sample sizes were adequate, as the calibration and equating were performed using nearly all (96–99%, depending on grade level) of the New York State public and non-public school student population data in each tested grade. As shown in Tables 16A. ELA Grades 3 and 4 Demographic Statistics–16C and Tables 17A–17C, the 2015 operational test samples were generally comparable to 2014 populations in terms of NRC, student race and ethnicity, proportions of English language learners, proportions of students with disabilities, and proportions of students using testing accommodations.

Table 16A. ELA Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2014 Population	2015 Sample	2014 Population	2015 Sample
Gender	Female	49.00	49.14	49.23	49.16
	Male	51.00	50.86	50.77	50.84
Ethnicity	Asian	8.58	9.50	8.77	9.69
	Black	17.81	19.33	17.53	19.09
	Hispanic	25.46	27.74	24.19	26.42
	American Indian	0.58	0.61	0.55	0.58
	Multiracial	1.83	1.98	1.47	1.61
	Pacific Islander	0.25	0.27	0.22	0.23
	White	45.48	40.57	47.27	42.38
NRC	New York	35.53	39.84	33.84	37.93
	Big 4 Cities	3.90	4.34	3.79	4.24
	Urban/Suburban	7.58	8.17	7.14	7.78
	High Needs Rural	5.25	5.49	5.14	5.35
	Average Needs	25.92	24.44	25.82	24.74
	Low Needs	12.19	11.14	12.74	11.54
	Charter	4.16	4.67	3.45	3.87
	Non-Public	5.36	1.92	7.99	4.56
SWD	No	84.97	84.98	84.59	84.98
	Yes	15.03	15.02	15.41	15.02
SUA	No	87.62	88.36	87.69	88.36
	Yes	12.38	11.64	12.31	11.64
ELL	No	92.12	91.36	92.00	91.36
	Yes	7.88	8.64	8.00	8.64

Table 16B. ELA Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2014 Population	2015 Sample	2014 Population	2015 Sample
Gender	Female	49.09	49.13	49.29	48.95
	Male	50.91	50.87	50.71	51.05
Ethnicity	Asian	8.58	9.62	8.42	9.49
	Black	17.79	19.36	18.51	20.04
	Hispanic	23.96	26.16	23.92	26.06
	American Indian	0.57	0.58	0.56	0.54
	Multiracial	1.30	1.41	1.18	1.31
	Pacific Islander	0.23	0.22	0.20	0.22
	White	47.57	42.65	47.22	42.35
NRC	New York	33.58	38.12	32.57	37.01
	Big 4 Cities	3.83	4.30	3.73	4.17
	Urban/Suburban	7.27	7.97	7.10	7.69
	High Needs Rural	5.29	5.33	5.29	5.21
	Average Needs	27.19	25.22	26.48	24.44
	Low Needs	13.32	12.54	13.12	12.10
	Charter	4.05	4.60	4.05	4.61
	Non-Public	5.36	1.92	7.56	4.76
SWD	No	83.82	83.27	84.20	83.27
	Yes	16.18	16.73	15.80	16.73
SUA	No	87.03	87.52	87.34	87.52
	Yes	12.97	12.48	12.66	12.48
ELL	No	93.38	92.73	93.35	92.73
	Yes	6.62	7.27	6.65	7.27

Table 16C. ELA Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2014 Population	2015 Sample	2014 Population	2015 Sample
Gender	Female	48.96	48.80	48.82	48.38
	Male	51.04	51.20	51.18	51.62
Ethnicity	Asian	8.28	9.36	8.69	9.83
	Black	18.69	20.34	18.72	20.36
	Hispanic	23.45	25.56	23.09	24.89
	American Indian	0.56	0.58	0.53	0.55
	Multiracial	1.04	1.11	0.97	1.05
	Pacific Islander	0.19	0.21	0.21	0.23
	White	47.78	42.84	47.79	43.09
NRC	New York	33.43	38.36	33.55	38.36
	Big 4 Cities	3.80	4.28	3.56	4.01
	Urban/Suburban	7.22	7.89	6.85	7.60
	High Needs Rural	5.61	5.64	5.35	5.70
	Average Needs	27.01	24.51	26.30	23.87
	Low Needs	14.15	13.57	14.04	13.28
	Charter	3.23	3.71	2.46	2.82
	Non-Public	5.43	2.04	7.78	4.36
SWD	No	84.07	83.82	85.03	83.82
	Yes	15.93	16.18	14.97	16.18
SUA	No	88.04	88.79	88.11	88.79
	Yes	11.96	11.21	11.89	11.21
ELL	No	94.49	93.83	94.75	93.83
	Yes	5.51	6.17	5.25	6.17

Table 17A. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2014 Population	2015 Sample	2014 Population	2015 Sample
Gender	Female	48.88	48.98	49.14	48.86
	Male	51.12	51.02	50.86	51.14
Ethnicity	Asian	8.80	9.72	8.97	9.90
	Black	17.77	19.23	17.50	19.02
	Hispanic	25.84	28.09	24.50	26.71
	American Indian	0.58	0.61	0.55	0.57
	Multiracial	1.80	1.96	1.45	1.60
	Pacific Islander	0.26	0.27	0.23	0.23
	White	44.94	40.12	46.80	41.97
NRC	New York	36.04	40.27	34.28	38.33
	Big 4 Cities	3.95	4.40	3.85	4.30
	Urban/Suburban	7.60	8.16	7.17	7.85
	High Needs Rural	5.17	5.37	5.07	5.22
	Average Needs	25.55	24.17	25.46	24.32
	Low Needs	12.08	11.04	12.62	11.55
	Charter	4.14	4.63	3.43	3.84
	Non-Public	5.37	1.95	8.01	4.60
SWD	No	85.11	84.98	84.74	84.98
	Yes	14.89	15.02	15.26	15.02
SUA	No	85.54	88.36	84.37	88.36
	Yes	14.46	11.64	15.63	11.64
ELL	No	91.16	91.36	91.06	91.36
	Yes	8.84	8.64	8.94	8.64

Table 17B. Mathematics Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2014 Population	2015 Sample	2014 Population	2015 Sample
Gender	Female	49.02	48.78	49.11	48.68
	Male	50.98	51.22	50.89	51.32
Ethnicity	Asian	8.81	9.84	8.68	9.77
	Black	17.79	19.30	18.56	20.11
	Hispanic	24.38	26.50	24.43	26.50
	American Indian	0.57	0.56	0.55	0.53
	Multiracial	1.29	1.42	1.16	1.28
	Pacific Islander	0.23	0.23	0.22	0.23
	White	46.93	42.15	46.41	41.58
NRC	New York	34.19	38.67	33.39	37.88
	Big 4 Cities	3.87	4.36	3.80	4.24
	Urban/Suburban	7.30	7.94	7.13	7.68
	High Needs Rural	5.21	5.16	5.20	5.15
	Average Needs	26.59	24.79	25.70	23.65
	Low Needs	13.16	12.52	12.92	11.94
	Charter	4.06	4.60	4.07	4.62
	Non-Public	5.51	1.97	7.69	4.85
SWD	No	84.01	83.27	84.44	83.27
	Yes	15.99	16.73	15.56	16.73
SUA	No	85.45	87.52	85.76	87.52
	Yes	14.55	12.48	14.24	12.48
ELL	No	92.39	92.73	92.34	92.73
	Yes	7.61	7.27	7.66	7.27

Table 17C. Mathematics Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2014 Population	2015 Sample	2014 Population	2015 Sample
Gender	Female	48.78	48.36	47.96	47.37
	Male	51.22	51.64	52.04	52.63
Ethnicity	Asian	8.57	9.67	7.48	8.41
	Black	18.89	20.38	21.38	23.04
	Hispanic	24.13	26.15	26.63	28.45
	American Indian	0.55	0.55	0.56	0.56
	Multiracial	1.01	1.05	0.92	1.00
	Pacific Islander	0.20	0.22	0.20	0.22
	White	46.65	41.97	42.82	38.32
NRC	New York	34.71	39.69	36.85	41.65
	Big 4 Cities	3.88	4.31	4.31	4.79
	Urban/Suburban	7.29	7.87	7.67	8.34
	High Needs Rural	5.52	5.34	5.63	5.87
	Average Needs	26.00	23.64	23.12	20.66
	Low Needs	13.61	13.23	9.97	9.72
	Charter	3.29	3.77	3.01	3.42
	Non-Public	5.58	2.15	9.34	5.55
SWD	No	84.26	83.82	82.30	83.82
	Yes	15.74	16.18	17.70	16.18
SUA	No	87.38	88.79	87.10	88.79
	Yes	12.62	11.21	12.90	11.21
ELL	No	93.37	93.83	92.18	93.83
	Yes	6.63	6.17	7.82	6.17

Calibration Process

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

The calibration of NYSTP 2015 Grades 3–8 Common Core ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were on the original theta scale, and all of the items were well within the prescribed parameter ranges. For both the Grades 3–8 Common Core ELA and Mathematics Tests, all calibration estimation results were reasonable. The summaries of the calibration results are presented in Table 18 for ELA and Table 19 for mathematics. Additional details, including individual item parameter estimates, may be found in Appendix O, in Tables O13–O24. The parameter estimates are expressed on the theta metric and are defined below:

- for MC items:
    - *a*-parameter is a discrimination parameter;
    - *b*-parameter is a difficulty parameter; and
    - *c*-parameter is a guessing parameter; and
- for CR items:
    - *alpha* is a discrimination parameter; and
    - *step* is a difficulty parameter for category m_j.

As described in Section 6 "IRT Calibration and Equating," above in subsection "IRT Models and Rationale for Use," m_j denotes the number of score levels for the *j*th item, and, typically, the highest score level is assigned (m_j - 1) score points. Note that for the 2PPC model there are m_j - 1 independent steps and one alpha, for a total of m_j independent parameters estimated for each item, while there is one *a*-parameter and one *b*-parameter per item in the 3PL model.

Table 18. ELA Calibration Results

Grade	Item-level			Student-Level		
	Largest a-Parameter	Range of b- / Step Parameters		n	Theta Est.*	
					Mean	SD
3	1.135	-1.934	1.135	156,480	0.01	0.94
4	1.199	-1.990	3.403	151,880	0.01	0.94
5	1.283	-2.082	1.832	149,480	0.00	0.95
6	1.102	-1.884	2.096	148,800	0.00	0.95
7	1.374	-1.850	2.124	137,660	0.01	0.95
8	1.378	-3.119	1.656	133,180	0.00	0.94

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

Table 19. Mathematics Calibration Results

Grade	Item-level			Student-Level		
	Largest a-Parameter	Range of b- / Step Parameters		n	Theta Est.*	
					Mean	SD
3	1.621	-2.227	1.621	155,200	0.00	0.94
4	1.472	-1.512	1.472	148,800	0.00	0.93
5	1.940	-2.023	1.940	143,920	0.01	0.94
6	1.968	-1.580	1.968	141,880	0.02	0.94
7	2.397	-2.578	2.397	126,260	0.02	0.94
8	1.795	-2.282	1.795	95,280	0.03	0.92

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

Item-Model Fit

Item fit statistics provide evidence of the appropriateness of using an item in the 3PL or 2PPC model. The Q_1 procedure described by Yen (1981) was used to measure fit to the three-parameter model. Students are rank-ordered on the basis of $\hat{\theta}$ values and sorted into ten cells with 10% of the sample in each cell. For each item, the number of students in cell k who answered item i, N_{ik}, and the number of students in that cell who answered item i correctly, R_{ik}, were determined. The observed proportion in cell k passing item i, O_{ik}, is R_{ik}/N_{ik}. The fit index for item i is:

$$Q_{1i} = \sum_{k=1}^{10} \frac{N_{ik}(O_{ik} - E_{ik})^2}{E_{ik}(1 - E_{ik})},$$

with:

$$E_{ik} = \frac{1}{N_{ik}} \sum_{j \varepsilon \text{ cell } k}^{N_{ik}} P_i(\hat{\theta}_j).$$

A modification of this procedure was used to measure fit to the 2PPC model. For the 2PPC model, Q_{1j} was assumed to have an approximate chi-square distribution with the following degrees of freedom (*df*):

$$df = I(m_j - 1) - m_j,$$

where I is the total number of cells (usually 10) and m_j is the possible number of score levels for item j.

To adjust for differences in degrees of freedom among items, Q_1 was transformed to Z_{Q_1} where:

$$\mathrm{Z}_{Q_1} = (Q_1 - df)/(2df)^{1/2}.$$

The value of Z increases with sample size, when all else is equal. To use this standardized statistic to flag items for potential poor fit, it has been a common practice to vary the critical value for Z as a function of sample size. For the tests that have large calibration sample sizes, the criterion $\mathrm{Z}_{Q_1}Crit$ was used to flag items and was calculated using the expression

$$\mathrm{Z}_{Q_1}Crit = \left(\frac{N}{1500}\right) * 4,$$

where N is the calibration sample size.

To compute the Q_1 and related statistics, a stratified sampling procedure was implemented in a way that a representative sample with the size of approximately 70,000 students was drawn at each grade level. Items were considered to have poor fit if the value of the obtained Z_{Q_1} was

greater than the value of Z_{Q1} critical. If the obtained Z_{Q1} was less than Z_{Q1} critical, the items were rated as having acceptable fit. The fact that the majority of the items in the NYSTP 2015 Grades 3–8 Common Core ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models. Item fit statistics are presented in Appendix O, in Tables O1–O12.

Local Independence

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

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

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

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

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

The generalization to items with multiple response categories uses

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

where

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

If a substantial number of items in the test demonstrate local dependence, these items may need to be calibrated separately. All pairs of items with Q_3 values greater than .20 were classified as significant for local dependency. The maximum value for this index is 1.00. When item pairs are flagged by Q_3, the content of the flagged items is examined to identify possible sources of the local dependence. The primary concern about locally dependent items is that they contribute less psychometric information about examinee proficiency than do locally independent items, and therefore inflate score reliability estimates.

The Q_3 statistics were examined for all unique pairs of ELA and mathematics items. Items that were found to be significant in local dependency vary, depending on the subject and grade: two pairs of items each were found in ELA Grades 7 and 8; and one pair of items was found in Grade

5. When reviewing the results for mathematics, two pairs of items exceeded a correlation of .20 in mathematics Grades 3 and 4 and one pair of items each in Grades 5, 6, and 7. The magnitudes of these statistics were not sufficient to warrant further concern or action (with the Q_3 values ranging from .20 to .25 for ELA tests and from .20 to .25, with one value of .51 for mathematics tests).

Equating and Scaling

With the new assessments being implemented in 2013, a new scale was established after the data were collected. The purpose of equating was to place the 2015 item parameters and proficiency estimates on the same scale as those in 2014. The following steps constitute the equating process for each subject and grade:

1. Operational items as well as non-scored (i.e., external) anchor items were calibrated in IRTPRO.

2. The 2015 item parameter estimates for all anchor items—both scored and non-scored—enabled the establishment of the equating relationship via a test characteristic curve (TCC) method (Stocking and Lord, 1983; implemented in STUIRT, Kim, & Kolen, 2004) to the 2014 theta scale, using the established 2014 item parameter estimates for those same items. The resulting equating coefficients are displayed in Table 20 and Table 21, and the following parameters were equated using the formulas below:

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

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

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

where

M_1^E is defined as the multiplicative adjustment for equating and M_2^E is the additive adjustment for equating. Also note that the superscript "*E*" denotes equated item parameter estimates, while the superscript "*C*" denotes calibrated item parameter estimates.

Table 20. ELA Equating Coefficients

Grade	M_1^E	M_2^E
3	1.047	-0.023
4	1.053	0.017
5	1.047	-0.063
6	1.009	-0.010
7	1.003	-0.146
8	1.037	-0.021

Table 21. Mathematics Equating Coefficients

Grade	M_1^E	M_2^E
3	1.075	0.158
4	1.148	0.148
5	1.124	0.251
6	1.138	0.149
7	1.133	0.141
8	1.082	-0.183

3. A raw-score-to-theta conversion chart was produced using the test characteristic curve (TCC) method (Stocking and Lord, 1983; see the "Scoring Procedure
4. " subsection for more detail) and implemented in POLYEQUATE (Kolen & Cui, 2004). The theta estimates associated with the TCC method ($\hat{\theta}_{TCC}$) must be equated back to the underlying theta scale established in the prior year (Spring 2014), and are computed as follows:

$$\theta^E = \left(M_1^E \cdot \hat{\theta}_{TCC}\right) + M_2^E$$

5. The TCC method does not produce theta estimates for raw scores below chance level or above the perfect score (highest obtainable raw score). In addition, for the scores at the low and high ends of the scale, some raw scores tended to have large theta estimates (for example, -7.999). Typically, the first obtainable theta value on a test corresponds to a very extreme theta value. The following adjustment/interpolation was conducted:

   For any equated theta estimates (θ^E) that are outside of the range of -2.5 to 3, at the lower end of the scale, 0.25 was subtracted from the preceding theta value that is within the range; at the higher end of the scale, 0.25 was added to the previous theta value that is within the range, thus resulting in an adjusted theta estimate (θ^A) for those extremes. See the table below for an example at the lower end of the scale. Such an adjustment helps contain the theta scale within a reasonable range, and is standard practice in testing.

Raw score	θ^E	θ^A
6	-5.30263	-3.37458
7	-3.66491	-3.12458
8	-3.03055	-2.87458
9	-2.76782	-2.62458
10	-2.37458	-2.37458

6. Once theta values were either estimated or interpolated for all raw scores, the raw-score-to-theta relationship was applied to each student, yielding a theta estimate corresponding to his or her raw score.

7. The adjusted theta estimates were then scaled using the established scaling coefficients from the prior year (Spring 2014) and displayed in Table 22 and Table 23 according to the following formula:

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

where

M_1^S is defined as the multiplicative scaling coefficient, and M_2^S is the additive scaling coefficient. M_1^S and M_2^S are applied to a true score (i.e., the post-equated theta estimate) in order to obtain a scale score.

Table 22. ELA Scaling Coefficients

Grade	M_1^S	M_2^S
3	31.814	301.495
4	32.036	300.762
5	32.016	300.954
6	32.258	300.673
7	31.926	300.801
8	31.627	300.980

Table 23. Mathematics Scaling Coefficients

Grade	M_1^S	M_2^S
3	32.249	299.856
4	32.698	300.176
5	32.220	300.693
6	32.421	300.377
7	31.229	301.144
8	31.868	301.143

8. Scale scores range, approximately, from 100 to 400 across grades. The lowest and highest observed scale score (LOSS and HOSS, respectively) may vary by grade.

9. A series of anchor set stability checks were performed before finalizing the anchor set for each subject and grade; see the "Anchor Set Evaluation" subsection, which follows this one.

10. For conditional standard error of measurement (CSEM), the scale scores (both estimated and interpolated) were used to compute the information function and CSEM.

Throughout this process, NYSED psychometricians have reviewed, and a senior scientist from HumRRO has independently verified, the results generated by Pearson psychometricians.

Anchor Set Evaluation

In order to determine if each item from the anchor set performs similarly to when it was administered in the prior year, comparisons of individual item characteristic curves (ICCs) and item parameter estimates from the previous and current administrations were made. Initial comparisons included a graphical inspection of the linearity of relationships between equated item parameter estimates from the 2014 and 2015 administrations. These revealed approximately linear relationships as well as similarities in item functions, and therefore provided support for the selected post-equating method used herein. Additional analyses of the correlations between equated item parameter estimates also provided evidence of strong linear relationships.

A formal process for validating the anchor set by using an objective criterion was used to determine if any items ought to be considered for removal from the anchor set. The equated item parameter estimates were used to calculate a weighted, squared deviation of the current ICC from the previous ICC, across the range of ability (i.e., theta, or θ) and under a hypothetical normal distribution for θ. For a given item i, that quantity, called "d squared," is given by

$$d_i^2 = \sum_k \left\{ \left[P_{i1}(\theta_k) - P_{i0}(\theta_k) \right]^2 \cdot g(\theta_k) \right\},$$

where i indexes anchor items; k indexes quadrature points for θ; $P_{i1}(\cdot)$ is the probability of a correct response to item i under the current calibration, while $P_{i0}(\cdot)$ is the same quantity under the previous calibration; and $g(\theta_k)$ are weights for the quadrature points.

Historically and as recently as the 2014 operational equating, a fixed criterion on this metric ($d_i^2 \geq 0.05$) has been used for flagging items to be considered for removal from equating. The same approach and criterion were used for the equating of the 2015 operational forms to the 2014 scale score scale. This procedure minimizes the weighted squared differences between the two ICCs for each MC item: one based on 2014 item parameter estimates and the other on 2015 estimates. The differential item performance was evaluated by examining previous and current item parameters. The following steps were taken:

1. Before the iterative procedures start, the initial equating was performed, using all of the eligible anchor items as an anchor set, as described in the "Equating and Scaling" section. The initial equating coefficients (M_1^E and M_2^E) were obtained through the Stocking-Lord method.

2. The following process was repeated for at least five iterations or until the largest $d_i^2 < 0.05$ is reached, whichever was greater:

a. For each anchor item, d_i^2 was calculated as a weighted sum of the squared deviations between the ICCs based on old (2014) and new (2015) parameter estimates at each quadrature point and assuming a normal theta distribution.
b. The item having the largest d_i^2 was identified and removed from the anchor set.
c. The equating procedures described in the "Equating and Scaling" subsection were performed with the newly reduced anchor set.
d. New raw-score-to-scale-score tables were prepared as described in the "Scoring Procedure" subsection.

3. Select the equating coefficients (M_1^E and M_2^E) associated with the iteration selected in step 2 above.

The items that are implicitly proposed for removal from the anchor set, based on the process described above, were summarized and evaluated. The only subject where items were proposed and ultimately approved for removal from the anchor set was mathematics, and one item each was removed from the anchor sets for Grades 5, 6, and 7.

Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2015 TCCs were generated using final item parameters for all reporting test items administered in Spring 2015. TCCs are the summation of all the item characteristic curves (ICCs) for items that contribute to the scale score. Conditional standard error of measurement (CSEM) curves graphically show the amount of measurement error at different proficiency levels. The TCCs and CSEM curves are presented in Figure 1 though Figure 24.

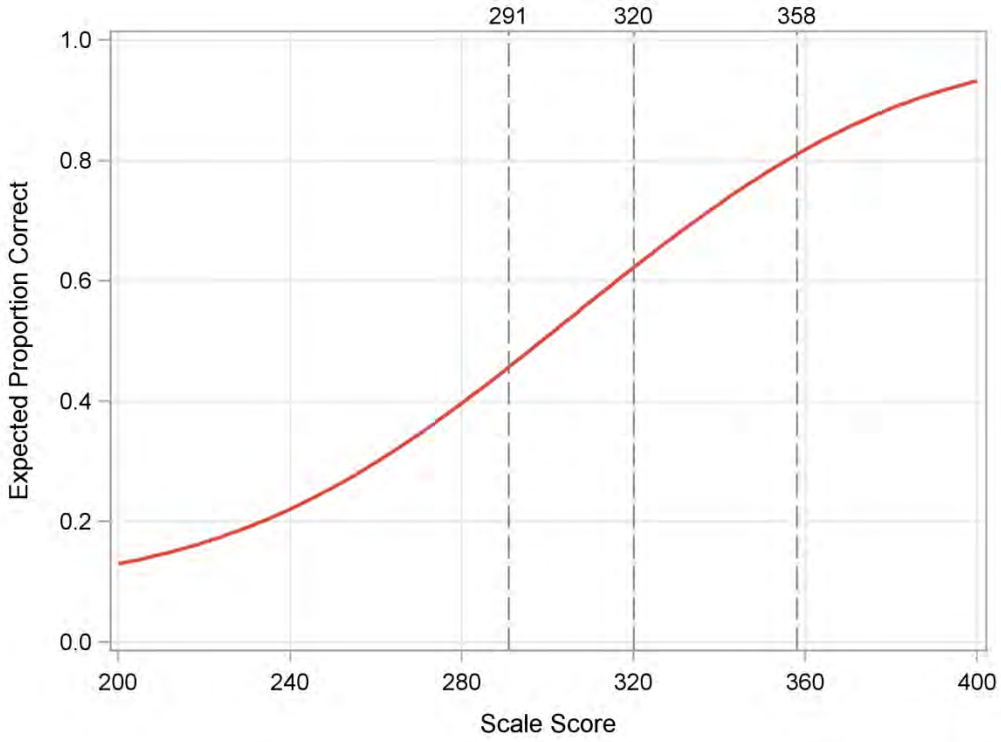

Figure 1. ELA Grade 3 TCC

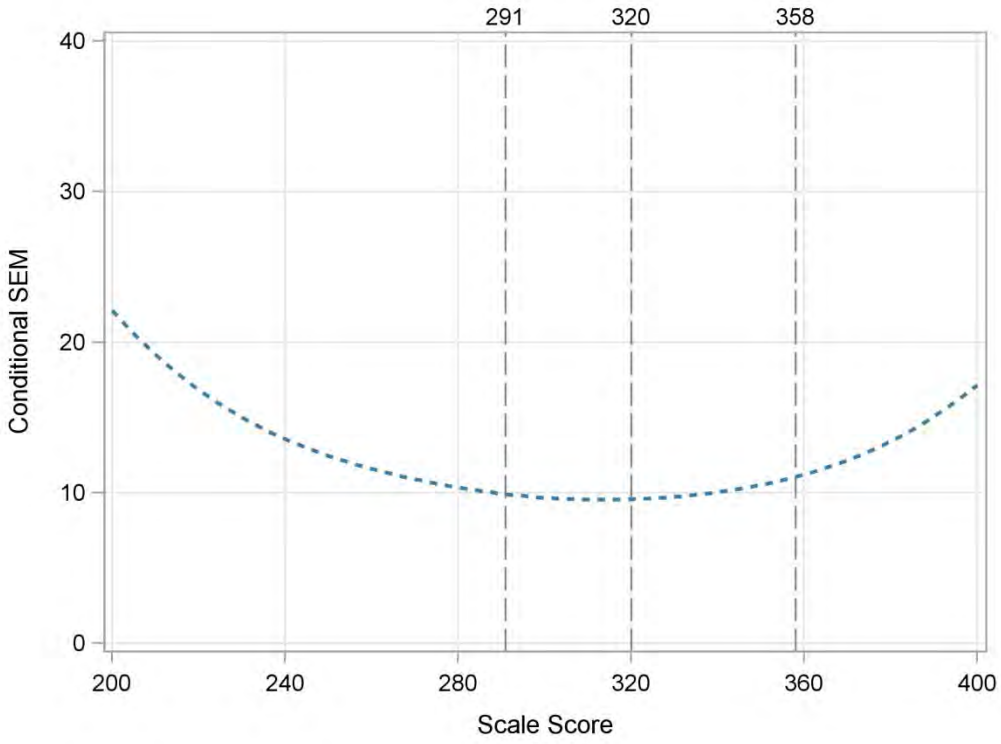

Figure 2. ELA Grade 3 CSEM Curve

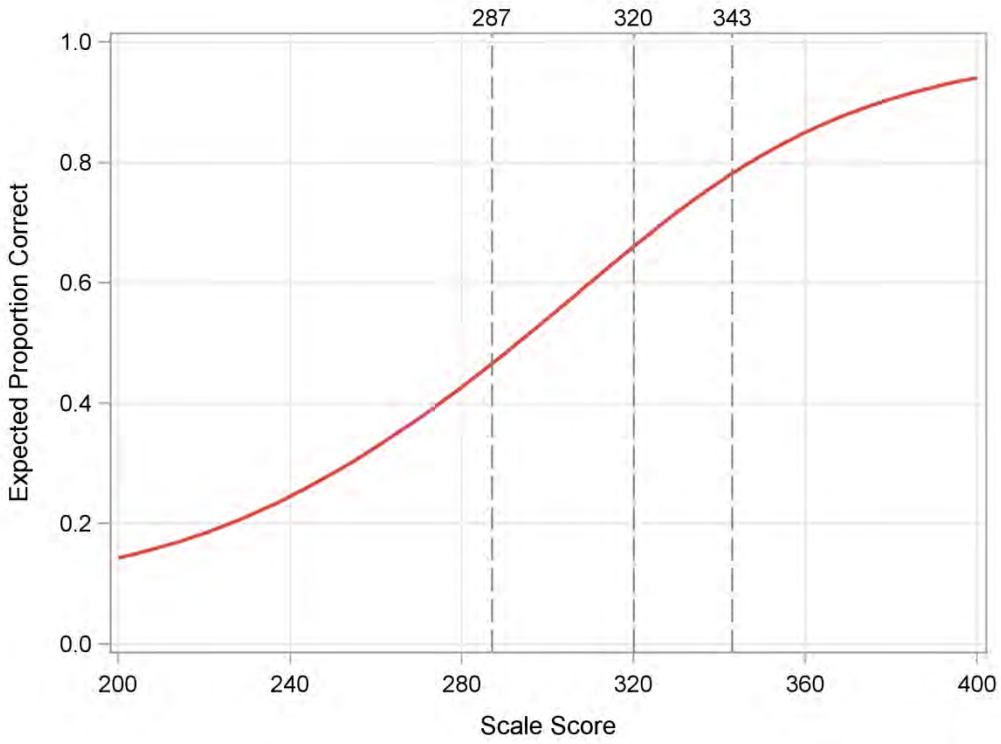

Figure 3. ELA Grade 4 TCC

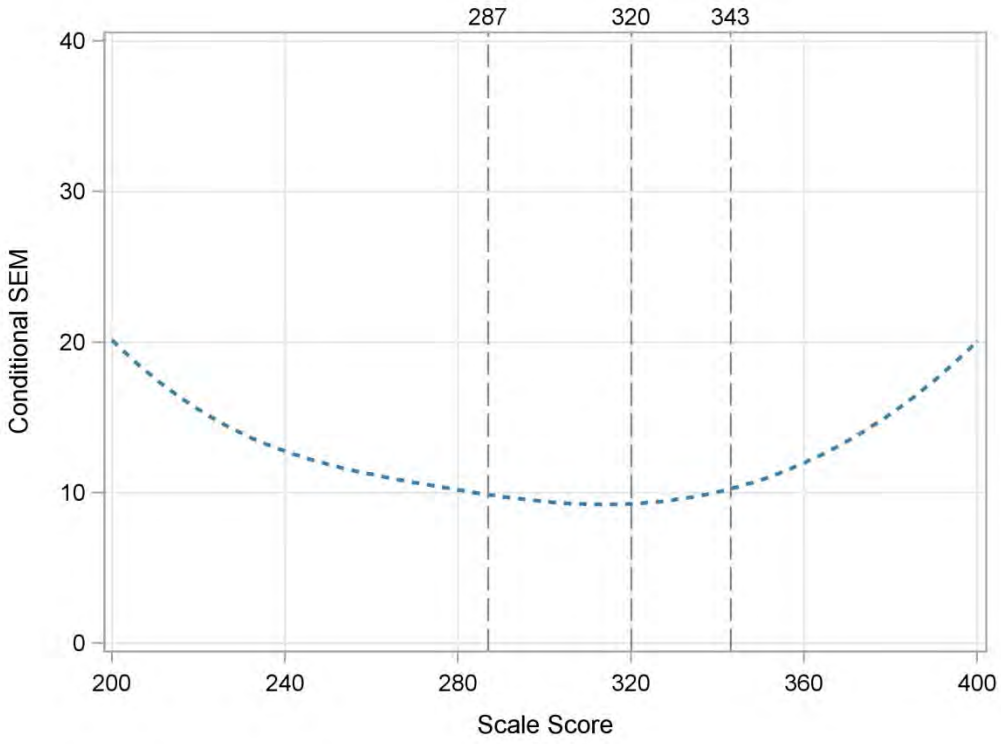

Figure 4. ELA Grade 4 CSEM Curve

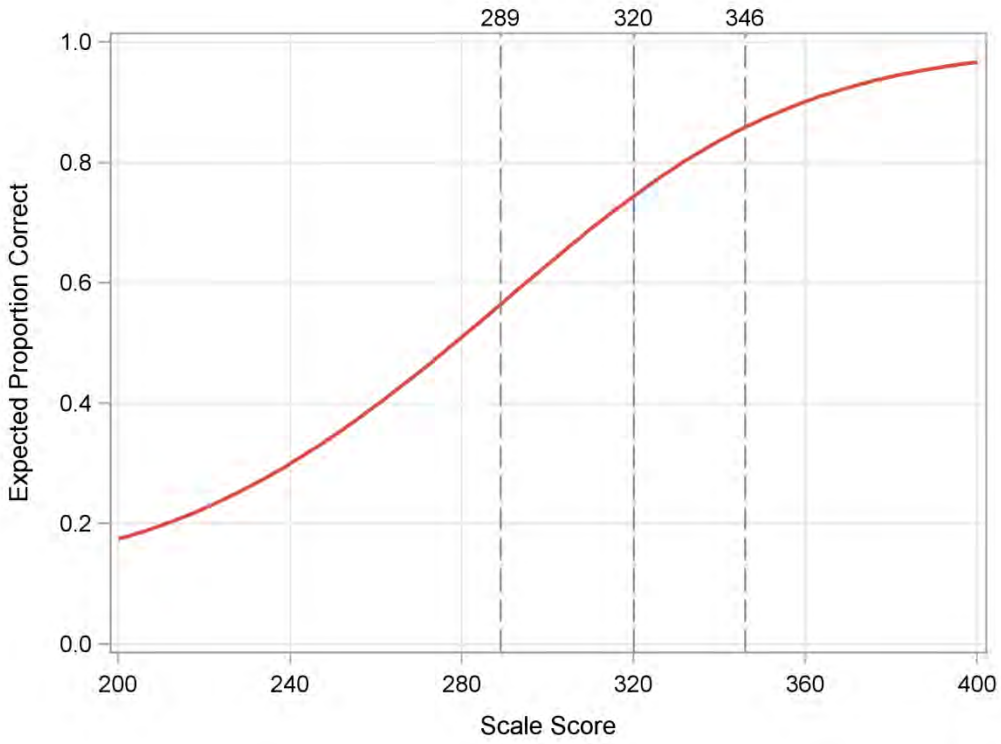

Figure 5. ELA Grade 5 TCC

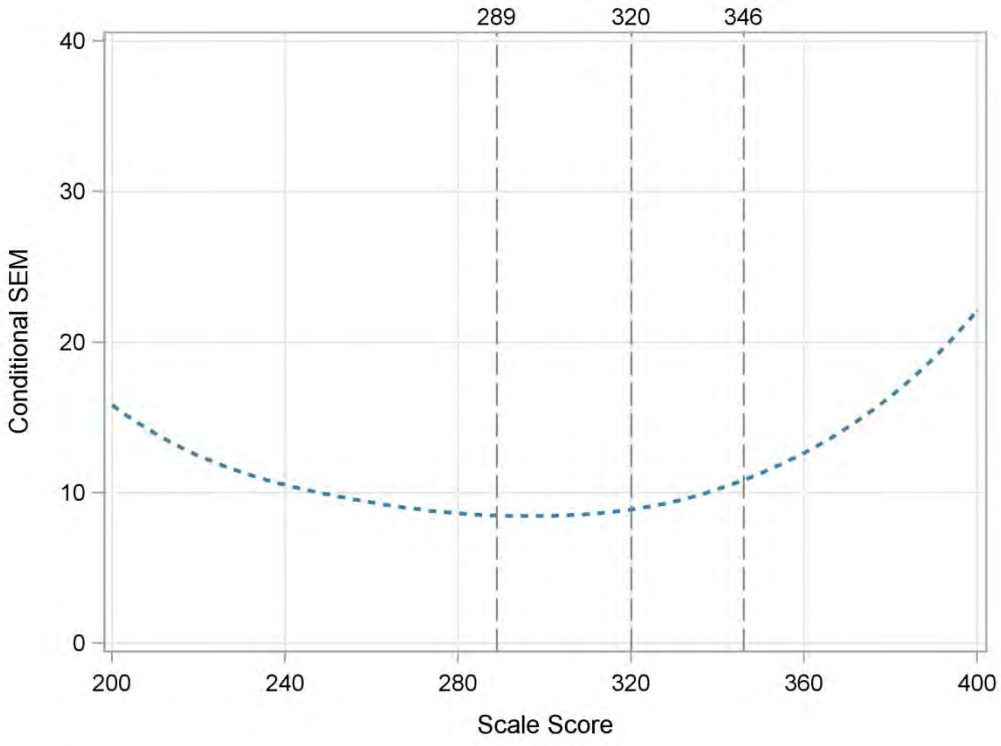

Figure 6. ELA Grade 5 CSEM Curve

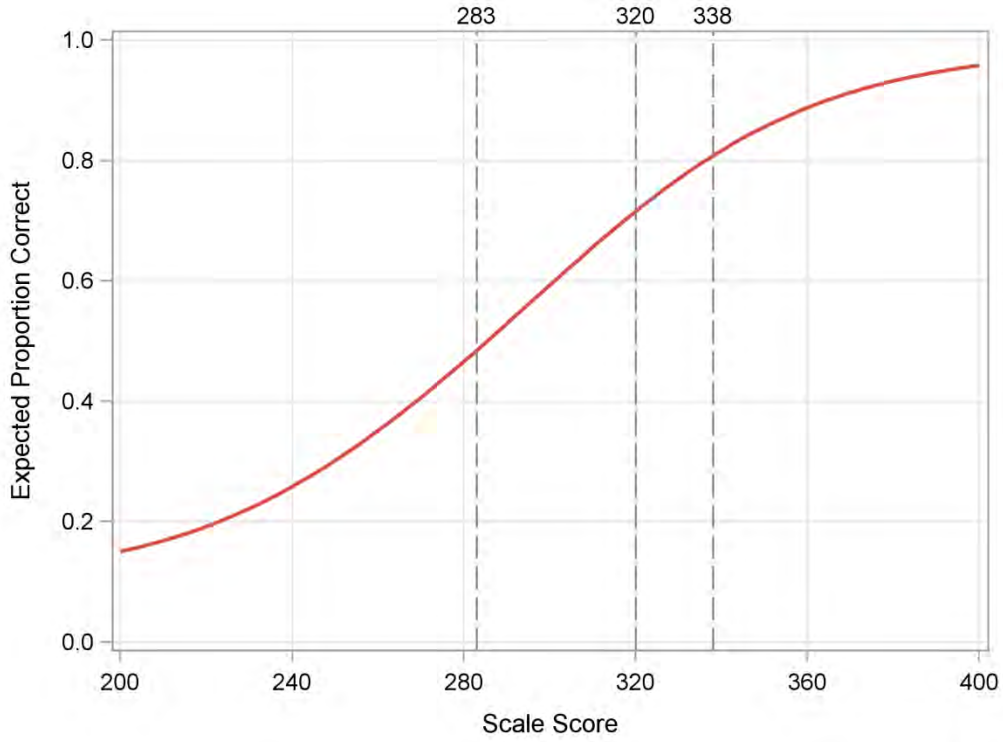

Figure 7. ELA Grade 6 TCC

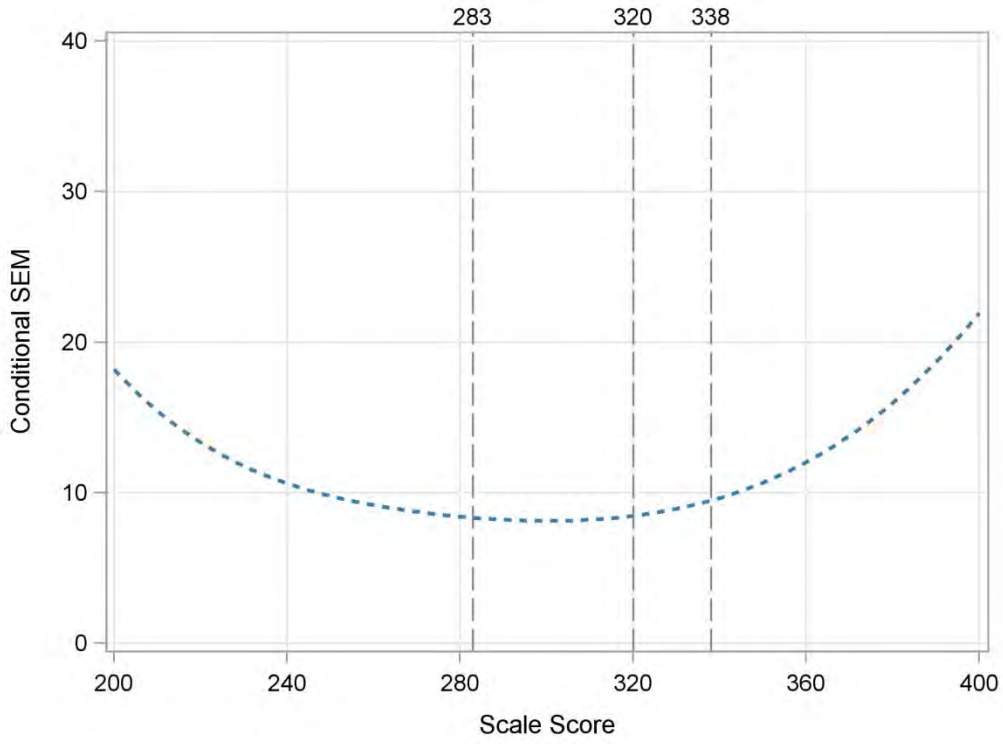

Figure 8. ELA Grade 6 CSEM Curve

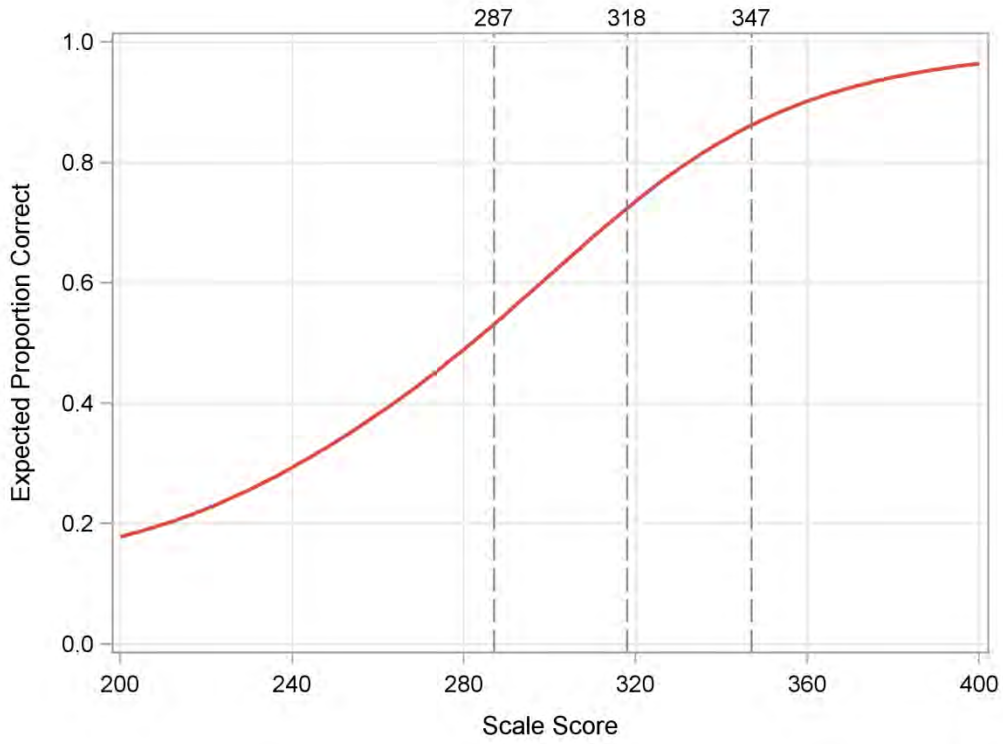

Figure 9. ELA Grade 7 TCC

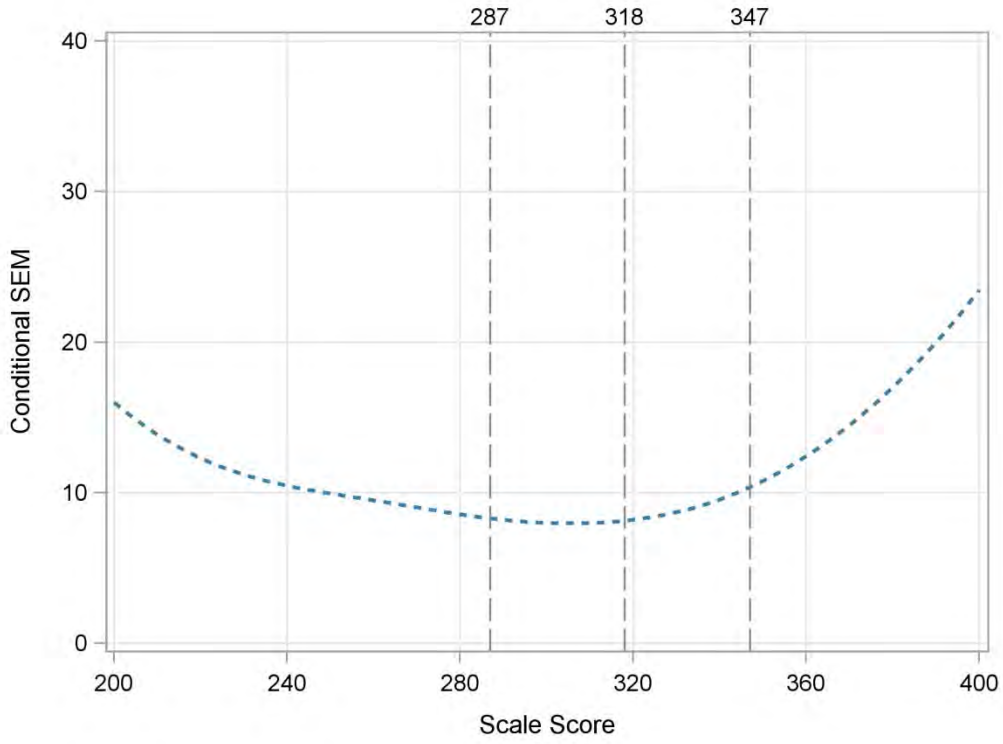

Figure 10. ELA Grade 7 CSEM Curve

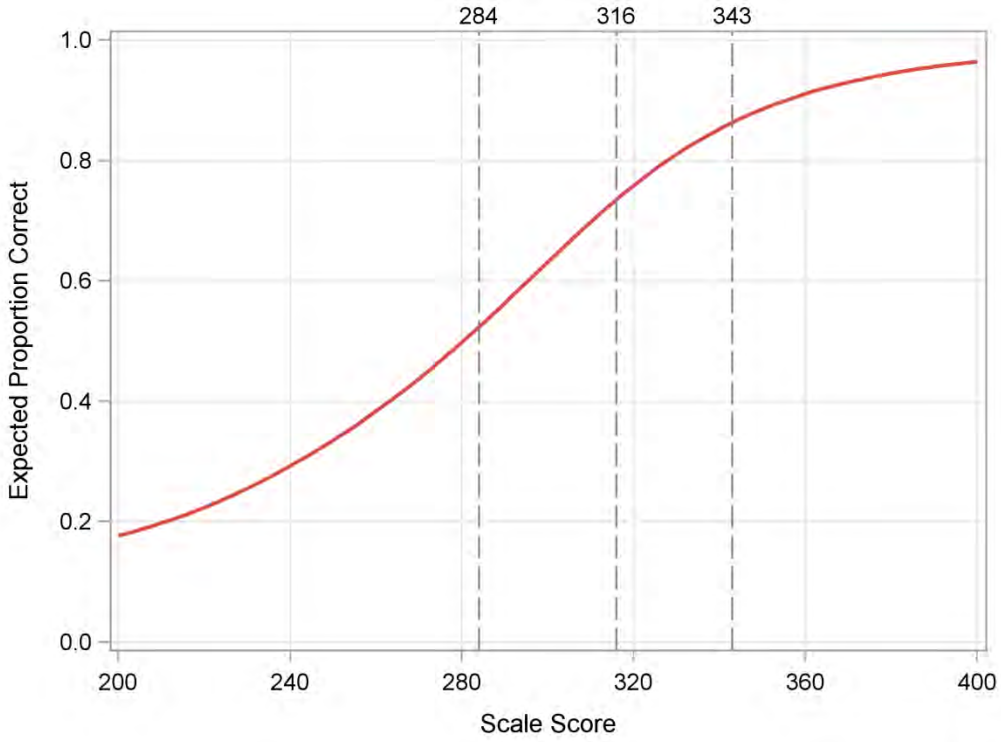

Figure 11. ELA Grade 8 TCC

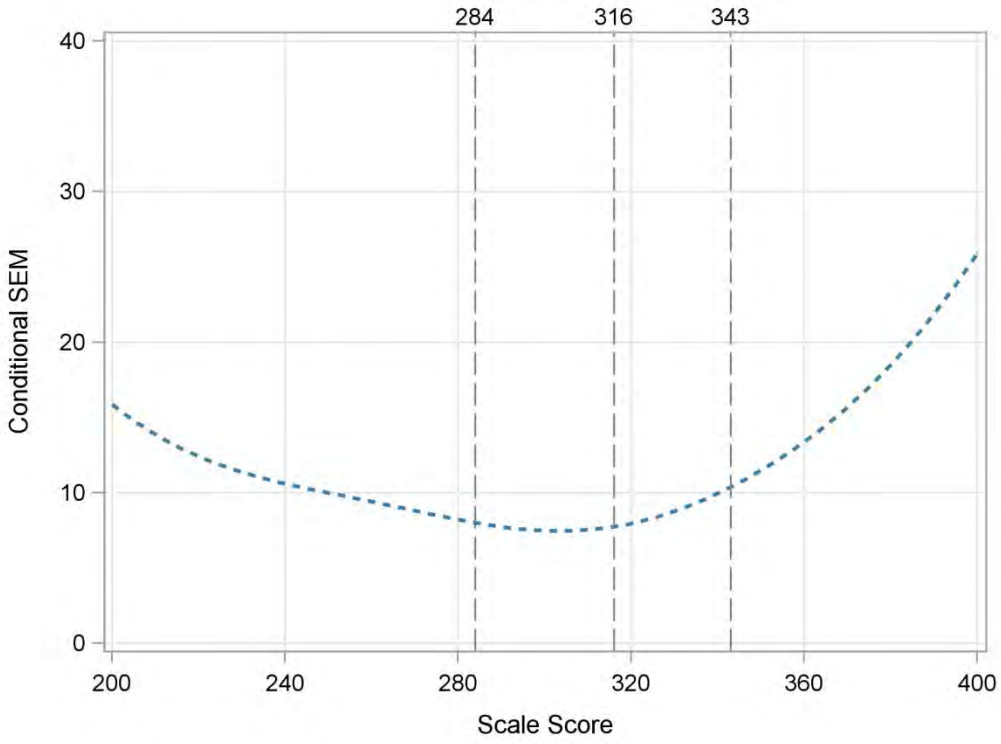

Figure 12. ELA Grade 8 CSEM Curve

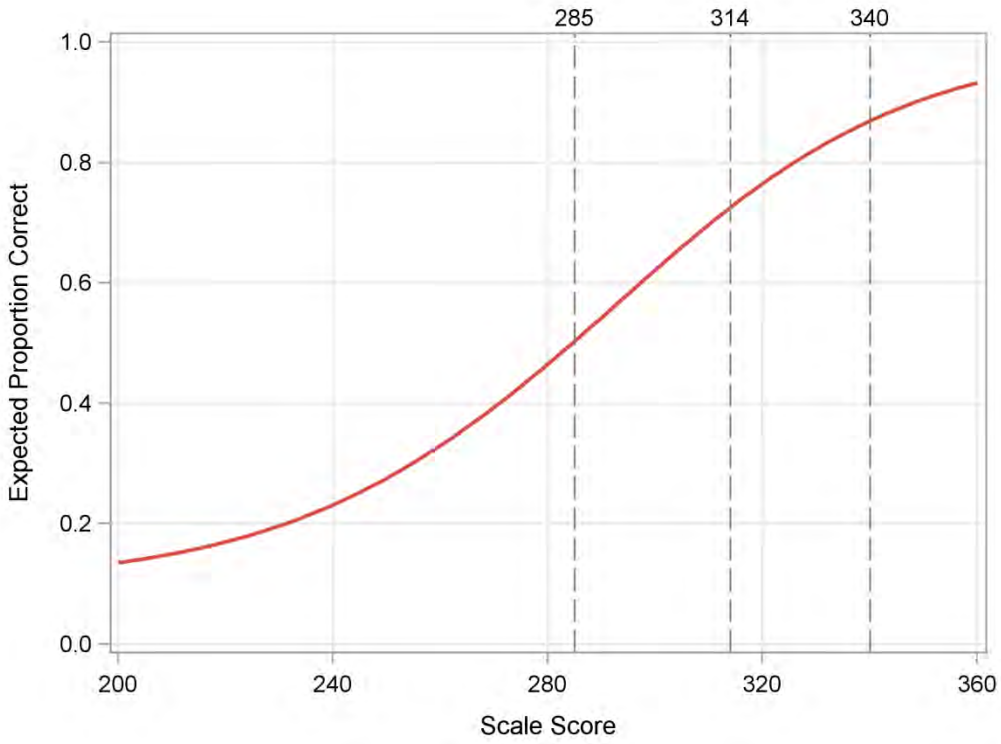

Figure 13. Mathematics Grade 3 TCC

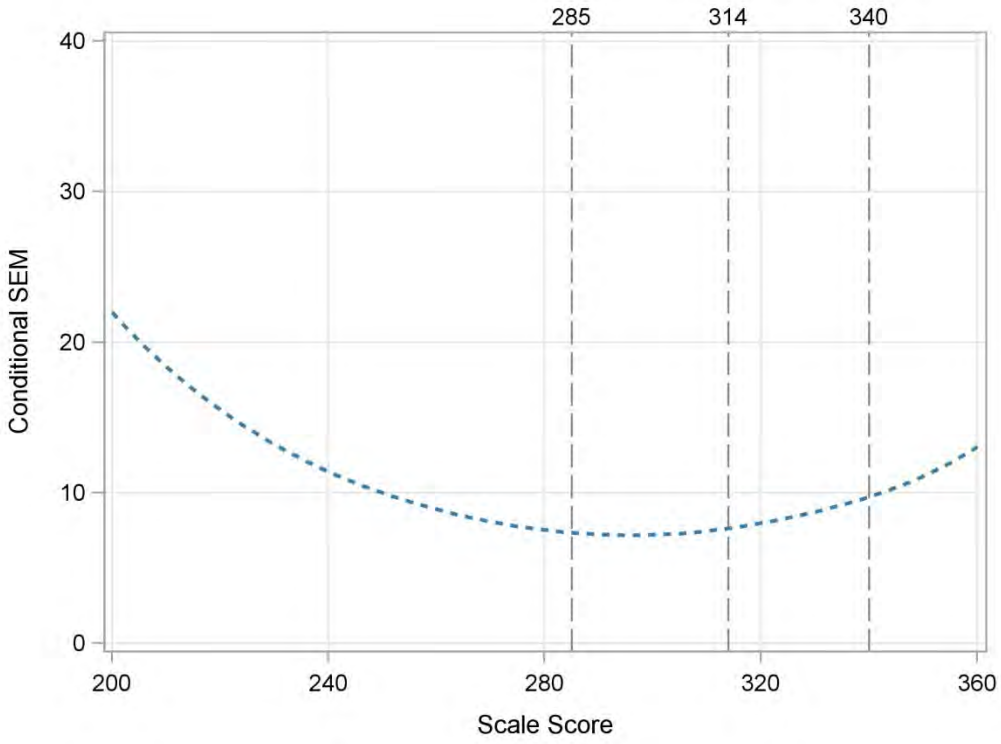

Figure 14. Mathematics Grade 3 CSEM Curve

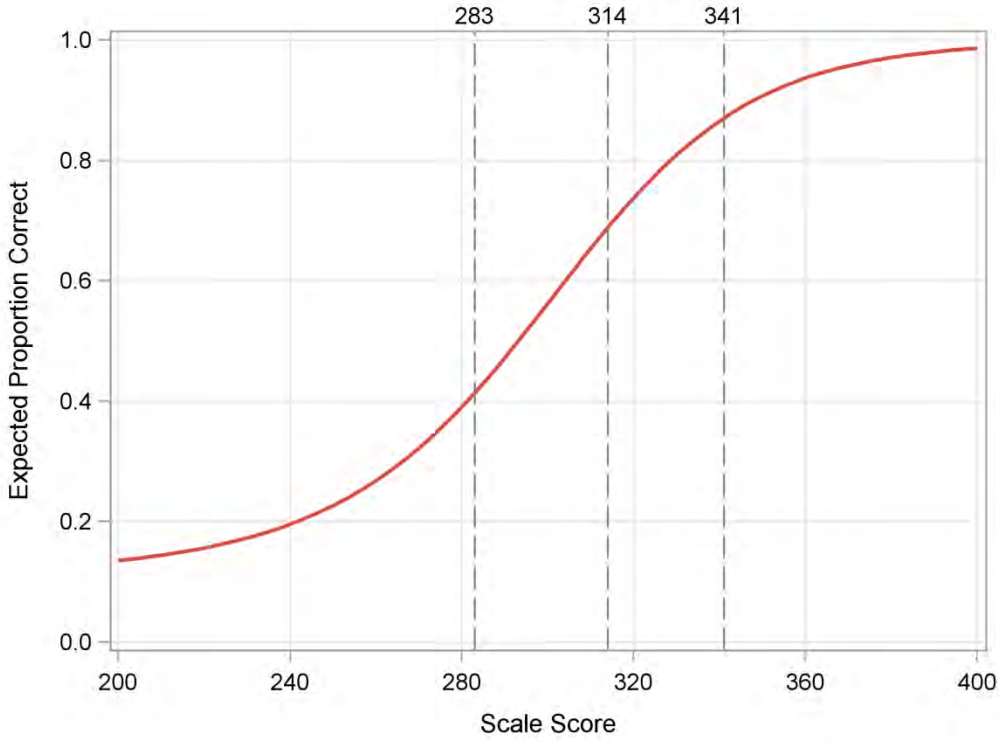

Figure 15. Mathematics Grade 4 TCC

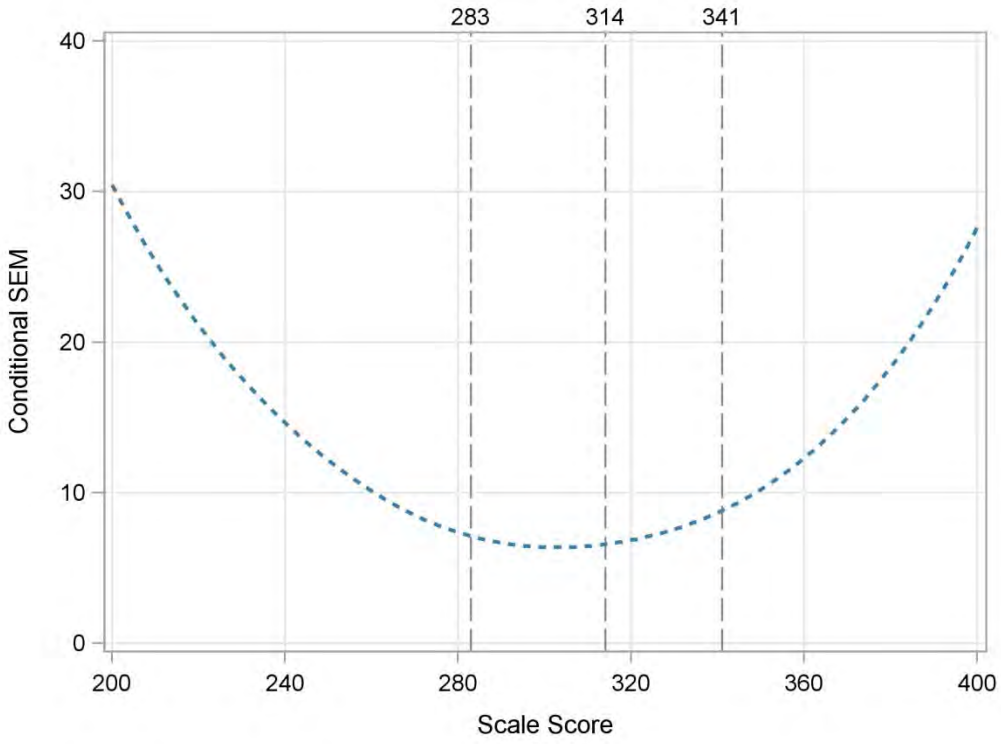

Figure 16. Mathematics Grade 4 CSEM Curve

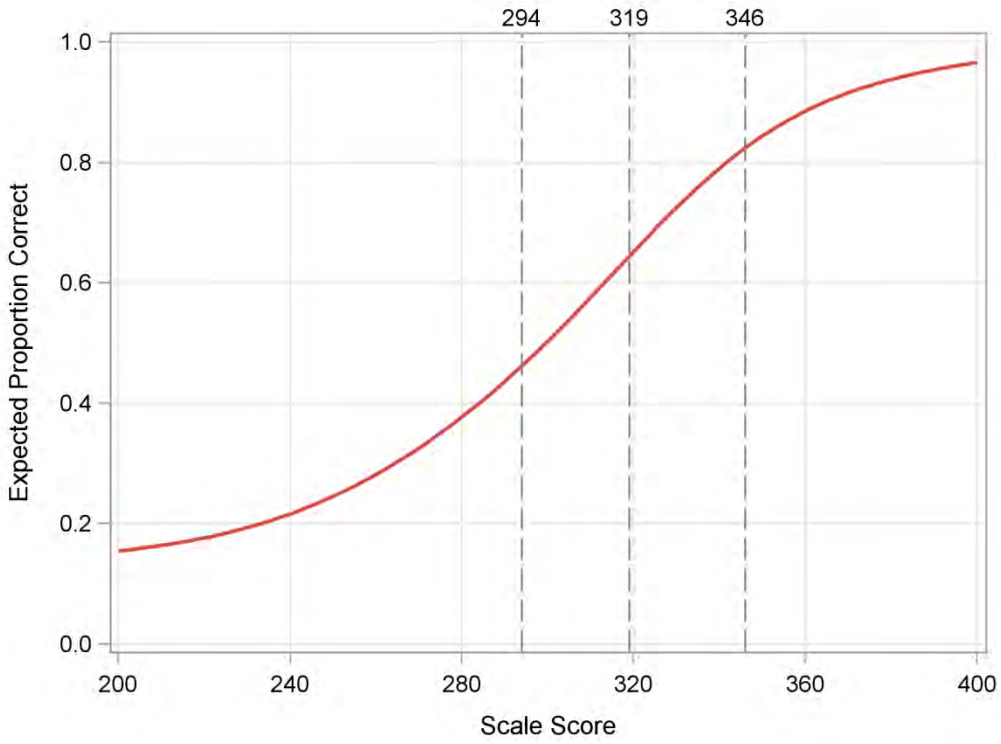

Figure 17. Mathematics Grade 5 TCC

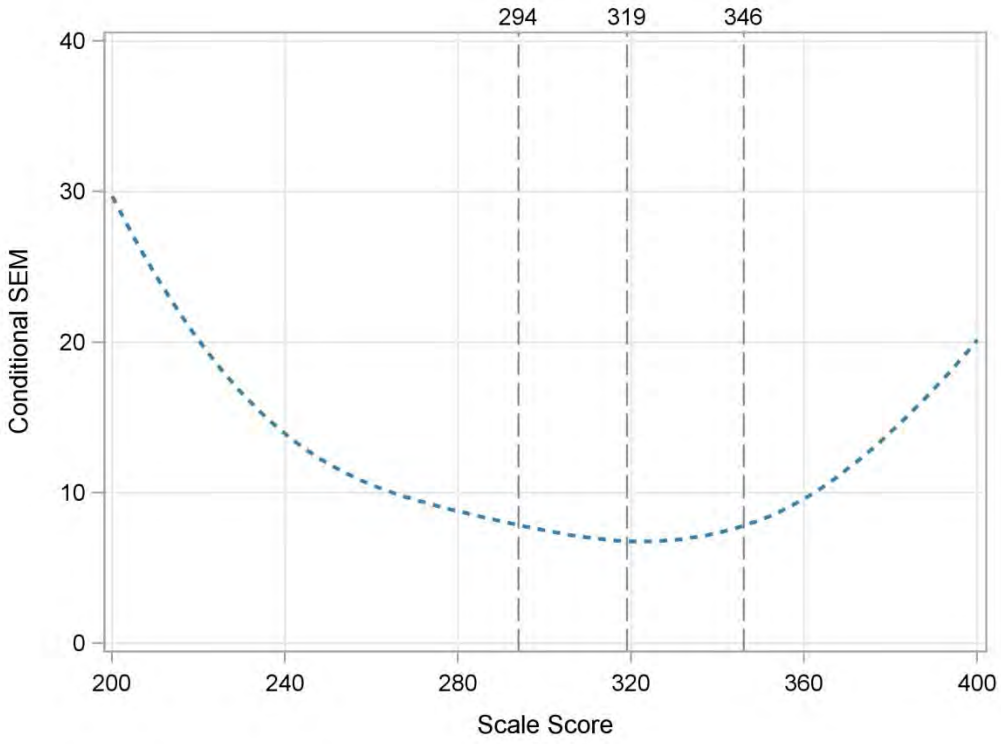

Figure 18. Mathematics Grade 5 CSEM Curve

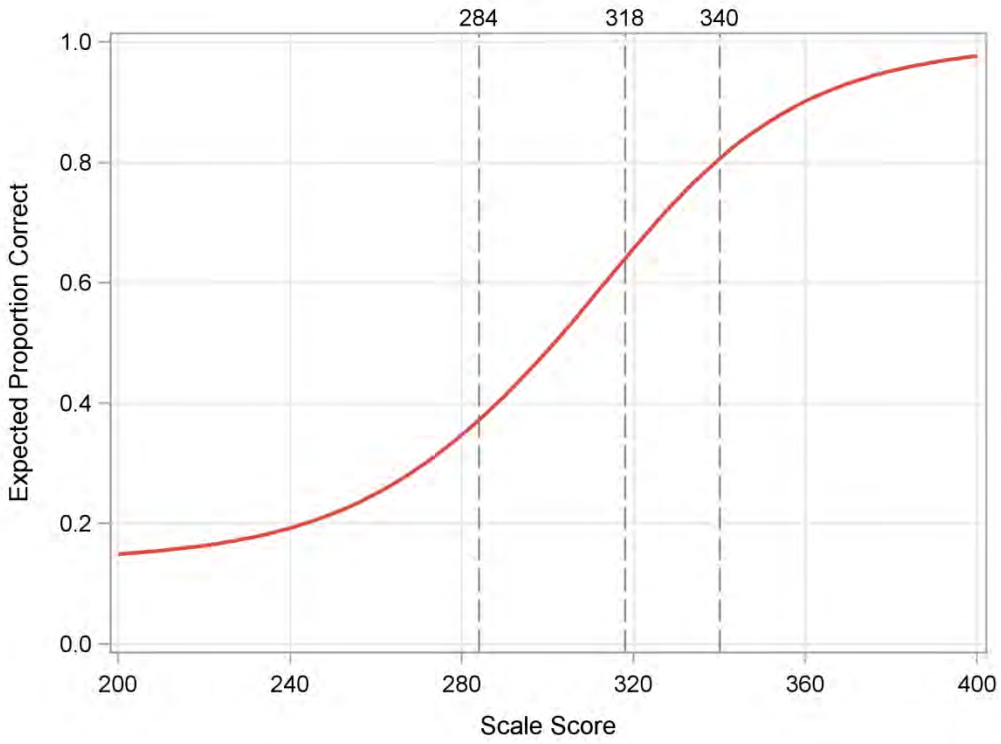

Figure 19. Mathematics Grade 6 TCC

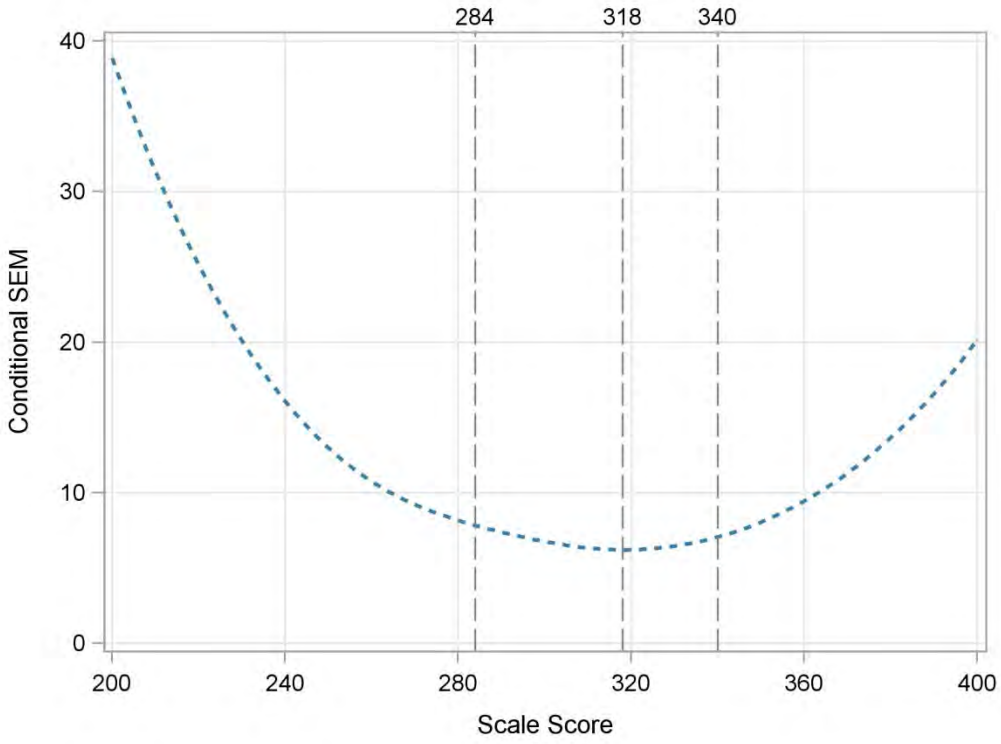

Figure 20. Mathematics Grade 6 CSEM Curve

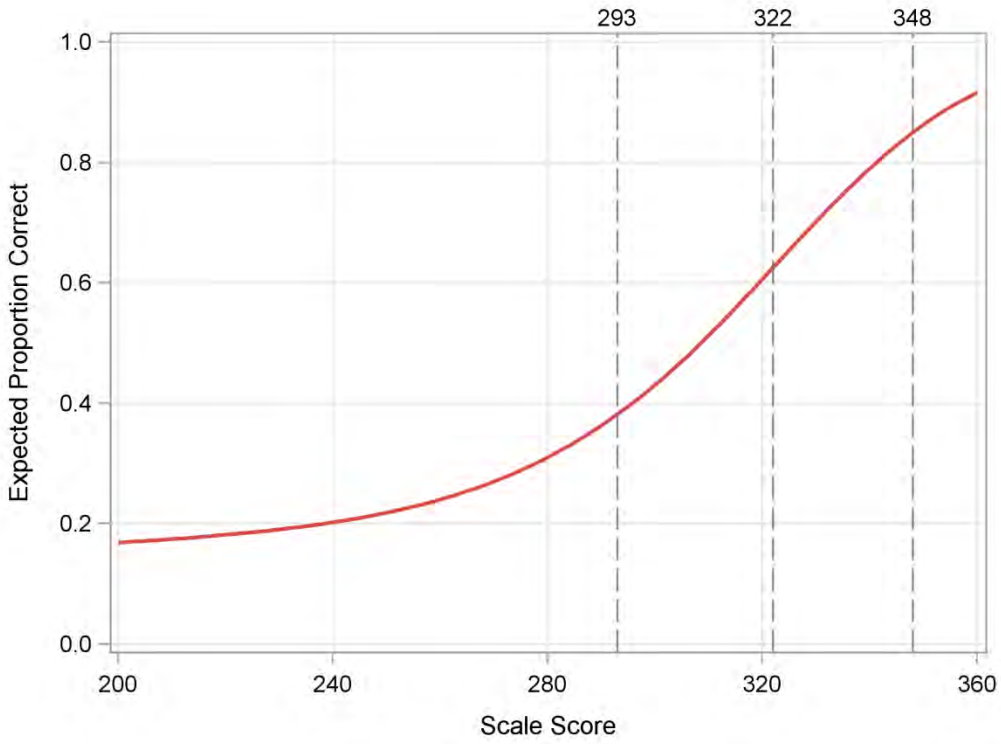

Figure 21. Mathematics Grade 7 TCC

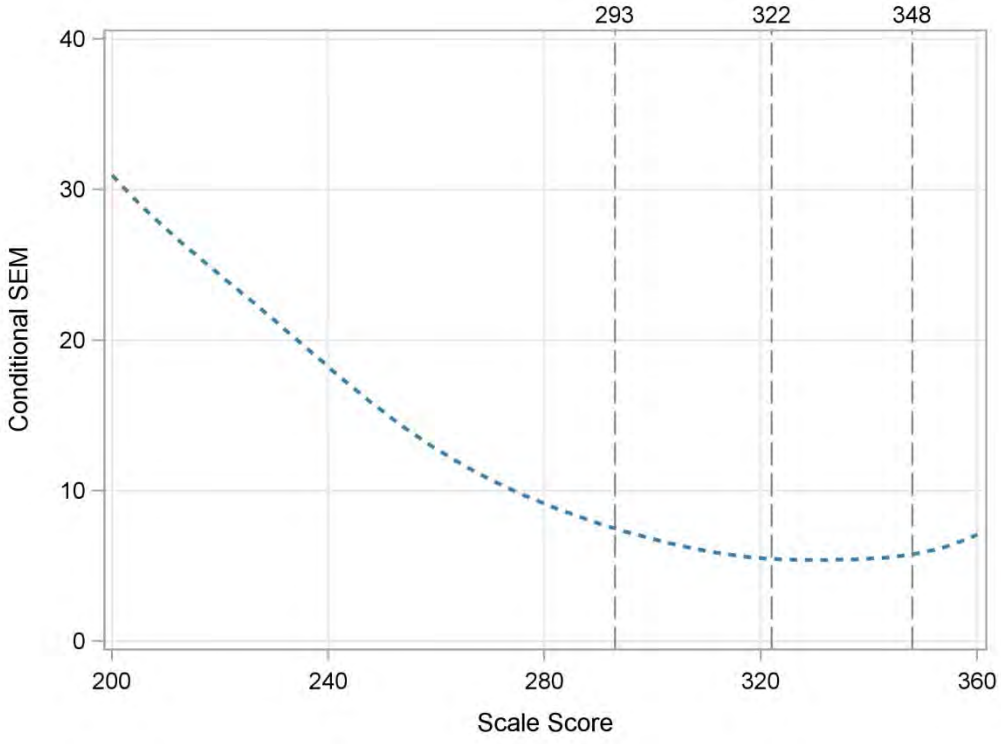

Figure 22. Mathematics Grade 7 CSEM Curve

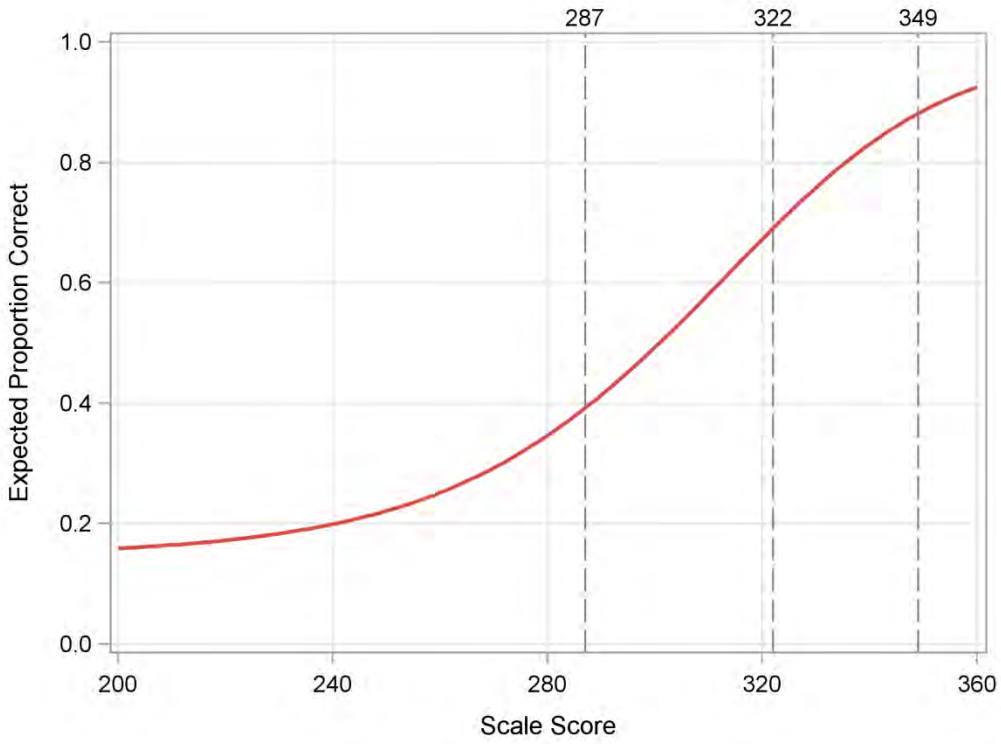

Figure 23. Mathematics Grade 8 TCC

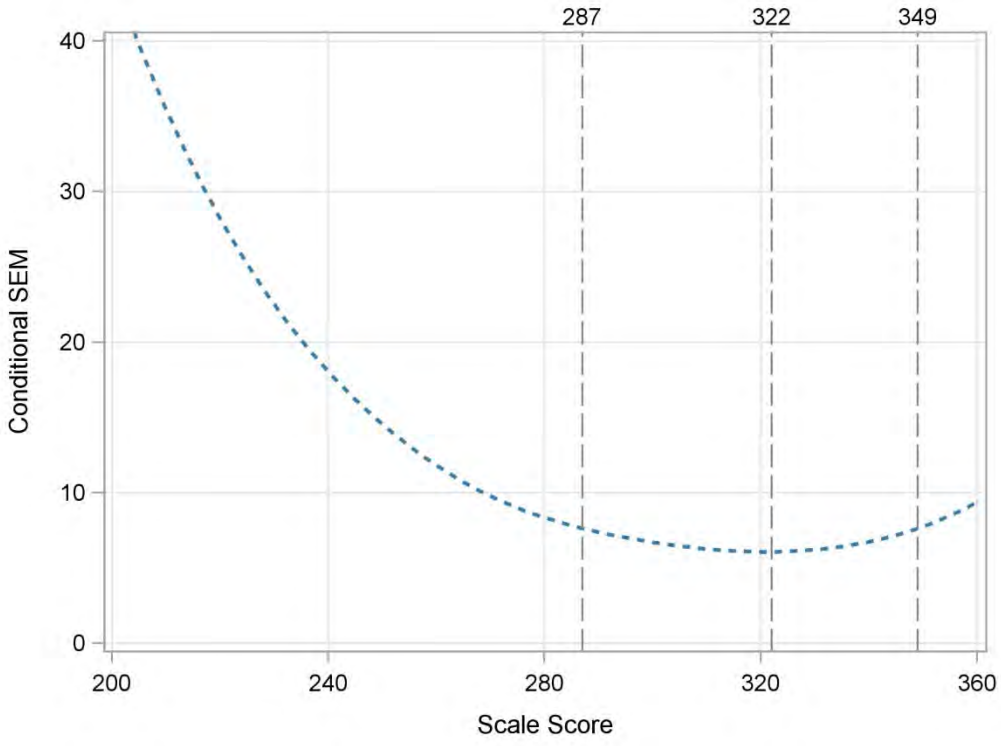

Figure 24. Mathematics Grade 8 CSEM Curve

Scoring Procedure

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

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

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

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

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

where:

x_i is a student's observed raw score on item *i,*

v_i is a non-optimal weight specified in a scoring process (v_i = 1 if no weights are specified), and

$\tilde{\theta}$ is a trait estimate.

It should be noted that potential differences in test form difficulty at different proficiency levels are accounted for in the equating and in the resulting raw score-to-scale score conversion tables, so that students of the same proficiency are expected to obtain the same scale score, regardless of which form they took.

Raw Score-to-Scale Score and SEM Conversion Tables

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

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

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

where

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

$I(\theta)$ is the amount of information provided by the test at a given proficiency level.

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

The equating procedure described above does not guarantee that the same scale score scale points selected as performance-level cut scores will be observed. It was important to appropriately reflect the performance levels set by the standard setting panel and approved by the Commissioner in the Summer of 2013. To that end, if a given scale score cut was not observed in the 2015 RSSS table, the nearest, but lower, scale score value was rounded up to the established scale score cut. In this way, the approved scale score cuts set in 2013 were maintained for 2015.

Table 24 and Table 25 for ELA and mathematics, respectively, show the raw- and scale-score performance level cut scores.

Table 24. ELA Performance-Level Cut Scores

Performance Level	Raw Score Cut (Scale Score Cut)					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	25 (291)	25 (287)	37 (289)	32 (283)	35 (287)	34 (284)
NYS Level III	34 (320)	36 (320)	49 (320)	47 (320)	47 (318)	48 (316)
NYS Level IV	44 (358)	43 (343)	56 (346)	53 (338)	57 (347)	57 (343)

Table 25. Mathematics Performance-Level Cut Scores

Performance Level	Raw Score Cut (Scale Score Cut)					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	30 (285)	27 (283)	30 (294)	27 (284)	27 (293)	28 (287)
NYS Level III	43 (314)	45 (314)	42 (319)	46 (318)	45 (322)	50 (322)
NYS Level IV	52 (340)	57 (341)	54 (346)	58 (340)	61 (348)	63 (349)

Section 7: Reliability and Standard Error of Measurement

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

Test Reliability

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

Test Statistics and Reliability for Total Test

Test statistics including raw-score (RS) means and raw-score standard deviations (SDs) are presented in Table 26 and Table 28 for ELA and mathematics, respectively. These statistics give the necessary context for Table 27 and Table 29, which present the case counts (n-count), number of test items (# Items), Cronbach's alpha and associated SEM, and Feldt-Raju coefficient and associated SEM obtained for the total ELA and mathematics tests. Reliability coefficients provide measures of internal consistency that range from zero to one. High reliability indicates that scores are consistent and not unduly influenced by random error. Overall test reliability is a very good indication of each test's internal consistency.

Grades 3–8 ELA reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from .89 to .93. Grades 3–8 mathematics reliability estimates ranged from .93 to .96 (Cronbach's alpha and Feldt-Raju). The lowest reliabilities were observed for ELA Grade 3 and Grade 4; however, as those tests had the fewest score points, it was reasonable that its reliability would not be as high as the other grade-level tests. The highest reliability was observed for mathematics Grades 4 and 6. All reliabilities were at least .89 across grades, which is a good indication that the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are acceptably reliable.

Table 26. ELA Test Form Statistics

Grade	Item-level			Student-Level			
	P-value			n	Raw Score		
	Mean	Min.	Max.		Max	Mean	SD
3	0.53	0.26	0.83	156,480	55	27.94	9.91
4	0.52	0.19	0.81	151,880	55	30.21	10.15
5	0.60	0.25	0.86	149,480	66	40.62	12.11
6	0.58	0.31	0.91	148,800	66	38.71	12.48
7	0.56	0.17	0.87	137,660	66	38.68	12.06
8	0.58	0.23	0.89	133,180	66	41.15	12.55

Table 27. 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	156,480	41	55	.89	3.22	.90	3.11
4	151,880	41	55	.89	3.29	.90	3.15
5	149,480	52	66	.92	3.50	.92	3.38
6	148,800	52	66	.91	3.65	.92	3.48
7	137,660	52	66	.91	3.63	.92	3.49
8	133,180	52	66	.92	3.57	.93	3.41

Table 28. Mathematics Test Form Statistics

Grade	Item-level			Student-Level			
	P-value			n	Raw Score		
	Mean	Min.	Max.		Max	Mean	SD
3	0.65	0.21	0.92	155,200	60	38.18	13.00
4	0.63	0.36	0.96	148,800	66	39.56	16.38
5	0.58	0.24	0.92	143,920	66	38.01	14.14
6	0.56	0.18	0.89	141,880	72	39.44	16.91
7	0.48	0.04	0.96	126,260	72	37.47	16.73
8	0.49	0.12	0.91	95,280	72	35.40	16.17

Table 29. 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	155,200	49	60	.93	3.34	.94	3.15
4	148,800	52	66	.95	3.73	.96	3.46
5	143,920	52	66	.93	3.60	.94	3.36
6	141,880	58	72	.95	3.90	.95	3.64
7	126,260	58	72	.94	3.98	.95	3.67
8	95,280	58	72	.93	4.12	.94	3.86

Reliability of MC Items

In addition to overall test reliability, Cronbach's alpha and Feldt-Raju coefficient were computed separately for MC and CR item sets. It is important to recognize that reliability is directly affected by test length; therefore, reliability estimates for tests by item type will always be lower than reliability estimates for the overall test form. Table 30 and Table 31 present reliabilities for the subsets of MC items.

Table 30. ELA MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	156,480	31	.83	2.49	.83	2.48
4	151,880	31	.81	2.50	.81	2.49
5	149,480	42	.87	2.83	.87	2.82
6	148,800	42	.86	2.90	.86	2.89
7	137,660	42	.85	2.93	.86	2.92
8	133,180	42	.87	2.86	.88	2.84

Table 31. Mathematics MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	155,200	41	.91	2.52	.91	2.50
4	148,800	42	.93	2.61	.93	2.59
5	143,920	42	.91	2.55	.91	2.53
6	141,880	48	.93	2.89	.93	2.87
7	126,260	48	.91	2.95	.92	2.93
8	95,280	48	.91	3.00	.91	2.99

Reliability of CR Items

Reliability coefficients were also computed for the subsets of CR items. The results are presented in Table 32 and Table 33.

Table 32. 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	156,480	10	24	.85	1.84	.86	1.77
4	151,880	10	24	.86	1.92	.87	1.85
5	149,480	10	24	.87	1.81	.88	1.74
6	148,800	10	24	.89	1.90	.90	1.81
7	137,660	10	24	.89	1.81	.90	1.74
8	133,180	10	24	.88	1.85	.90	1.74

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

Table 33. 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	155,200	8	19	.85	1.94	.86	1.89
4	148,800	10	24	.89	2.35	.90	2.25
5	143,920	10	24	.87	2.27	.88	2.19
6	141,880	10	24	.88	2.31	.89	2.22
7	126,260	10	24	.90	2.25	.91	2.14
8	95,280	10	24	.87	2.47	.88	2.41

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

Test Reliability for Reporting Categories

In this section, reliability coefficients that were estimated for the population and subgroups are presented. The reporting categories include the following: gender, ethnicity, NRC, ELL, all SWD, all SUA, students with disabilities using accommodations falling under 504 Plan (SWD/SUA), and English language learners using accommodations specific to their ELL status (ELL/SUA). Accommodations available to students under the 504 Plan include the following: Flexibility in Scheduling/Timing, Flexibility in Setting, Method of Presentation (excluding braille), Method of Response, Braille and Large-type, and others. Accommodations available to English language learners are Time Extension, Separate Location, Third Reading of Listening Selection, and Bilingual Dictionaries and Glossaries.

As shown in Table 34A–36F and Table 35 A–37F, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Cronbach's alpha reliability coefficients were all at least .79. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least .80 each. These indicate a very good test internal consistency (reliability) for analyzed subgroups of examinees.

Table 34A. ELA Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	156,480	.89	3.22	.90	3.11
Gender	Female	76,902	.89	3.22	.90	3.11
	Male	79,578	.89	3.20	.90	3.10
Ethnicity	Asian	14,868	.88	3.16	.89	3.06
	Black	30,246	.89	3.24	.89	3.13
	Hispanic	43,405	.88	3.22	.89	3.13
	American Indian	959	.89	3.24	.89	3.13
	Multiracial	3,098	.90	3.21	.91	3.08
	Pacific Islander	421	.89	3.19	.89	3.07
	White	63,483	.89	3.19	.90	3.08
NRC	New York	62,276	.89	3.21	.90	3.10
	Big 4 Cities	6,785	.90	3.19	.90	3.08
	Urban/Suburban	12,765	.88	3.20	.89	3.12
	Rural	8,581	.89	3.19	.90	3.10
	Average Needs	38,197	.89	3.20	.90	3.10
	Low Needs	17,416	.87	3.13	.88	3.05
	Charter	7,298	.87	3.27	.88	3.18
	Non-Public	2,997	.89	3.27	.90	3.13
SWD	All Codes	23,510	.88	3.15	.89	3.06
SUA	All Codes	18,218	.88	3.18	.89	3.08
ELL	ELL=Y	13,513	.85	3.19	.86	3.10
SWD/SUA	SUA=504 plan codes	9,536	.87	3.09	.87	3.01
ELL/SUA	SUA & ELL codes	4,890	.84	3.18	.85	3.10

Table 34B. ELA Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	151,880	.89	3.29	.90	3.15
Gender	Female	74,667	.89	3.29	.90	3.14
	Male	77,213	.90	3.29	.91	3.15
Ethnicity	Asian	14,713	.89	3.19	.90	3.05
	Black	29,000	.88	3.31	.89	3.19
	Hispanic	40,129	.88	3.30	.89	3.18
	American Indian	883	.89	3.31	.90	3.16
	Multiracial	2,438	.90	3.29	.91	3.12
	Pacific Islander	346	.88	3.25	.89	3.12
	White	64,371	.89	3.27	.90	3.12
NRC	New York	57,551	.90	3.27	.90	3.12
	Big 4 Cities	6,432	.89	3.30	.90	3.15
	Urban/Suburban	11,799	.88	3.29	.89	3.18
	Rural	8,119	.89	3.32	.90	3.18
	Average Needs	37,550	.89	3.29	.90	3.15
	Low Needs	17,506	.86	3.18	.87	3.07
	Charter	5,876	.85	3.31	.86	3.22
	Non-Public	6,916	.87	3.34	.88	3.20
SWD	All Codes	23,756	.87	3.23	.88	3.12
SUA	All Codes	17,993	.88	3.26	.89	3.14
ELL	ELL=Y	11,563	.85	3.25	.86	3.14
SWD/SUA	SUA=504 plan codes	10,203	.86	3.18	.87	3.09
ELL/SUA	SUA & ELL codes	4,147	.84	3.24	.85	3.14

Table 34C. ELA Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	149,480	.92	3.50	.92	3.38
Gender	Female	73,442	.91	3.46	.92	3.35
	Male	76,038	.92	3.52	.92	3.40
Ethnicity	Asian	14,376	.91	3.30	.92	3.18
	Black	28,940	.91	3.60	.91	3.48
	Hispanic	39,110	.91	3.57	.91	3.46
	American Indian	867	.91	3.57	.92	3.44
	Multiracial	2,110	.92	3.45	.93	3.32
	Pacific Islander	331	.90	3.47	.91	3.35
	White	63,746	.91	3.43	.92	3.31
NRC	New York	56,927	.92	3.49	.92	3.37
	Big 4 Cities	6,425	.92	3.65	.92	3.51
	Urban/Suburban	11,895	.91	3.60	.92	3.48
	Rural	7,954	.91	3.57	.92	3.46
	Average Needs	37,665	.91	3.47	.92	3.36
	Low Needs	18,730	.89	3.31	.89	3.22
	Charter	6,866	.89	3.52	.89	3.44
	Non-Public	2,870	.92	3.51	.92	3.35
SWD	All Codes	25,010	.90	3.63	.90	3.51
SUA	All Codes	18,649	.90	3.63	.91	3.51
ELL	ELL=Y	10,867	.88	3.67	.89	3.54
SWD/SUA	SUA=504 plan codes	11,088	.89	3.60	.90	3.50
ELL/SUA	SUA & ELL codes	3,834	.87	3.65	.88	3.54

Table 34D. ELA Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	148,800	.91	3.65	.92	3.48
Gender	Female	72,832	.91	3.61	.92	3.46
	Male	75,968	.92	3.66	.92	3.50
Ethnicity	Asian	14,117	.91	3.44	.92	3.30
	Black	29,822	.90	3.72	.91	3.57
	Hispanic	38,772	.90	3.70	.91	3.55
	American Indian	798	.91	3.69	.91	3.53
	Multiracial	1,952	.92	3.64	.93	3.45
	Pacific Islander	327	.91	3.61	.92	3.47
	White	63,012	.91	3.58	.92	3.43
NRC	New York	55,019	.91	3.64	.92	3.47
	Big 4 Cities	6,193	.91	3.74	.92	3.57
	Urban/Suburban	11,437	.91	3.72	.91	3.57
	Rural	7,747	.91	3.72	.92	3.54
	Average Needs	36,332	.91	3.63	.92	3.47
	Low Needs	17,982	.89	3.46	.90	3.35
	Charter	6,856	.89	3.71	.89	3.59
	Non-Public	7,075	.90	3.61	.91	3.47
SWD	All Codes	23,841	.88	3.69	.89	3.55
SUA	All Codes	17,212	.89	3.70	.90	3.56
ELL	ELL=Y	9,492	.85	3.69	.86	3.55
SWD/SUA	SUA=504 plan codes	10,655	.87	3.65	.88	3.53
ELL/SUA	SUA & ELL codes	3,087	.83	3.67	.85	3.55

Table 34E. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	137,660	.91	3.63	.92	3.49
Gender	Female	67,182	.90	3.57	.91	3.45
	Male	70,478	.91	3.64	.92	3.50
Ethnicity	Asian	12,881	.91	3.43	.92	3.30
	Black	28,005	.89	3.70	.90	3.57
	Hispanic	35,180	.89	3.67	.90	3.55
	American Indian	799	.91	3.70	.91	3.53
	Multiracial	1,524	.92	3.63	.93	3.45
	Pacific Islander	294	.91	3.58	.91	3.45
	White	58,977	.91	3.57	.92	3.43
NRC	New York	52,766	.90	3.60	.91	3.47
	Big 4 Cities	5,880	.90	3.73	.91	3.57
	Urban/Suburban	10,852	.90	3.70	.90	3.57
	Rural	7,760	.91	3.68	.91	3.53
	Average Needs	33,707	.91	3.62	.91	3.48
	Low Needs	18,665	.89	3.45	.90	3.36
	Charter	5,104	.87	3.64	.88	3.56
	Non-Public	2,808	.91	3.64	.92	3.46
SWD	All Codes	22,276	.87	3.67	.88	3.54
SUA	All Codes	15,434	.88	3.68	.89	3.55
ELL	ELL=Y	8,492	.82	3.67	.84	3.53
SWD/SUA	SUA=504 plan codes	10,047	.86	3.65	.87	3.53
ELL/SUA	SUA & ELL codes	2,387	.81	3.64	.82	3.51

Table 34F. ELA Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	133,180	.92	3.57	.93	3.41
Gender	Female	64,436	.91	3.50	.92	3.36
	Male	68,744	.92	3.61	.93	3.44
Ethnicity	Asian	13,096	.92	3.32	.92	3.18
	Black	27,114	.91	3.68	.91	3.53
	Hispanic	33,146	.90	3.66	.91	3.52
	American Indian	735	.92	3.67	.92	3.48
	Multiracial	1,404	.93	3.54	.94	3.36
	Pacific Islander	303	.91	3.43	.92	3.28
	White	57,382	.92	3.48	.93	3.33
NRC	New York	51,028	.91	3.56	.92	3.41
	Big 4 Cities	5,332	.91	3.77	.92	3.58
	Urban/Suburban	10,112	.91	3.71	.92	3.55
	Rural	7,583	.92	3.63	.93	3.46
	Average Needs	31,762	.92	3.55	.93	3.39
	Low Needs	17,672	.91	3.34	.91	3.23
	Charter	3,752	.88	3.59	.88	3.51
	Non-Public	5,794	.91	3.49	.91	3.35
SWD	All Codes	20,575	.89	3.68	.90	3.54
SUA	All Codes	14,480	.90	3.70	.91	3.54
ELL	ELL=Y	7,420	.85	3.70	.86	3.54
SWD/SUA	SUA=504 plan codes	9,504	.89	3.67	.90	3.53
ELL/SUA	SUA & ELL codes	2,037	.84	3.67	.86	3.52

Table 35A. Mathematics Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	155,200	.93	3.34	.94	3.15
Gender	Female	76,023	.93	3.33	.94	3.15
	Male	79,177	.94	3.34	.94	3.15
Ethnicity	Asian	15,093	.93	3.04	.94	2.84
	Black	29,844	.93	3.41	.94	3.27
	Hispanic	43,592	.93	3.41	.93	3.27
	American Indian	944	.93	3.36	.93	3.21
	Multiracial	3,045	.93	3.31	.94	3.13
	Pacific Islander	421	.93	3.23	.94	3.05
	White	62,261	.93	3.24	.93	3.08
NRC	New York	62,433	.93	3.36	.94	3.17
	Big 4 Cities	6,822	.93	3.43	.94	3.29
	Urban/Suburban	12,659	.93	3.42	.93	3.28
	Rural	8,330	.93	3.37	.93	3.22
	Average Needs	37,473	.93	3.29	.93	3.13
	Low Needs	17,124	.92	3.10	.93	2.93
	Charter	7,180	.93	3.21	.94	3.02
	Non-Public	3,022	.93	3.34	.93	3.17
SWD	All Codes	23,340	.93	3.46	.93	3.33
SUA	All Codes	18,594	.93	3.45	.93	3.32
ELL	ELL=Y	15,096	.92	3.42	.93	3.30
SWD/SUA	SUA=504 plan codes	9,506	.92	3.45	.93	3.34
ELL/SUA	SUA & ELL codes	5,641	.92	3.42	.92	3.31
ELL Test Language	English	12,162	.92	3.42	.93	3.30
	Chinese	502	.92	3.23	.93	3.02
	Haitian-Creole	40	.85	3.49	.86	3.40
	Korean	18	.92	2.94	.94	2.68
	Russian	67	.93	3.41	.94	3.19
	Spanish	2,297	.90	3.39	.91	3.31
	All Translations	2,924	.93	3.42	.94	3.27

Table 35B. Mathematics Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	148,800	.95	3.73	.96	3.46
Gender	Female	72,708	.95	3.73	.95	3.46
	Male	76,092	.95	3.73	.96	3.45
Ethnicity	Asian	14,737	.94	3.43	.95	3.14
	Black	28,296	.94	3.76	.95	3.54
	Hispanic	39,740	.94	3.79	.95	3.56
	American Indian	855	.95	3.76	.95	3.50
	Multiracial	2,375	.95	3.69	.96	3.40
	Pacific Islander	344	.94	3.74	.95	3.46
	White	62,453	.94	3.65	.95	3.40
NRC	New York	56,989	.95	3.75	.96	3.46
	Big 4 Cities	6,386	.94	3.69	.95	3.47
	Urban/Suburban	11,669	.94	3.77	.95	3.55
	Rural	7,756	.94	3.78	.95	3.54
	Average Needs	36,159	.94	3.70	.95	3.44
	Low Needs	17,170	.93	3.50	.94	3.26
	Charter	5,706	.94	3.73	.95	3.46
	Non-Public	6,846	.94	3.74	.94	3.53
SWD	All Codes	23,146	.94	3.66	.94	3.49
SUA	All Codes	17,557	.94	3.69	.94	3.51
ELL	ELL=Y	12,951	.94	3.67	.94	3.49
SWD/SUA	SUA=504 plan codes	9,669	.93	3.60	.94	3.45
ELL/SUA	SUA & ELL codes	4,638	.93	3.59	.93	3.45
ELL Test Language	English	9,907	.93	3.67	.94	3.51
	Chinese	429	.93	3.65	.94	3.36
	Haitian-Creole	41	.85	3.60	.86	3.50
	Korean	24	.90	3.38	.92	3.15
	Russian	66	.95	3.84	.95	3.55
	Spanish	2,245	.91	3.52	.91	3.42
	All Translations	2,805	.94	3.63	.95	3.44

Table 35C. Mathematics Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	143,920	.94	3.60	.94	3.36
Gender	Female	70,206	.93	3.59	.94	3.35
	Male	73,714	.94	3.61	.95	3.36
Ethnicity	Asian	14,161	.93	3.36	.94	3.08
	Black	27,779	.92	3.58	.93	3.40
	Hispanic	38,145	.92	3.60	.93	3.40
	American Indian	804	.93	3.60	.94	3.38
	Multiracial	2,039	.94	3.61	.94	3.35
	Pacific Islander	329	.93	3.58	.94	3.31
	White	60,663	.93	3.57	.94	3.34
NRC	New York	55,601	.94	3.58	.95	3.32
	Big 4 Cities	6,267	.93	3.54	.93	3.37
	Urban/Suburban	11,415	.93	3.59	.93	3.40
	Rural	7,413	.93	3.63	.93	3.42
	Average Needs	35,645	.93	3.61	.94	3.38
	Low Needs	18,006	.92	3.47	.93	3.26
	Charter	6,616	.92	3.59	.93	3.37
	Non-Public	2,830	.93	3.64	.94	3.41
SWD	All Codes	23,808	.92	3.51	.92	3.37
SUA	All Codes	17,796	.92	3.54	.93	3.38
ELL	ELL=Y	12,049	.92	3.50	.93	3.35
SWD/SUA	SUA=504 plan codes	10,293	.91	3.46	.92	3.34
ELL/SUA	SUA & ELL codes	4,162	.91	3.44	.92	3.31
ELL Test Language	English	9,201	.91	3.49	.92	3.36
	Chinese	396	.92	3.61	.94	3.32
	Haitian-Creole	33	.83	3.45	.84	3.34
	Korean	25	.88	3.17	.90	2.85
	Russian	77	.92	3.59	.93	3.40
	Spanish	2,165	.89	3.38	.89	3.28
	All Translations	2,696	.93	3.49	.93	3.32

Table 35D. Mathematics Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	141,880	.95	3.90	.95	3.64
Gender	Female	69,067	.95	3.91	.95	3.65
	Male	72,813	.95	3.88	.96	3.63
Ethnicity	Asian	13,859	.95	3.59	.95	3.32
	Black	28,525	.93	3.90	.94	3.71
	Hispanic	37,601	.94	3.93	.94	3.72
	American Indian	746	.94	3.90	.94	3.69
	Multiracial	1,819	.95	3.88	.96	3.61
	Pacific Islander	333	.95	3.86	.96	3.58
	White	58,997	.94	3.85	.95	3.62
NRC	New York	53,695	.95	3.90	.96	3.63
	Big 4 Cities	6,004	.93	3.82	.94	3.64
	Urban/Suburban	10,883	.93	3.87	.94	3.69
	Rural	7,297	.93	3.93	.94	3.74
	Average Needs	33,517	.94	3.89	.95	3.66
	Low Needs	16,919	.93	3.69	.94	3.47
	Charter	6,552	.94	3.90	.95	3.67
	Non-Public	6,875	.93	3.96	.94	3.73
SWD	All Codes	22,230	.92	3.75	.92	3.62
SUA	All Codes	16,963	.93	3.80	.93	3.65
ELL	ELL=Y	10,865	.92	3.77	.93	3.62
SWD/SUA	SUA=504 plan codes	10,159	.91	3.72	.92	3.61
ELL/SUA	SUA & ELL codes	3,726	.92	3.72	.92	3.59
ELL Test Language	English	7,563	.92	3.78	.92	3.63
	Chinese	480	.94	3.81	.94	3.57
	Haitian-Creole	77	.87	3.54	.87	3.45
	Korean	27	.86	3.29	.87	3.18
	Russian	72	.94	3.96	.94	3.74
	Spanish	2,597	.86	3.66	.87	3.57
	All Translations	3,253	.93	3.76	.94	3.59

Table 35E. Mathematics Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	126,260	.94	3.98	.95	3.67
Gender	Female	61,061	.94	3.97	.95	3.66
	Male	65,199	.94	3.97	.95	3.66
Ethnicity	Asian	12,210	.95	3.66	.96	3.37
	Black	25,734	.92	3.91	.93	3.69
	Hispanic	33,018	.93	3.95	.93	3.71
	American Indian	696	.93	3.97	.94	3.70
	Multiracial	1,328	.94	3.98	.95	3.69
	Pacific Islander	279	.95	3.93	.95	3.62
	White	52,995	.94	3.95	.95	3.67
NRC	New York	50,074	.95	3.94	.96	3.62
	Big 4 Cities	5,443	.91	3.84	.92	3.63
	Urban/Suburban	9,932	.91	3.94	.92	3.70
	Rural	6,733	.92	4.02	.93	3.75
	Average Needs	29,819	.93	4.01	.94	3.72
	Low Needs	16,696	.93	3.82	.94	3.57
	Charter	4,753	.93	3.94	.94	3.71
	Non-Public	2,710	.94	3.98	.95	3.69
SWD	All Codes	19,914	.89	3.73	.90	3.58
SUA	All Codes	14,557	.91	3.81	.92	3.63
ELL	ELL=Y	9,460	.91	3.73	.92	3.57
SWD/SUA	SUA=504 plan codes	8,877	.88	3.71	.89	3.58
ELL/SUA	SUA & ELL codes	3,038	.91	3.67	.92	3.52
ELL Test Language	English	6,607	.90	3.70	.90	3.57
	Chinese	423	.93	3.75	.94	3.54
	Haitian-Creole	63	.79	3.59	.80	3.51
	Korean	27	.96	3.53	.96	3.24
	Russian	101	.92	4.05	.93	3.80
	Spanish	2,218	.83	3.61	.84	3.52
	All Translations	2,832	.93	3.78	.94	3.56

Table 35F. Mathematics Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Students	95,280	.94	4.12	.94	3.86
Gender	Female	45,138	.93	4.13	.94	3.88
	Male	50,142	.94	4.10	.94	3.84
Ethnicity	Asian	8,010	.95	3.93	.95	3.63
	Black	21,956	.92	4.03	.93	3.83
	Hispanic	27,105	.92	4.09	.93	3.87
	American Indian	537	.93	4.09	.93	3.87
	Multiracial	955	.94	4.11	.95	3.84
	Pacific Islander	206	.94	4.10	.95	3.78
	White	36,511	.93	4.14	.94	3.91
NRC	New York	39,652	.94	4.10	.95	3.82
	Big 4 Cities	4,561	.92	3.86	.93	3.67
	Urban/Suburban	7,935	.91	4.01	.91	3.84
	Rural	5,586	.91	4.13	.92	3.93
	Average Needs	19,669	.92	4.17	.93	3.95
	Low Needs	9,257	.93	4.07	.94	3.84
	Charter	3,253	.94	4.07	.94	3.83
	Non-Public	5,284	.94	4.10	.95	3.84
SWD	All Codes	17,262	.90	3.83	.90	3.69
SUA	All Codes	11,860	.91	3.89	.92	3.72
ELL	ELL=Y	7,954	.92	3.85	.93	3.67
SWD/SUA	SUA=504 plan codes	7,556	.90	3.82	.90	3.68
ELL/SUA	SUA & ELL codes	2,205	.89	3.67	.90	3.55
ELL Test Language	English	5,344	.91	3.83	.92	3.68
	Chinese	456	.94	3.90	.95	3.59
	Haitian-Creole	57	.85	3.73	.86	3.59
	Korean	20	.93	4.04	.94	3.76
	Russian	73	.92	3.97	.93	3.78
	Spanish	1,974	.86	3.69	.87	3.61
	All Translations	2,580	.93	3.87	.94	3.64

Standard Error of Measurement

The SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Table 27 and Table 29 for ELA and mathematics, respectively. The SEMs ranged from 3.11 to 4.12 across subjects, grades, and the two methods of estimation, which is reasonable and small. The SEMs are directly related to reliability: the higher the reliability, the lower the standard error. As discussed, the reliability of these tests is relatively high, so it was expected that the SEMs would be very low.

The SEMs for subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Table 34A–36F and Table 35A–37F. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations ranged from 2.68 to 4.17, which is acceptably close to those for the entire population. This narrow range indicates that across the Grades 3–8 Common Core ELA and Mathematics Tests, all students' test scores are reasonably reliable with minimal error.

Performance Level Classification Consistency and Accuracy

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

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

Classification consistency is most relevant for students whose proficiency is near the pass/fail cut score. For example, consider the cut score delineating Levels II and III or simply the "Level III Cut." Students whose proficiency is far above or far below that cut score are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Examinees whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency; these numbers show the number of students who are most likely to be misclassified. Scoring tables with SEMs are located in Section 6, "IRT Calibration and Scaling," and student scale score frequency distributions are located in Appendix Q. Classification consistency and accuracy were estimated using the IRT procedure

suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000). Appendix P includes a description of the calculations and procedure based on the paper by Lee et al. (2002).

Consistency

The results for classifying students into four performance levels are separated from results based solely on the Level III cut. Table 36 and Table 37 include case counts (n-count), classification consistency (Agreement), classification inconsistency (Inconsistency), and Cohen's kappa (Kappa). Consistency indicates the rate that a second administration would yield the same performance category designation (or a different designation for the inconsistency rate). The agreement index is a sum of the diagonal element in the contingency table. Kappa is similar, but corrects for chance agreement. The inconsistency index is equal to the "1 - agreement index."

Table 36 depicts the ELA and mathematics consistency study results based on the range of performance levels for all grades. For ELA, 70–74% of students were estimated to be classified consistently to one of the four performance categories with a hypothetical second administration. Kappa–which corrects for chance agreement–ranged from 0.58 to 0.63. These are between "moderate" and "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 74% and 80% of students were estimated to be classified consistently to one of the four performance categories. Kappa, which indicates the consistency of the placement in the absence of chance, ranged from 0.66 to 0.71. These are all considered "substantial" agreement, by Landis and Koch's (1977) rules of thumb for kappa. As mentioned above and for all tests, there is an acceptable amount of measurement error that all scores contain and by random chance, students testing twice may be classified first, for example, as a Level III and second as a Level IV. This is expected to occur more often for students scoring around the selected cut score and less so for students closer to the middle of the performance level (i.e., close to the mid-point of two adjacent cut scores).

Table 36. Decision Consistency (All Cuts)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	156,480	71%	29%	0.59
4	151,880	70%	30%	0.58
5	149,480	73%	27%	0.61
6	148,800	73%	27%	0.61
7	137,660	74%	26%	0.62
8	133,180	73%	27%	0.63
Mathematics				
3	155,200	74%	26%	0.66
4	148,800	78%	22%	0.70
5	143,920	76%	24%	0.68
6	141,880	78%	22%	0.70
7	126,260	79%	21%	0.71
8	95,280	80%	20%	0.70

Table 37 depicts the ELA and mathematics consistency study results based on two performance levels (NYS Level II and NYS Level III) as defined by the Level III cut. For ELA, 88% and 90% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.73 to 0.77. These are considered "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. In addition, 90% and 94% of the mathematics classifications of individual students are estimated to remain stable with a second administration. Kappa coefficients for classification consistency based on the Level III cut ranged from 0.80 to 0.85. As with ELA, these statistics indicate at least "substantial" agreement (where kappa > 0.60) and some indicating "almost perfect" agreement (where kappa > 0.80), as per Landis and Koch's (1977) rules of thumb for kappa.

Table 37. Decision Consistency (Level III Cut)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	156,480	88%	12%	0.73
4	151,880	89%	11%	0.74
5	149,480	90%	10%	0.75
6	148,800	90%	10%	0.75
7	137,660	90%	10%	0.76
8	133,180	90%	10%	0.77
Mathematics				
3	155,200	90%	10%	0.80
4	148,800	92%	8%	0.84
5	143,920	91%	9%	0.82
6	141,880	93%	7%	0.84
7	126,260	93%	7%	0.85
8	95,280	94%	6%	0.82

Accuracy

The results of classification accuracy for ELA and mathematics across all grades are presented in Table 38. Included in the table are case counts (n-count) and classification accuracy (Accuracy) for all performance levels (All Cuts) and for the Level III cut score. Note that, by definition, accuracy associated with the Level III cut is at least as great as that with the entire set of cut scores because there are only two categories for the former, as opposed to the latter, which has four.

For ELA, the estimated accuracy rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her underlying proficiency from 77% to 80% of the time across all performance levels and 91% to 93% of the time in regard to the Level III cut score. For mathematics, the estimated accuracy rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her true proficiency

from 82% to 86% of the time across all performance levels and from 93% to 96% of the time in regard to the Level III cut score.

Table 38. Decision Agreement (Accuracy) Estimates

Grade	N-Count	Accuracy	
		All Cuts	Level III Cut
ELA			
3	156,480	78%	91%
4	151,880	77%	91%
5	149,480	79%	93%
6	148,800	80%	93%
7	137,660	80%	93%
8	133,180	80%	93%
Mathematics			
3	155,200	82%	93%
4	148,800	85%	95%
5	143,920	83%	94%
6	141,880	84%	94%
7	126,260	86%	96%
8	95,280	86%	96%

Section 8: Summary of Operational Test Results

This section summarizes the distribution of scale score scale score results on the NYSTP 2014 Grades 3–8 Common Core ELA and Mathematics Tests. These include the scale score means, standard deviations, percentile ranks, and performance level distributions for each grade's population and specific subgroups. Gender, ethnic identification, NRC, ELL, SWD, and SUA variables were used to calculate the results of subgroups required for federal reporting and test equity purposes for both the ELA and mathematics tests. Additionally, the ELL/SUA subgroup is defined as English language learners who use one or more ELL-related accommodations. The SWD/SUA subgroup is defined as examinees with disabilities using one or more disability-related accommodations falling under the 504 Plan. For the mathematics analyses, the test translation language is also indicated. (Recall that the ELA tests are not translated, as they are a measure of mastery of the English language.) ELA and mathematics data include examinees with valid scores from all public, private, and charter schools. Note that complete scale score frequency distribution tables for ELA and mathematics are located in Appendix Q.

Scale Score Distribution Summary

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

ELA Scale Score and Subscore Distributions

Table 39 shows some key statistics characterizing the distribution of ELA scale scores, while Table 40 summarizes the ELA subscores derived from the test in each grade. Table 41A–43F break down the scale scores by selected subgroups. Some general observations from these tables include: Females outperformed Males; Asian and White students outperformed their peers from other reported ethnic groups; students from Low Needs (as identified by NRC) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter); and ELL students, SWD, and/or SUA achieved below the State population (All Students) in every percentile rank. This pattern of achievement was consistent across all grades.

Table 39. ELA Scale Score Distribution Summary

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10th	25th	50th	75th	90th
3	177,519	297.55	37.57	249	274	300	323	344
4	178,492	299.22	37.25	252	278	304	325	343
5	170,998	297.09	37.70	251	276	301	323	339
6	171,859	298.39	36.24	250	276	302	325	340
7	164,563	293.47	36.91	246	271	297	319	336
8	163,167	297.46	37.64	250	278	301	323	339

Table 40. ELA Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Reading	177,519	31	17.48	6.07
	Writing	177,519	24	10.21	4.84
4	Reading	178,492	31	17.64	5.75
	Writing	178,492	24	12.29	5.27
5	Reading	170,998	42	25.58	7.93
	Writing	170,998	24	14.74	5.23
6	Reading	171,859	42	23.79	7.80
	Writing	171,859	24	14.60	5.79
7	Reading	164,563	42	23.13	7.73
	Writing	164,563	24	15.16	5.66
8	Reading	163,167	42	24.47	8.17
	Writing	163,167	24	16.12	5.66

ELA Grade 3

Scale score statistics and n-counts of demographic subgroups for Grade 3 are presented in Table 41A. The population scale score mean was 297.55 with a standard deviation of 37.57. Female students tended to outperform male students by around 10 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (314.74). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about two-thirds of a standard deviation below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and English language learners (ELL) subgroups scored, on average, 0.87 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed, scoring about 43 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (300): Female (306), Asian (316), Multiracial (303), Pacific Islander (306), and White (306) students and students from Average (303) and Low Needs (320) districts and Charter schools (306).

Table 41A. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	177,519	297.55	37.57	249	274	300	323	344
Gender	Female	87,135	302.69	36.24	254	281	306	326	347
	Male	90,384	292.59	38.15	239	270	297	320	336
Ethnicity	Asian	17,473	314.74	34.27	270	297	316	336	358
	Black	33,584	288.14	36.85	239	266	291	313	333
	Hispanic	50,097	289.41	35.15	244	270	294	313	333
	American Indian	1,175	294.03	36.40	244	274	297	320	336
	Multiracial	3,914	300.14	38.86	249	277	303	326	347
	Pacific Islander	630	303.17	35.62	258	284	306	329	344
	White	70,646	303.41	37.68	254	284	306	329	347
NRC	New York	70,267	298.16	36.67	249	277	300	323	344
	Big 4 Cities	7,533	273.04	40.12	219	244	274	300	323
	Urban/Suburban	13,989	282.64	36.93	233	258	287	310	326
	Rural	8,960	287.42	38.45	233	262	291	316	333
	Average Needs	39,365	299.68	36.61	249	277	303	326	344
	Low Needs	17,907	314.84	32.34	274	297	320	336	351
	Charter	9,227	304.56	32.15	262	284	306	326	344
	Non-Public	10,088	295.96	39.51	244	274	300	323	340
SWD	All Codes	26,818	263.71	38.54	211	239	266	291	313
SUA	All Codes	20,804	268.85	39.15	219	244	270	297	316
ELL	ELL=Y	16,454	268.80	34.52	219	244	270	294	310
SWD/SUA	SUA=504 plan codes	10,944	254.23	38.18	203	227	254	281	303
ELL/SUA	SUA & ELL codes	5,506	266.06	33.44	219	244	266	291	306

ELA Grade 4

Table 41B contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 299.22 with a standard deviation of 37.25. Female students tended to outperform male students by around 9 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Average and Low Needs districts and Charter and non-Public schools. Across ethnic groups, Asian students earned the highest mean score (316.83). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. Students with 504 plans were the lowest performing subgroup analyzed, scoring about 44 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (304): Female (307), Asian (320), Multiracial (307), Pacific Islander (310), White (310) students and those from Average (307) or Low Needs (320), districts or Non-Public (307) schools.

Table 41B. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	178,492	299.22	37.25	252	278	304	325	343
Gender	Female	87,880	303.86	35.72	256	282	307	329	347
	Male	90,612	294.72	38.14	242	271	298	322	339
Ethnicity	Asian	17,351	316.83	34.79	271	298	320	339	356
	Black	34,311	288.10	35.62	242	268	291	313	332
	Hispanic	48,884	289.79	34.79	242	268	291	313	332
	American Indian	1,100	293.95	36.82	247	271	295	321	339
	Multiracial	3,225	302.81	39.86	252	278	307	332	351
	Pacific Islander	520	305.22	34.91	260	287	310	329	347
	White	73,101	306.43	36.78	256	287	310	332	347
NRC	New York	69,644	298.81	36.62	252	275	301	322	343
	Big 4 Cities	7,114	272.87	39.76	218	247	275	301	322
	Urban/Suburban	13,131	283.03	36.34	237	260	287	310	325
	Rural	8,433	288.91	38.29	237	268	291	316	335
	Average Needs	38,546	302.54	36.17	256	282	307	329	343
	Low Needs	18,168	317.78	31.43	278	301	320	339	351
	Charter	8,023	301.14	30.27	260	282	304	322	335
	Non-Public	15,287	301.52	37.35	252	278	307	329	343
SWD	All Codes	28,137	264.57	37.42	210	237	264	291	313
SUA	All Codes	20,707	267.10	38.21	218	242	268	295	316
ELL	ELL=Y	14,771	266.11	34.56	218	242	268	291	310
SWD/SUA	SUA=504 plan codes	11,785	255.65	37.45	202	232	256	282	304
ELL/SUA	SUA & ELL codes	4,648	262.14	33.59	218	242	264	287	304

ELA Grade 5

Scale score summary statistics by key demographic subgroups for Grade 5 students are in Table 41C. The population scale score mean was 297.09 with a standard deviation of 37.70. Female students tended to outperform male students by around 11 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs, and Low Needs districts. Across all ethnic groups, Asian students earned the highest mean score (316.08). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, one standard deviations below the mean scale score for the population. Students with 504 plans were the lowest performing subgroup analyzed, scoring about 45 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (301): Female (306), Asian (320), Multiracial (306), Pacific Islander (306), and White (309) students and those from Average (303) or Low Needs (317) districts.

Table 41C. ELA Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	170,998	297.09	37.70	251	276	301	323	339
Gender	Female	84,049	302.62	35.54	257	281	306	326	346
	Male	86,949	291.74	38.93	241	271	296	320	336
Ethnicity	Asian	17,510	316.08	35.19	273	298	320	339	357
	Black	33,109	285.73	36.06	241	265	289	309	329
	Hispanic	45,428	288.11	35.37	244	268	291	311	329
	American Indian	1,052	291.39	37.91	247	273	293	317	332
	Multiracial	2,573	301.55	39.57	251	278	306	329	348
	Pacific Islander	427	303.75	33.22	262	286	306	326	339
	White	70,899	303.33	37.24	257	283	309	329	346
NRC	New York	66,086	298.50	36.57	254	276	301	323	346
	Big 4 Cities	6,843	267.65	41.59	209	244	271	296	320
	Urban/Suburban	12,373	281.44	37.85	233	260	283	309	326
	Rural	8,182	286.64	37.47	237	265	291	311	332
	Average Needs	38,476	299.94	36.07	254	281	303	326	339
	Low Needs	19,273	315.39	31.16	278	298	317	336	352
	Charter	9,141	295.95	30.30	257	278	298	317	332
	Non-Public	10,466	290.98	42.68	233	271	298	320	336
SWD	All Codes	28,543	261.30	38.78	209	241	265	289	306
SUA	All Codes	21,094	263.01	39.88	209	241	268	291	311
ELL	ELL=Y	13,356	258.28	36.73	209	237	262	283	301
SWD/SUA	SUA=504 plan codes	12,542	252.22	39.69	193	229	257	281	301
ELL/SUA	SUA & ELL codes	4,173	253.61	36.47	201	233	257	278	296

ELA Grade 6

Table 41D contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 298.39 with a standard deviation of 36.24. Female students tended to outperform male students by around 12 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts. Across ethnic groups, Asian students earned the highest mean score (317.23). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 43 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (302): Female (307), Asian (320), Multiracial (304), Pacific Islander (304), and White (309) students and those from Average (304) or Low Needs (320) districts or Non-Public (304) schools.

Table 41D. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	171,859	298.39	36.24	250	276	302	325	340
Gender	Female	84,061	304.35	34.20	259	283	307	328	344
	Male	87,798	292.68	37.20	243	268	295	320	338
Ethnicity	Asian	17,104	317.23	34.46	270	297	320	340	357
	Black	33,882	287.29	34.29	243	265	290	312	328
	Hispanic	45,550	289.40	33.19	247	268	290	312	331
	American Indian	1,135	292.42	35.23	247	273	295	317	338
	Multiracial	2,384	301.25	38.28	250	276	304	331	348
	Pacific Islander	488	302.66	35.64	259	283	304	328	344
	White	71,316	304.85	36.09	257	285	309	331	348
NRC	New York	64,744	298.87	35.43	253	276	300	322	344
	Big 4 Cities	6,684	272.72	37.33	227	247	273	300	320
	Urban/Suburban	12,059	282.55	35.41	236	259	285	307	325
	Rural	8,124	288.96	36.43	240	268	293	314	334
	Average Needs	37,218	301.31	35.33	253	281	304	325	344
	Low Needs	18,602	315.94	30.97	276	300	320	338	352
	Charter	9,195	296.18	30.45	257	276	297	317	334
	Non-Public	15,049	298.05	38.11	247	278	304	325	340
SWD	All Codes	27,626	264.01	33.91	222	243	265	285	307
SUA	All Codes	19,822	264.93	35.44	222	243	265	288	309
ELL	ELL=Y	11,985	259.33	30.98	222	240	259	281	297
SWD/SUA	SUA=504 plan codes	12,348	256.71	34.34	214	236	257	281	300
ELL/SUA	SUA & ELL codes	3,452	255.77	29.83	222	236	257	276	293

ELA Grade 7

Scale score statistics and n-counts of demographic subgroups for Grade 7 are presented in Table 41E. The population scale score mean was 293.47 with a standard deviation of 36.91. Female students tended to outperform male students by around 13 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from New York City, Average and Low Needs districts. Across ethnic groups, Asian students earned the highest mean score (312.65). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 45 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (297): Female (301), Asian (318), Multiracial (299), Pacific Islander (304), and White (306) students and those from Average (299) and Low Needs (318) districts.

Table 41E. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	164,563	293.47	36.91	246	271	297	319	336
Gender	Female	80,270	299.94	33.99	256	279	301	324	340
	Male	84,293	287.31	38.49	239	265	292	313	333
Ethnicity	Asian	16,353	312.65	33.94	268	294	318	336	352
	Black	33,845	281.99	34.49	239	262	284	306	321
	Hispanic	43,591	285.13	33.08	242	265	287	306	324
	American Indian	1,001	282.77	38.56	235	262	287	309	327
	Multiracial	1,861	292.46	41.91	239	268	299	321	340
	Pacific Islander	404	301.81	34.08	259	282	304	324	340
	White	67,508	300.10	37.62	250	279	306	327	343
NRC	New York	65,317	294.79	34.35	253	274	297	319	336
	Big 4 Cities	6,448	264.14	40.28	210	241	265	292	313
	Urban/Suburban	11,566	276.22	36.87	230	253	279	301	319
	Rural	7,993	284.53	38.06	235	262	289	311	330
	Average Needs	34,962	295.21	36.88	250	274	299	321	336
	Low Needs	19,520	312.10	31.42	274	294	318	333	347
	Charter	8,039	293.23	29.22	256	277	294	313	327
	Non-Public	10,582	289.10	41.80	230	271	297	318	333
SWD	All Codes	26,880	260.84	35.55	218	242	262	284	301
SUA	All Codes	18,246	260.00	37.76	210	239	262	287	306
ELL	ELL=Y	11,239	253.58	33.56	210	235	256	277	292
SWD/SUA	SUA=504 plan codes	11,896	252.54	36.94	202	235	256	277	297
ELL/SUA	SUA & ELL codes	2,730	248.08	31.74	210	230	250	268	287

ELA Grade 8

Table 41F contains Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 297.46 with a standard deviation of 37.64. Female students tended to outperform male students by around 13 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts and Charter and Non-Public schools. Across ethnic groups, Asian students earned the highest mean score (316.13). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 51 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (301): Female (308), Asian (320), Multiracial (303), Pacific Islander (310), and White (310) students and students enrolled in Average (303) and Low Needs (320) districts and Non-Public (305) schools.

Table 41F. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	163,167	297.46	37.64	250	278	301	323	339
Gender	Female	79,113	304.40	34.98	259	285	308	328	343
	Male	84,054	290.93	38.87	243	270	296	317	335
Ethnicity	Asian	16,494	316.13	34.62	273	299	320	339	353
	Black	34,363	286.08	35.33	243	268	290	310	325
	Hispanic	42,983	289.34	33.93	246	270	294	312	328
	American Indian	968	288.37	40.16	235	268	294	312	332
	Multiracial	1,684	297.73	41.73	246	275	303	325	343
	Pacific Islander	409	308.16	34.41	265	287	310	332	348
	White	66,266	304.05	38.45	256	285	310	328	348
NRC	New York	65,958	298.21	34.29	256	278	301	320	339
	Big 4 Cities	6,412	268.37	42.04	210	243	273	299	320
	Urban/Suburban	11,274	280.92	38.29	231	259	285	308	325
	Rural	7,991	288.32	40.05	235	268	294	316	332
	Average Needs	33,353	299.37	38.86	250	278	303	325	343
	Low Needs	18,377	315.56	33.62	273	299	320	335	353
	Charter	6,050	297.63	27.86	262	280	301	316	328
	Non-Public	13,563	297.69	39.68	246	280	305	323	339
SWD	All Codes	25,822	262.79	36.74	210	243	265	287	305
SUA	All Codes	17,487	261.70	39.80	210	239	265	290	310
ELL	ELL=Y	10,201	254.83	35.24	202	235	259	280	296
SWD/SUA	SUA=504 plan codes	11,518	254.66	38.93	202	231	259	280	301
ELL/SUA	SUA & ELL codes	2,417	246.59	34.35	202	226	246	270	290

Mathematics Scale Score Distributions

Table 42 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 43 summarizes the mathematics subscores derived from the test in each grade. Table 44A–44F break down the scale scores by selected subgroups. Some general observations from the mathematics data are as follows: Female and Male students performed fairly consistently; Asian students scored considerably higher than other reported ethnic groups; Low and Average Needs schools (as identified by the NRC) outperformed most other school types (New York City, Big 4 Cities, High Needs Urban/Suburban, and Rural districts), with Private and Charter Schools sometimes also outperforming other school types. Students taking the Chinese and Korean translations met or exceeded the population at most reported percentile ranks, whereas the other translation subgroups (Haitian-Creole, Spanish, and Russian) were below the population scale score at each percentile rank; and ELLs, SWDs, and/or SUAs achieved below the State mean in most percentile rank ranks. This pattern of achievement was consistent across all grades.

Table 42. Mathematics Scale Score Distribution Summary

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10th	25th	50th	75th	90th
3	176,720	304.36	37.10	258	280	304	329	354
4	176,807	303.84	40.13	250	278	305	331	351
5	167,821	307.64	39.17	258	286	310	333	354
6	166,508	303.89	39.86	252	280	306	332	353
7	156,113	303.72	38.50	252	282	307	331	350
8	124,506	293.23	38.37	244	270	296	319	338

Table 43. Mathematics Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Operations and Algebraic Thinking	176,720	27	17.47	6.50
	Number and Operations—Fractions	176,720	12	6.89	3.02
	Measurement and Data	176,720	12	7.43	2.70
4	Operations and Algebraic Thinking	176,807	11	6.57	3.31
	Number and Operations in Base Ten	176,807	17	10.95	4.38
	Number and Operations—Fractions	176,807	18	9.74	4.98
5	Number and Operations in Base Ten	167,821	18	11.10	4.09
	Number and Operations—Fractions	167,821	25	11.74	5.55
	Measurement and Data	167,821	10	6.88	2.67
6	Ratios and Proportional Relationships	166,508	18	10.26	4.75
	The Number System	166,508	12	6.87	3.12
	Expressions and Equations	166,508	28	14.24	7.18
7	Ratios and Proportional Relationships	156,113	20	9.00	5.34
	The Number System	156,113	14	7.86	3.54
	Expressions and Equations	156,113	22	10.60	5.52
8	Expressions and Equations	124,506	30	14.36	7.39
	Functions	124,506	19	8.44	4.51
	Geometry	124,506	12	6.17	3.48

Mathematics Grade 3

Scale score statistics and n-counts of demographic subgroups for Grade 3 are presented in Table 44A. The population scale score mean was 304.36 with a standard deviation of 37.10. 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 Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (325.87). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.74 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 34 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (304): Asian (326), Multiracial (308), Pacific Islander (310), and White (314) students and those enrolled in Average (310) or Low Needs (323) districts or Charter (314) schools. In terms of the 50th-percentile ranks for students using translated forms, they ranged from 269 (Spanish, n = 2,611) to 329 (Korean, n = 24).

Table 44A. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	176,720	304.36	37.10	258	280	304	329	354
Gender	Female	86,474	304.43	35.70	261	280	304	329	348
	Male	90,246	304.29	38.40	255	280	304	329	354
Ethnicity	Asian	17,971	325.87	35.66	280	302	326	348	369
	Black	33,418	291.68	35.76	248	269	291	315	336
	Hispanic	50,556	293.71	34.07	252	273	293	315	336
	American Indian	1,170	300.44	35.56	255	278	299	323	344
	Multiracial	3,787	308.06	38.06	261	285	308	332	360
	Pacific Islander	637	310.81	35.21	266	287	310	336	360
	White	69,181	312.49	35.49	269	291	314	336	354
NRC	New York	71,492	302.52	36.84	258	278	302	326	348
	Big 4 Cities	7,580	279.18	36.79	236	255	278	304	326
	Urban/Suburban	13,747	289.32	34.53	248	266	289	314	332
	Rural	8,625	299.97	34.38	258	278	302	323	344
	Average Needs	38,098	308.66	34.75	266	287	310	332	354
	Low Needs	17,530	324.21	33.52	282	304	323	348	369
	Charter	9,212	315.76	35.56	271	291	314	340	360
	Non-Public	10,264	299.68	37.52	252	278	302	323	344
SWD	All Codes	26,467	276.86	36.77	231	252	276	299	323
SUA	All Codes	20,970	279.15	36.39	236	255	280	302	326
ELL	ELL=Y	18,488	279.38	35.22	236	258	278	301	323
SWD/SUA	SUA=504 plan codes	10,740	270.21	36.38	225	244	269	293	315
ELL/SUA	SUA & ELL codes	6,289	274.06	33.57	231	252	273	295	315
ELL Test Language	English	15,119	279.64	34.86	236	258	278	302	323
	Chinese	615	316.59	32.88	273	295	315	340	360
	Haitian-Creole	42	270.12	24.09	240	258	273	280	302
	Korean	24	330.04	31.70	287	306	329	360	369
	Russian	77	300.29	33.44	258	276	295	318	348
	Spanish	2,611	268.19	30.85	231	248	269	289	306
	All Translations	3,369	278.23	36.76	236	255	276	299	326

Mathematics Grade 4

Table 44B contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 303.84 with a standard deviation of 40.13. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (329.09). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about three-quarters of a standard deviation

below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.81 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 39 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (305): Male (306), Asian (331), Multiracial (311), Pacific Islander (310), and White (315) students and those enrolled in Average (314, Low (327) Needs districts or Charter (311) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 261 (Spanish, n = 2,645) to 325 (Korean, n = 31).

Table 44B. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	176,807	303.84	40.13	250	278	305	331	351
Gender	Female	86,670	303.81	39.15	254	280	305	329	351
	Male	90,137	303.87	41.04	250	278	306	331	355
Ethnicity	Asian	17,838	329.09	37.81	283	306	331	355	379
	Black	33,918	287.38	38.51	234	264	288	314	334
	Hispanic	49,337	291.92	37.30	246	269	293	317	337
	American Indian	1,081	299.91	40.48	246	276	301	327	351
	Multiracial	3,061	308.78	41.16	254	284	311	337	361
	Pacific Islander	524	309.17	36.84	264	286	310	334	355
	White	71,048	313.45	37.20	267	291	315	337	355
NRC	New York	70,868	301.99	40.85	250	276	301	329	355
	Big 4 Cities	7,017	274.73	40.83	218	246	276	303	327
	Urban/Suburban	12,716	284.89	38.82	234	261	286	311	331
	Rural	7,974	298.52	37.99	250	276	301	325	343
	Average Needs	36,960	310.04	36.95	264	288	314	334	355
	Low Needs	17,829	325.81	34.04	284	306	327	346	368
	Charter	8,008	311.07	37.00	264	288	311	334	355
	Non-Public	15,302	299.99	36.30	254	278	301	323	343
SWD	All Codes	27,689	271.61	38.34	218	246	272	296	321
SUA	All Codes	20,028	273.37	39.07	218	246	274	300	323
ELL	ELL=Y	16,791	274.53	37.94	226	250	274	300	323
SWD/SUA	SUA=504 plan codes	11,045	264.68	38.06	218	240	264	291	314
ELL/SUA	SUA & ELL codes	5,157	266.27	37.07	218	240	267	290	314
ELL Test Language	English	13,459	275.32	37.08	226	250	276	300	321
	Chinese	523	320.56	34.60	278	298	321	343	361
	Haitian-Creole	48	265.31	26.42	226	252	269	286	298
	Korean	31	328.19	33.69	291	308	325	351	368
	Russian	85	295.69	38.81	246	267	296	321	346
	Spanish	2,645	260.27	34.01	210	240	261	284	303
	All Translations	3,332	271.34	41.08	218	246	269	296	325

Mathematics Grade 5

Grade 5 demographic subgroup n-counts and scale score statistics are presented in Table 44C. The population scale score mean was 307.64 with a standard deviation of 39.17. 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 and Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (334.45). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by 0.84 standard deviations below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about 0.87 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 41 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (310): Asian (335), Multiracial (314), Pacific Islander (319), and White (320) students and those enrolled in Average (316) and Low Needs (330) districts. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 268 (Spanish, n = 2,492) to 335 (Korean, n = 33).

Table 44C. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	167,821	307.64	39.17	258	286	310	333	354
Gender	Female	82,165	308.94	36.79	261	288	310	333	351
	Male	85,656	306.40	41.29	254	283	310	335	354
Ethnicity	Asian	17,893	334.45	37.10	288	314	335	358	381
	Black	32,496	291.96	36.20	244	272	295	316	335
	Hispanic	45,632	296.28	35.83	249	275	300	320	340
	American Indian	992	302.73	36.96	258	280	304	326	348
	Multiracial	2,416	312.35	40.26	261	290	314	337	358
	Pacific Islander	432	315.99	35.75	268	295	319	337	358
	White	67,960	315.56	37.35	268	295	320	340	358
NRC	New York	67,140	308.14	39.01	258	283	308	333	354
	Big 4 Cities	6,732	276.24	40.60	223	249	278	304	328
	Urban/Suburban	11,835	288.78	37.99	238	265	294	314	333
	Rural	7,580	300.67	37.11	254	280	304	324	346
	Average Needs	36,298	312.32	36.06	268	294	316	335	354
	Low Needs	18,538	328.46	32.70	290	310	330	348	368
	Charter	9,152	307.87	33.03	265	288	308	330	348
	Non-Public	10,409	297.71	42.09	238	275	304	326	346
SWD	All Codes	27,493	274.33	38.28	223	249	278	300	322
SUA	All Codes	19,890	276.12	39.66	223	249	278	304	324
ELL	ELL=Y	15,219	276.00	38.89	223	254	278	300	324
SWD/SUA	SUA=504 plan codes	11,481	266.65	38.77	215	238	268	294	314
ELL/SUA	SUA & ELL codes	4,565	268.62	37.99	215	244	272	294	314

Table 44C. Mathematics Grade 5 Scale Score Distribution by Subgroup (cont.)

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
ELL Test Language	English	12,051	276.11	38.22	223	254	278	300	322
	Chinese	508	321.19	34.78	278	297	324	344	362
	Haitian-Creole	44	270.14	28.96	238	258	272	287	304
	Korean	33	336.85	37.59	297	319	335	358	381
	Russian	91	295.40	36.49	254	278	295	319	342
	Spanish	2,492	264.84	35.20	215	244	268	288	306
	All Translations	3,168	275.58	41.31	223	249	278	302	328

Mathematics Grade 6

Table 44D contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 303.89 with a standard deviation of 39.86. 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 enrolled in Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (331.35). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.87 standard deviations below the mean scale score for the population. Students with 504 plans were the lowest-performing subgroup analyzed for English forms, scoring about 39 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (306): Female (308), Asian (334), Multiracial (311), Pacific Islander (311), and White (316) students and those enrolled in Average (313), Low (330) Needs districts. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 254 (Haitian-Creole, n = 84) to 342 (Korean, n = 36).

Table 44D. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	166,508	303.89	39.86	252	280	306	332	353
Gender	Female	81,249	306.12	38.49	256	282	308	332	353
	Male	85,259	301.77	41.01	247	275	305	330	353
Ethnicity	Asian	17,384	331.35	37.30	284	310	334	357	376
	Black	32,928	286.75	38.01	241	263	289	313	334
	Hispanic	45,293	291.50	36.91	247	270	292	316	336
	American Indian	1,081	297.84	37.95	252	275	300	321	345
	Multiracial	2,197	309.27	41.17	256	284	311	338	360
	Pacific Islander	491	310.24	39.48	263	286	311	336	360
	White	67,134	313.42	36.44	267	292	316	338	357

Table 44D. Mathematics Grade 6 Scale Score Distribution by Subgroup (cont.)

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
NRC	New York	65,739	301.60	41.28	252	275	301	330	353
	Big 4 Cities	6,414	275.97	40.09	223	252	278	303	326
	Urban/Suburban	11,185	283.96	38.31	233	260	286	310	332
	Rural	7,445	297.66	35.62	252	278	301	321	340
	Average Needs	34,054	310.15	35.73	263	291	313	334	353
	Low Needs	17,505	327.94	33.39	286	310	330	350	370
	Charter	9,136	305.27	34.52	263	284	306	328	347
	Non-Public	14,865	300.83	37.24	252	280	303	326	345
SWD	All Codes	26,115	268.02	37.43	214	247	270	292	315
SUA	All Codes	19,215	273.58	38.95	223	252	275	300	323
ELL	ELL=Y	13,788	270.05	37.92	223	247	270	294	318
SWD/SUA	SUA=504 plan codes	11,496	265.32	37.55	214	241	267	291	311
ELL/SUA	SUA & ELL codes	4,122	267.01	37.51	214	247	267	291	315
ELL Test Language	English	9,975	269.78	36.88	223	247	270	294	316
	Chinese	593	319.93	34.32	280	300	319	342	360
	Haitian-Creole	84	249.62	36.05	206	223	254	272	294
	Korean	36	333.58	33.31	305	322	342	352	370
	Russian	96	299.35	38.48	252	278	303	324	350
	Spanish	3,004	259.99	32.80	214	241	263	282	301
	All Translations	3,813	270.77	40.51	214	247	270	296	323

Mathematics Grade 7

The n-counts and scale score statistics for key demographic subgroups of Grade 7 students are presented in Table 44E. The population scale score mean was 303.72 with a standard deviation of 38.50. Female students tended to outperform male students by around 4 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (330.66). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by 0.80 standard deviations below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.88 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 39 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (307): Female (309), Asian (335), Multiracial (310), Pacific Islander (316) and White (318) students and those enrolled in Average (313) and Needs (329) or Charter (310) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 264 (Spanish, n = 2,793) to 342 (Korean, n = 32).

Table 44E. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	156,113	303.72	38.50	252	282	307	331	350
Gender	Female	75,976	305.55	37.27	258	285	309	332	350
	Male	80,137	301.98	39.54	252	279	305	329	350
Ethnicity	Asian	16,383	330.66	35.27	285	312	335	354	368
	Black	32,201	288.28	36.33	233	268	293	313	332
	Hispanic	42,925	291.87	35.98	244	272	294	316	335
	American Indian	902	293.36	37.92	244	268	296	319	339
	Multiracial	1,634	305.26	39.76	252	282	310	334	352
	Pacific Islander	396	312.36	37.57	264	293	316	339	354
	White	61,672	312.92	35.45	268	294	318	337	352
NRC	New York	66,103	302.65	39.28	252	279	304	329	352
	Big 4 Cities	5,952	273.66	38.74	218	252	276	300	321
	Urban/Suburban	10,332	283.31	36.94	226	264	287	309	328
	Rural	6,922	296.37	35.39	252	276	302	321	337
	Average Needs	30,548	308.31	35.10	264	290	313	332	348
	Low Needs	17,468	325.87	31.04	287	310	329	346	360
	Charter	7,992	307.15	32.78	264	287	310	329	346
	Non-Public	10,682	299.84	37.30	252	279	305	325	342
SWD	All Codes	24,783	269.52	36.35	218	244	272	294	315
SUA	All Codes	17,150	273.51	38.73	218	252	276	300	321
ELL	ELL=Y	13,228	269.71	38.08	218	244	272	294	318
SWD/SUA	SUA=504 plan codes	10,421	266.27	36.80	218	244	268	293	312
ELL/SUA	SUA & ELL codes	3,561	264.44	39.18	210	233	264	290	315
ELL Test Language	English	9,574	268.26	36.62	218	244	272	294	312
	Chinese	627	323.95	29.93	287	309	326	346	357
	Haitian-Creole	72	259.36	32.04	210	226	266	284	296
	Korean	32	330.28	41.62	276	317	342	354	368
	Russian	130	295.42	34.84	244	276	302	321	339
	Spanish	2,793	260.87	33.87	218	233	264	285	304
	All Translations	3,654	273.50	41.43	218	244	272	302	329

Mathematics Grade 8

That table contains Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 293.23 with a standard deviation of 38.37. Female students tended to outperform male students by around 5 scale score points. Asian, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts and Charter and Non-Public schools. Across ethnic groups, Asian students earned the highest mean score (320.17). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score – by 0.79

standard deviations below the population mean. The SWD, SUA, and ELL subgroups scored, on average, three-quarters of a standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed for English forms, scoring about 38 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (296): Female (297), Asian (322), Multiracial (297), Pacific Islander (307), and White (306) students and those enrolled in Average (301) and Low (315) Needs districts and Charter (307) and Non-Public (297) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 257 (Haitian-Creole, n = 72) to 324 (Chinese, n = 662).

Table 44F. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	124,506	293.23	38.37	244	270	296	319	338
Gender	Female	59,426	295.84	36.99	250	273	297	319	340
	Male	65,080	290.85	39.45	237	267	294	318	338
Ethnicity	Asian	11,113	320.17	39.22	270	297	322	345	367
	Black	29,475	280.47	36.97	227	259	282	306	326
	Hispanic	37,581	286.06	35.67	237	263	288	310	329
	American Indian	755	286.50	38.22	227	263	290	313	329
	Multiracial	1,183	295.20	40.02	244	273	297	322	343
	Pacific Islander	318	307.09	38.93	255	284	307	331	362
	White	44,081	301.04	35.98	255	282	306	324	343
NRC	New York	56,638	293.72	38.51	244	270	294	319	343
	Big 4 Cities	5,545	263.67	40.25	211	237	263	288	316
	Urban/Suburban	8,429	274.37	35.61	227	255	278	299	316
	Rural	5,809	289.04	33.90	244	270	292	312	327
	Average Needs	20,191	296.05	33.43	250	278	301	318	332
	Low Needs	9,573	312.95	33.61	270	296	315	332	351
	Charter	5,228	304.09	35.51	259	282	307	327	349
	Non-Public	12,967	294.84	39.55	244	273	297	321	340
SWD	All Codes	22,664	264.17	35.34	219	244	267	288	309
SUA	All Codes	14,101	265.98	37.35	219	244	267	292	313
ELL	ELL=Y	11,723	268.62	38.73	219	244	267	294	318
SWD/SUA	SUA=504 plan codes	9,025	261.14	36.13	211	237	263	287	307
ELL/SUA	SUA & ELL codes	2,560	255.43	35.43	211	237	255	275	299

Table 44F. Mathematics Grade 8 Scale Score Distribution by Subgroup (cont.)

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
ELL Test Language	English	8,216	266.63	37.67	219	244	267	292	313
	Chinese	662	322.09	35.97	278	299	324	345	367
	Haitian-Creole	72	256.76	33.64	211	237	257	280	296
	Korean	25	314.96	31.03	263	302	316	332	345
	Russian	122	287.16	34.55	244	267	291	310	326
	Spanish	2,626	260.39	31.57	219	244	263	282	301
	All Translations	3,507	273.28	40.74	219	244	270	297	329

Performance Level Distribution Summary

Students are classified as NYS Level I, NYS Level II, NYS Level III, and NYS Level IV. The cut scores were established in 2013 during the standard-setting. Table 24 and Table 25 show the ELA and mathematics cut scores, respectively, used for classification of students into the four performance-level categories in 2015. Please note that it is inappropriate to compare scale scores across grades as they neither measure the same content, nor are they on the same scale. During the standard-setting process, while cut scores were set separately for different grades within a subject, additional care was taken to vertically articulate performance levels; see Section 8 and Appendix P in the 2013 technical report (NYSED, 2014) for details. While vertical articulation helps to build consistent meaning to the performance levels, the very nature of grade-specific content, differing performance expectations, and panel set cut scores result in cut score differences across grades.

ELA Test Performance Level Distributions

Table 45 shows the performance level distribution for all examinees from public, charter, and private schools with valid ELA scores. Performance level data for selected subgroups of students were also examined. In general, these distributions reflect the same achievement trends in the scale score summary discussion. More Female students were classified in Level III and above categories than Male students. Similarly, more Asian and White students were classified in Level III and above categories than their peers from other reported ethnic groups. Consistent with the pattern shown in scale score distribution across the subgroups, students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, Urban/Suburban, and Rural). The Level III and above rates for students in the ELL, SWD, and SUA subgroups were low compared to the total population of examinees.

Table 45. ELA Test Performance Level Distributions

Grade	N-Count	Level I	Level II	Level III	Level IV	Level III & IV
3	177,519	37.08	31.93	26.17	4.81	30.99
4	178,492	30.57	36.47	21.55	11.42	32.97
5	170,998	35.30	35.08	19.98	9.63	29.61
6	171,859	30.18	39.05	16.35	14.41	30.76
7	164,563	37.62	33.19	23.68	5.52	29.19
8	163,167	28.99	36.00	25.32	9.70	35.02

ELA Grade 3

Performance level distributions and n-counts of demographic subgroups for ELA Grade 3 are presented in Table 46A. Statewide, a combined 31% of students achieved Level III and Level IV. About 36% of Female students were at Level III or above, as compared to 27% of Male students The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. About 49% of Asian students and 51% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–22% of students in those same performance categories. Only 7% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (31%): Female (36%), Asian (49%), Multiracial (34%), Pacific Islander (37%), and White (38%) students and those enrolled in Average (33%) and Low (51%) Needs districts or Charter (36%) schools.

Table 46A. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	177,519	37.08	31.93	26.17	4.81	30.99
Gender	Female	87,135	32.27	32.18	29.15	6.40	35.55
	Male	90,384	41.72	31.69	23.31	3.28	26.59
Ethnicity	Asian	17,473	20.00	30.65	38.50	10.86	49.36
	Black	33,584	47.77	31.10	18.74	2.38	21.12
	Hispanic	50,097	46.00	33.28	18.55	2.17	20.72
	American Indian	1,175	41.11	31.91	23.57	3.40	26.98
	Multiracial	3,914	34.29	31.25	28.51	5.95	34.47
	Pacific Islander	630	31.90	31.27	31.27	5.56	36.83
	White	70,646	30.03	31.72	31.94	6.30	38.24
NRC	New York	70,267	37.26	32.49	24.96	5.30	30.26
	Big 4 Cities	7,533	63.39	23.40	11.84	1.37	13.21
	Urban/Suburban	13,989	53.20	30.20	15.10	1.50	16.61
	Rural	8,960	47.30	30.46	19.97	2.28	22.24
	Average Needs	39,365	33.82	33.30	28.20	4.69	32.88
	Low Needs	17,907	18.26	31.07	42.23	8.44	50.67
	Charter	9,227	29.63	34.71	30.53	5.13	35.66
	Non-Public	10,088	37.61	31.74	25.97	4.68	30.65
SWD	All Codes	26,818	73.13	19.27	7.03	0.57	7.60
SUA	All Codes	20,804	67.18	23.13	8.96	0.73	9.69
ELL	ELL=Y	16,454	71.13	22.15	6.29	0.43	6.72
SWD/SUA	SUA=504 plan codes	10,944	80.63	14.75	4.42	0.20	4.62
ELL/SUA	SUA & ELL codes	5,506	74.63	20.36	4.72	0.29	5.01

ELA Grade 4

Performance level distributions and n-counts of demographic subgroups for ELA Grade 4 are presented in B. Statewide, a combined 31% of students achieved Level III and Level IV. About 36% of Female students were at Level III or above, as compared to 27% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. About 49% of Asian students and 51% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–22% of students in those same performance categories. Only 7% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (31%): Female (36%), Asian (49%), Multiracial (34%), Pacific Islander (37%), and White (38%) students and those enrolled in Average (33%) or Low (51%) Needs districts or Charter (36%) schools.

Table 46B. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	178,492	30.57	36.47	21.55	11.42	32.97
Gender	Female	87,880	26.07	36.73	23.67	13.53	37.20
	Male	90,612	34.93	36.21	19.49	9.37	28.86
Ethnicity	Asian	17,351	15.48	30.95	29.88	23.69	53.57
	Black	34,311	41.65	37.75	15.52	5.07	20.59
	Hispanic	48,884	39.47	39.53	15.78	5.22	21.00
	American Indian	1,100	36.64	36.27	18.09	9.00	27.09
	Multiracial	3,225	28.19	33.30	22.42	16.09	38.51
	Pacific Islander	520	24.23	37.50	25.58	12.69	38.27
	White	73,101	23.05	35.26	26.25	15.45	41.69
NRC	New York	69,644	31.82	36.87	19.86	11.46	31.32
	Big 4 Cities	7,114	58.94	28.21	9.25	3.60	12.85
	Urban/Suburban	13,131	46.88	36.54	12.60	3.98	16.58
	Rural	8,433	40.06	36.77	16.67	6.50	23.17
	Average Needs	38,546	26.31	37.18	24.43	12.08	36.51
	Low Needs	18,168	12.34	33.23	32.59	21.84	54.43
	Charter	8,023	25.55	43.28	24.09	7.08	31.17
	Non-Public	15,287	27.40	36.68	23.65	12.26	35.91
SWD	All Codes	28,137	68.25	24.43	5.88	1.44	7.32
SUA	All Codes	20,707	64.37	27.05	6.99	1.59	8.58
ELL	ELL=Y	14,771	68.35	25.73	5.02	0.91	5.92
SWD/SUA	SUA=504 plan codes	11,785	76.04	19.47	3.86	0.64	4.50
ELL/SUA	SUA & ELL codes	4,648	72.76	23.15	3.61	0.47	4.09

ELA Grade 5

Performance level distributions and n-counts of demographic subgroups for ELA Grade 5 are presented in Table 46C. Statewide, a combined 30% of students achieved Level III and Level IV. About 34% of Female students were at Level III or above, as compared to 25% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. About 52% of Asian students and 50% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 10–19% of students in those same performance categories. Only 5% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (30%): Female (34%), Asian (52%), Multiracial (36%), Pacific Islander (35%), and White (37%) students and those enrolled in Average (32%) or Low (50%) Needs districts.

Table 46C. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	170,998	35.30	35.08	19.98	9.63	29.61
Gender	Female	84,049	29.79	35.84	22.48	11.89	34.37
	Male	86,949	40.63	34.35	17.57	7.45	25.02
Ethnicity	Asian	17,510	16.54	31.55	29.95	21.95	51.91
	Black	33,109	48.07	34.81	13.09	4.04	17.12
	Hispanic	45,428	44.87	36.39	14.46	4.27	18.73
	American Indian	1,052	41.16	36.88	15.59	6.37	21.96
	Multiracial	2,573	30.82	33.07	22.11	13.99	36.11
	Pacific Islander	427	25.29	39.58	25.53	9.60	35.13
	White	70,899	27.98	35.26	24.23	12.53	36.76
NRC	New York	66,086	34.75	35.43	19.32	10.50	29.82
	Big 4 Cities	6,843	66.93	22.93	7.64	2.50	10.14
	Urban/Suburban	12,373	52.86	31.73	11.96	3.45	15.41
	Rural	8,182	46.48	34.80	14.07	4.66	18.72
	Average Needs	38,476	31.53	36.55	22.00	9.92	31.92
	Low Needs	19,273	15.44	34.80	31.66	18.10	49.76
	Charter	9,141	36.18	40.24	18.76	4.82	23.59
	Non-Public	10,466	38.14	35.63	18.55	7.68	26.23
SWD	All Codes	28,543	74.27	20.72	4.14	0.87	5.00
SUA	All Codes	21,094	71.51	22.21	5.35	0.92	6.28
ELL	ELL=Y	13,356	78.26	18.73	2.65	0.36	3.01
SWD/SUA	SUA=504 plan codes	12,542	81.49	15.45	2.65	0.41	3.06
ELL/SUA	SUA & ELL codes	4,173	82.99	15.05	1.77	0.19	1.97

ELA Grade 6

Performance level distributions and n-counts of demographic subgroups for ELA Grade 6 are presented in Table 46D. Statewide, a combined 31% of students achieved Level III and Level IV. About 36% of Female students were at Level III or above, as compared to 26% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. About 53% of Asian students and 50% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 11–22% of students in those same performance categories. Only 4% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (31%): Female (36%), Asian (53%), Multiracial (36%), Pacific Islander (36%), and White (39%) students and those enrolled in Average (34%) or Low (50%) Needs districts.

Table 46D. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	171,859	30.18	39.05	16.35	14.41	30.76
Gender	Female	84,061	23.98	39.97	18.55	17.50	36.05
	Male	87,798	36.12	38.17	14.24	11.46	25.71
Ethnicity	Asian	17,104	14.42	32.62	22.26	30.69	52.95
	Black	33,882	41.52	40.33	11.59	6.56	18.15
	Hispanic	45,550	38.73	42.05	12.14	7.08	19.23
	American Indian	1,135	34.80	42.11	12.78	10.31	23.08
	Multiracial	2,384	29.03	35.15	16.11	19.71	35.82
	Pacific Islander	488	23.36	40.98	19.26	16.39	35.66
	White	71,316	23.13	38.15	19.93	18.80	38.72
NRC	New York	64,744	30.56	39.25	15.31	14.87	30.19
	Big 4 Cities	6,684	59.01	30.00	7.21	3.79	11.00
	Urban/Suburban	12,059	47.24	37.19	9.93	5.64	15.57
	Rural	8,124	39.08	39.40	13.15	8.37	21.52
	Average Needs	37,218	26.49	39.69	18.11	15.71	33.82
	Low Needs	18,602	12.36	37.15	24.48	26.00	50.48
	Charter	9,195	30.38	45.56	15.57	8.49	24.07
	Non-Public	15,049	28.15	40.37	17.89	13.58	31.47
SWD	All Codes	27,626	70.14	25.20	3.29	1.37	4.66
SUA	All Codes	19,822	68.06	26.02	4.09	1.84	5.92
ELL	ELL=Y	11,985	76.92	20.78	1.81	0.48	2.29
SWD/SUA	SUA=504 plan codes	12,348	77.02	20.02	2.21	0.75	2.96
ELL/SUA	SUA & ELL codes	3,452	81.46	16.89	1.36	0.29	1.65

ELA Grade 7

Performance level distributions and n-counts of demographic subgroups for ELA Grade 7 are presented in Table 46E. Statewide, a combined 29% of students achieved Level III and Level IV. About 35% of Female students were at Level III or above, as compared to 24% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. About 53% of Asian students and 51% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 9–21% of students in those same performance categories. Only 4% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (29%): Female (35%), Asian (53%), Multiracial (32%), Pacific Islander (36%), and White (38%) students and those enrolled in Average (32%) or Low (51%) Needs districts.

Table 46E. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	164,563	37.62	33.19	23.68	5.52	29.19
Gender	Female	80,270	30.80	34.64	27.55	7.01	34.56
	Male	84,293	44.12	31.80	19.99	4.10	24.08
Ethnicity	Asian	16,353	18.00	29.41	38.73	13.86	52.60
	Black	33,845	51.18	33.31	13.78	1.72	15.51
	Hispanic	43,591	47.25	35.49	15.36	1.91	17.27
	American Indian	1,001	48.15	32.47	15.88	3.50	19.38
	Multiracial	1,861	37.94	30.47	24.61	6.99	31.60
	Pacific Islander	404	28.47	35.64	27.72	8.17	35.89
	White	67,508	29.25	32.62	30.43	7.70	38.13
NRC	New York	65,317	37.55	34.16	22.53	5.76	28.29
	Big 4 Cities	6,448	69.60	21.23	7.97	1.19	9.17
	Urban/Suburban	11,566	57.45	29.14	11.91	1.50	13.41
	Rural	7,993	46.38	32.52	18.15	2.95	21.11
	Average Needs	34,962	34.83	33.54	26.10	5.53	31.63
	Low Needs	19,520	17.29	31.66	39.49	11.56	51.05
	Charter	8,039	36.85	40.76	19.82	2.57	22.39
	Non-Public	10,582	37.41	35.33	23.20	4.05	27.25
SWD	All Codes	26,880	75.94	19.80	3.90	0.37	4.26
SUA	All Codes	18,246	74.77	19.80	5.02	0.41	5.43
ELL	ELL=Y	11,239	84.06	14.43	1.49	0.01	1.50
SWD/SUA	SUA=504 plan codes	11,896	82.24	15.09	2.55	0.13	2.67
ELL/SUA	SUA & ELL codes	2,730	89.63	9.30	1.06	0.00	1.06

ELA Grade 8

Performance level distributions and n-counts of demographic subgroups for ELA Grade 8 are presented in Table 46F. Statewide, a combined 35% of students achieved Level III and Level IV. About 41% of Female students were at Level III or above, as compared to 29% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 57% of Asian students and 58% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–28% of students in those same performance categories. Only 5% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (35%):, Female (41%), Asian (57%), Multiracial (38%), Pacific Islander (46%), and White (44%) and those enrolled in Average Needs (39%), Low Needs (58%), or Non-Public (37%) schools.

Table 46F. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	163,167	28.99	36.00	25.32	9.70	35.02
Gender	Female	79,113	22.27	36.24	28.87	12.61	41.48
	Male	84,054	35.30	35.77	21.98	6.95	28.93
Ethnicity	Asian	16,494	13.86	28.65	35.58	21.92	57.49
	Black	34,363	39.92	38.99	17.57	3.53	21.09
	Hispanic	42,983	35.65	41.39	19.06	3.90	22.96
	American Indian	968	34.71	40.39	19.01	5.89	24.90
	Multiracial	1,684	31.12	31.29	25.12	12.47	37.59
	Pacific Islander	409	20.05	34.23	27.87	17.85	45.72
	White	66,266	22.68	32.84	30.93	13.55	44.48
NRC	New York	65,958	28.40	38.53	23.97	9.09	33.07
	Big 4 Cities	6,412	59.45	27.70	10.29	2.56	12.85
	Urban/Suburban	11,274	46.31	34.90	15.38	3.41	18.79
	Rural	7,991	37.82	34.64	21.85	5.69	27.54
	Average Needs	33,353	27.13	34.36	27.39	11.12	38.51
	Low Needs	18,377	13.23	29.20	37.51	20.05	57.57
	Charter	6,050	25.11	46.56	24.48	3.85	28.33
	Non-Public	13,563	25.22	37.86	28.16	8.77	36.92
SWD	All Codes	25,822	67.98	26.17	5.16	0.69	5.85
SUA	All Codes	17,487	67.26	25.04	6.60	1.10	7.71
ELL	ELL=Y	10,201	77.00	20.28	2.48	0.24	2.72
SWD/SUA	SUA=504 plan codes	11,518	75.12	20.47	3.87	0.54	4.41
ELL/SUA	SUA & ELL codes	2,417	84.94	13.12	1.86	0.08	1.94

Mathematics Test Performance Level Distributions

Table 47 shows the performance level distributions for all examinees from public, charter, and private schools with valid scores, and presents mathematics performance level data for total populations of students in Grades 3–8. Performance level data for selected subgroups of students were also examined. In general, these summaries reflect the same achievement trends as in the scale score summary discussion. Male and Female students performed similarly across grades. More White, Native Hawaiian or Other Pacific Islander, and Asian students were classified in Level III and above, as compared to their peers from other ethnic subgroups. Students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, High Needs Urban/Suburban, and High Needs Rural), Private Schools, and Charter Schools. The subgroups that used the Korean or Chinese translations outperformed other test translation subgroups. The Level III and above rates for SWD and SUA subgroups were low, compared to the total population of examinees. Please note that the case counts for the Haitian-Creole, Korean, and Russian translation subgroups were very low, and the results might have been heavily influenced by very high and/or very low achieving individual students.

Table 47. Mathematics Test Performance Level Distributions

Grade	N-Count	Level I	Level II	Level III	Level IV	Level III & IV
3	176,720	27.92	30.31	23.68	18.09	41.77
4	176,807	26.91	30.42	23.96	18.71	42.67
5	167,821	31.12	26.63	26.15	16.10	42.25
6	166,508	28.29	33.17	19.94	18.60	38.54
7	156,113	33.47	31.99	22.84	11.70	34.54
8	124,506	39.11	38.65	15.53	6.71	22.24

Mathematics Grade 3

Performance level summaries and n-counts of demographic subgroups for Grade 3 are presented in Table 48A. Statewide, a combined 42% of students achieved Level III and Level IV. About 41% of Female students were at Level III or above, as compared to 42% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. About 66% each of Asian students and students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 19–37% of students in those same performance categories. Only 16% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (42%): Asian (66%), Multiracial (45%), Pacific Islander (50%), and White (52%) students and those enrolled in Average (47%) and Low (66%) Needs districts and Charter (52%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 2% (Haitian-Creole) to 67% (Korean).

Table 48A. Mathematics Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	176,720	27.92	30.31	23.68	18.09	41.77
Gender	Female	86,474	27.34	31.58	23.94	17.14	41.08
	Male	90,246	28.48	29.09	23.43	19.00	42.43
Ethnicity	Asian	17,971	11.06	22.95	28.63	37.37	66.00
	Black	33,418	41.19	31.71	17.35	9.75	27.10
	Hispanic	50,556	37.67	33.90	18.92	9.51	28.43
	American Indian	1,170	31.62	31.03	23.42	13.93	37.35
	Multiracial	3,787	24.56	30.24	23.61	21.60	45.21
	Pacific Islander	637	20.25	29.83	28.26	21.66	49.92
	White	69,181	18.96	28.92	28.89	23.22	52.11
NRC	New York	71,492	30.19	31.20	21.81	16.81	38.61
	Big 4 Cities	7,580	55.55	25.73	12.73	5.99	18.72
	Urban/Suburban	13,747	42.76	32.20	17.25	7.79	25.05
	Rural	8,625	30.50	32.71	23.81	12.97	36.79
	Average Needs	38,098	22.00	30.88	27.65	19.47	47.12
	Low Needs	17,530	10.34	24.09	31.60	33.96	65.57
	Charter	9,212	17.82	29.87	25.81	26.49	52.30
	Non-Public	10,264	30.75	31.80	23.06	14.39	37.45
SWD	All Codes	26,467	58.37	25.42	10.92	5.29	16.21
SUA	All Codes	20,970	55.12	27.46	11.86	5.57	17.42
ELL	ELL=Y	18,488	56.36	27.40	10.69	5.55	16.25
SWD/SUA	SUA=504 plan codes	10,740	65.14	22.73	8.49	3.64	12.13
ELL/SUA	SUA & ELL codes	6,289	62.63	25.66	8.11	3.59	11.70
ELL Test Language	English	15,119	55.84	28.06	10.78	5.32	16.10
	Chinese	615	15.77	28.46	28.94	26.83	55.77
	Haitian-Creole	42	76.19	21.43	2.38	0.00	2.38
	Korean	24	4.17	29.17	25.00	41.67	66.67
	Russian	77	35.06	28.57	22.08	14.29	36.36
	Spanish	2,611	69.71	23.32	5.55	1.42	6.97
	All Translations	3,369	58.68	24.40	10.30	6.62	16.92

Mathematics Grade 4

Performance level summaries and n-counts of demographic subgroups for Grade 4 are presented in Table 48B. Statewide, a combined 43% of students achieved Level III and Level IV. About 42% of Female students were at Level III or above, as compared to 43% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 69% of Asian students and 68% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–38% of students in those same performance categories. Only 14% of

the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (43%): Asian (69%), Multiracial (49%), Pacific Islander (48%), and White (54%) students and those enrolled in Average (50%) and Low (68%) Needs districts and Charter (48%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 0% (Haitian-Creole) to 71% (Korean).

Table 48B. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	176,807	26.91	30.42	23.96	18.71	42.67
Gender	Female	86,670	26.30	31.32	24.53	17.85	42.38
	Male	90,137	27.49	29.56	23.41	19.54	42.95
Ethnicity	Asian	17,838	9.74	20.92	28.30	41.04	69.34
	Black	33,918	42.21	32.63	16.85	8.31	25.16
	Hispanic	49,337	36.71	34.79	18.89	9.60	28.49
	American Indian	1,081	30.53	32.19	20.63	16.65	37.28
	Multiracial	3,061	23.42	27.83	25.12	23.62	48.74
	Pacific Islander	524	21.56	30.92	26.53	20.99	47.52
	White	71,048	17.24	28.79	29.77	24.20	53.96
NRC	New York	70,868	29.93	30.83	20.92	18.32	39.24
	Big 4 Cities	7,017	55.42	26.74	12.54	5.30	17.84
	Urban/Suburban	12,716	44.03	32.32	16.31	7.34	23.65
	Rural	7,974	29.52	32.86	24.93	12.69	37.62
	Average Needs	36,960	19.55	30.15	29.03	21.27	50.30
	Low Needs	17,829	8.63	23.24	33.09	35.04	68.13
	Charter	8,008	19.78	31.87	26.41	21.94	48.35
	Non-Public	15,302	27.06	35.56	24.94	12.44	37.38
SWD	All Codes	27,689	59.75	26.43	9.88	3.94	13.82
SUA	All Codes	20,028	56.91	27.56	11.26	4.27	15.54
ELL	ELL=Y	16,791	56.98	28.04	10.37	4.60	14.98
SWD/SUA	SUA=504 plan codes	11,045	66.24	23.23	8.12	2.41	10.53
ELL/SUA	SUA & ELL codes	5,157	66.36	23.54	7.29	2.81	10.10
ELL Test Language	English	13,459	55.84	29.39	10.48	4.29	14.77
	Chinese	523	12.05	25.62	33.08	29.25	62.33
	Haitian-Creole	48	70.83	29.17	0.00	0.00	0.00
	Korean	31	6.45	22.58	32.26	38.71	70.97
	Russian	85	38.82	23.53	27.06	10.59	37.65
	Spanish	2,645	72.59	21.85	4.76	0.79	5.56
	All Translations	3,332	61.58	22.60	9.96	5.85	15.82

Mathematics Grade 5

Performance level summaries and n-counts of demographic subgroups for Grade 5 are presented in Table 48C. Statewide, a combined 39% of students achieved Level III and Level IV. About 43% of Female students were at Level III or above, as compared to 42% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 72% of Asian students and 67% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 16–34% of students in those same performance categories. Only 12% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (42%): Asian (72%), Multiracial (47%), Pacific Islander (51%), and White (53%) students and those from Average (48%) and Low (67%) Needs districts. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 5% (Haitian-Creole) to 76% (Korean).

Table 48C. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	167,821	31.12	26.63	26.15	16.10	42.25
Gender	Female	82,165	29.28	28.13	27.08	15.50	42.58
	Male	85,656	32.89	25.18	25.27	16.67	41.94
Ethnicity	Asian	17,893	11.37	17.09	31.45	40.09	71.54
	Black	32,496	46.80	29.36	17.89	5.96	23.85
	Hispanic	45,632	41.58	29.99	21.15	7.28	28.43
	American Indian	992	33.67	32.56	21.57	12.20	33.77
	Multiracial	2,416	26.74	25.95	26.95	20.36	47.31
	Pacific Islander	432	23.61	25.46	29.86	21.06	50.93
	White	67,960	21.98	25.52	32.08	20.43	52.51
NRC	New York	67,140	31.89	27.14	24.29	16.68	40.97
	Big 4 Cities	6,732	64.17	20.28	11.53	4.03	15.55
	Urban/Suburban	11,835	49.78	27.17	17.57	5.48	23.05
	Rural	7,580	36.21	29.35	24.39	10.04	34.43
	Average Needs	36,298	24.83	27.44	30.63	17.10	47.73
	Low Needs	18,538	11.13	21.51	37.04	30.32	67.36
	Charter	9,152	29.87	30.86	26.46	12.81	39.27
	Non-Public	10,409	38.63	27.26	23.38	10.73	34.11
SWD	All Codes	27,493	66.68	20.96	9.66	2.70	12.36
SUA	All Codes	19,890	64.13	21.36	11.14	3.38	14.51
ELL	ELL=Y	15,219	66.03	20.76	9.49	3.72	13.21
SWD/SUA	SUA=504 plan codes	11,481	73.54	17.38	7.46	1.63	9.08
ELL/SUA	SUA & ELL codes	4,565	74.37	16.80	6.13	2.69	8.83

Table 48C. Mathematics Grade 5 Performance Level Distribution by Subgroup (cont.)

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
ELL Test Language	English	12,051	65.95	21.28	9.45	3.31	12.76
	Chinese	508	19.09	24.21	31.69	25.00	56.69
	Haitian-Creole	44	81.82	13.64	4.55	0.00	4.55
	Korean	33	9.09	15.15	33.33	42.42	75.76
	Russian	91	41.76	32.97	17.58	7.69	25.27
	Spanish	2,492	77.33	17.30	4.61	0.76	5.38
	All Translations	3,168	66.32	18.78	9.63	5.27	14.90

Mathematics Grade 6

Performance level summaries and n-counts of demographic subgroups for Grade 6 are presented in Table 48D. Statewide, a combined 39% of students achieved Level III and Level IV. About 38% of Female students were at Level III or above, as compared to 40% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 68% of Asian students and 66% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 15–30% of students in those same performance categories. Only 10% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (39%): Female (40%), Asian (68%), Multiracial (44%), Pacific Islander (42%), and White (49%) students and those enrolled in Average (45%) and Low (66%) Needs districts. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 2% (Haitian-Creole) to 81% (Korean).

Table 48D. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	166,508	28.29	33.17	19.94	18.60	38.54
Gender	Female	81,249	25.94	33.90	20.90	19.26	40.17
	Male	85,259	30.53	32.49	19.02	17.96	36.99
Ethnicity	Asian	17,384	9.84	22.20	23.67	44.29	67.96
	Black	32,928	44.45	34.28	13.79	7.49	21.28
	Hispanic	45,293	38.88	36.59	15.66	8.88	24.53
	American Indian	1,081	32.28	38.67	16.74	12.30	29.05
	Multiracial	2,197	24.49	31.41	20.85	23.26	44.11
	Pacific Islander	491	22.61	35.23	20.37	21.79	42.16
	White	67,134	18.10	33.13	24.91	23.87	48.78

Table 48D. Mathematics Grade 6 Performance Level Distribution by Subgroup (cont.)

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	65,739	32.15	32.29	17.27	18.30	35.57
	Big 4 Cities	6,414	55.77	29.11	9.82	5.30	15.12
	Urban/Suburban	11,185	47.22	33.41	13.05	6.31	19.37
	Rural	7,445	29.91	40.04	19.53	10.52	30.05
	Average Needs	34,054	19.99	35.27	24.23	20.51	44.74
	Low Needs	17,505	8.37	25.57	28.19	37.86	66.06
	Charter	9,136	24.98	36.89	22.22	15.92	38.13
	Non-Public	14,865	28.71	36.99	20.65	13.66	34.31
SWD	All Codes	26,115	65.74	25.51	6.08	2.67	8.75
SUA	All Codes	19,215	59.06	28.25	8.46	4.22	12.68
ELL	ELL=Y	13,788	64.53	25.27	6.49	3.71	10.20
SWD/SUA	SUA=504 plan codes	11,496	67.85	24.48	5.56	2.11	7.67
ELL/SUA	SUA & ELL codes	4,122	68.46	22.44	5.80	3.30	9.10
ELL Test Language	English	9,975	64.56	26.05	6.39	3.01	9.39
	Chinese	593	12.98	32.72	26.31	27.99	54.30
	Haitian-Creole	84	83.33	14.29	2.38	0.00	2.38
	Korean	36	8.33	11.11	30.56	50.00	80.56
	Russian	96	32.29	34.38	18.75	14.58	33.33
	Spanish	3,004	75.80	21.40	2.36	0.43	2.80
	All Translations	3,813	64.46	23.24	6.77	5.53	12.30

Mathematics Grade 7

Performance level summaries and n-counts of demographic subgroups for Grade 7 are presented in Table 48E. Statewide, a combined 35% of students achieved Level III and Level IV. About 36% of Female students were at Level III or above, as compared to 33% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 67% of Asian students and 62% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 10–25% of students in those same performance categories. Only 8% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (35%): Female (36%), Asian (67%), Multiracial (37%), Pacific Islander (42%), and White (45%) students and those enrolled in Average (39%) and Low (62%) Needs districts. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 0% (Haitian-Creole) to 72% (Korean).

Table 48E. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	156,113	33.47	31.99	22.84	11.70	34.54
Gender	Female	75,976	31.33	32.89	23.92	11.86	35.78
	Male	80,137	35.50	31.13	21.82	11.55	33.37
Ethnicity	Asian	16,383	12.02	21.44	32.19	34.35	66.54
	Black	32,201	49.90	32.49	13.63	3.98	17.61
	Hispanic	42,925	45.10	34.49	15.95	4.47	20.41
	American Indian	902	45.12	31.49	16.85	6.54	23.39
	Multiracial	1,634	31.58	31.40	23.13	13.89	37.03
	Pacific Islander	396	24.75	33.08	25.51	16.67	42.17
	White	61,672	22.43	32.81	30.03	14.74	44.77
NRC	New York	66,103	36.45	30.96	19.70	12.89	32.59
	Big 4 Cities	5,952	65.56	24.50	8.08	1.86	9.95
	Urban/Suburban	10,332	54.37	31.51	11.56	2.56	14.11
	Rural	6,922	38.25	37.17	19.75	4.83	24.57
	Average Needs	30,548	26.38	34.48	28.31	10.83	39.14
	Low Needs	17,468	11.45	26.91	37.47	24.16	61.64
	Charter	7,992	27.40	37.84	25.35	9.41	34.76
	Non-Public	10,682	34.42	36.43	22.08	7.07	29.15
SWD	All Codes	24,783	71.53	21.83	5.40	1.24	6.64
SUA	All Codes	17,150	65.68	24.35	7.85	2.12	9.97
ELL	ELL=Y	13,228	71.18	20.41	6.20	2.21	8.41
SWD/SUA	SUA=504 plan codes	10,421	73.63	20.89	4.67	0.81	5.48
ELL/SUA	SUA & ELL codes	3,561	76.02	16.40	5.14	2.44	7.58
ELL Test Language	English	9,574	72.72	20.96	4.96	1.36	6.32
	Chinese	627	11.80	29.67	36.20	22.33	58.53
	Haitian-Creole	72	83.33	16.67	0.00	0.00	0.00
	Korean	32	12.50	15.63	34.38	37.50	71.88
	Russian	130	41.54	33.85	20.77	3.85	24.62
	Spanish	2,793	80.99	15.97	2.86	0.18	3.04
	All Translations	3,654	67.16	18.97	9.44	4.43	13.88

Mathematics Grade 8

Performance level summaries and n-counts of demographic subgroups for Grade 8 are presented in Table 48F. Statewide, a combined 22% of students achieved Level III and Level IV. About 23% of Female students were at Level III or above, as compared to 21% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 50% of Asian students and 41% of students from Low Needs districts were classified in Levels III and IV, whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 7–15% of students in those same performance categories. Only 6% of the

SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (22%): Female (23%), Asian (50%), Multiracial (25%), Pacific Islander (35%), and White (28%) students and those enrolled in Low Needs (41%) districts or Charter (32%) or Non-Public (24%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 0% (Haitian-Creole) to 55% (Chinese).

Table 48F. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	124,506	39.11	38.65	15.53	6.71	22.24
Gender	Female	59,426	36.23	40.45	16.32	7.00	23.32
	Male	65,080	41.74	37.01	14.80	6.45	21.25
Ethnicity	Asian	11,113	16.92	33.04	25.48	24.56	50.04
	Black	29,475	53.59	34.16	9.46	2.79	12.25
	Hispanic	37,581	46.99	38.04	11.39	3.59	14.97
	American Indian	755	45.17	36.42	15.10	3.31	18.41
	Multiracial	1,183	36.86	37.70	17.58	7.86	25.44
	Pacific Islander	318	27.04	38.05	20.44	14.47	34.91
	White	44,081	28.35	43.66	20.52	7.47	27.99
NRC	New York	56,638	40.66	36.75	14.53	8.06	22.59
	Big 4 Cities	5,545	71.43	20.65	5.97	1.95	7.92
	Urban/Suburban	8,429	59.44	33.29	6.15	1.13	7.27
	Rural	5,809	40.92	44.29	12.74	2.05	14.79
	Average Needs	20,191	31.74	47.17	17.47	3.62	21.09
	Low Needs	9,573	17.33	41.95	27.67	13.05	40.72
	Charter	5,228	28.00	40.40	21.29	10.31	31.60
	Non-Public	12,967	36.21	39.32	17.14	7.33	24.47
SWD	All Codes	22,664	72.76	22.37	4.08	0.79	4.88
SUA	All Codes	14,101	69.32	24.41	5.08	1.18	6.27
ELL	ELL=Y	11,723	68.20	23.43	5.73	2.64	8.37
SWD/SUA	SUA=504 plan codes	9,025	74.88	20.83	3.73	0.55	4.29
ELL/SUA	SUA & ELL codes	2,560	81.72	14.73	2.46	1.09	3.55
ELL Test Language	English	8,216	69.28	24.21	4.76	1.75	6.51
	Chinese	662	14.95	30.51	31.27	23.26	54.53
	Haitian-Creole	72	80.56	19.44	0.00	0.00	0.00
	Korean	25	16.00	44.00	32.00	8.00	40.00
	Russian	122	45.08	41.80	9.02	4.10	13.11
	Spanish	2,626	79.47	18.28	2.09	0.15	2.25
	All Translations	3,507	65.67	21.61	8.01	4.70	12.72

References

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

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

Bock, R.D. & M. Aitkin (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika* 46:443–459.

Cattell, R.B. (1966). The Screen Test for the Number of Factors. *Multivariate Behavioral Research* 1:245–276.

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

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

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

Fleiss J.L. & J. Cohen (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33: 613–619.

Green, D.R., W.M. Yen & G.R. Burket (1989). Experiences in the application of item response theory in test construction. *Applied Measurement in Education* 2:297–312.

Huynh, H. & C. Schneider (2004). *Vertically moderated standards as an alternative to vertical scaling: assumptions, practices, and an odyssey through NAEP.* Paper presented at the National Conference on Large-Scale Assessment. Boston, MA, June 21.

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

Johnson, N.L. & S. Kotz (1970). *Distributions in Statistics: Continuous Univariate Distributions,* Vol. 2. New York: John Wiley.

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

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

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

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

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

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

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

Linn, R.L. (1991). Linking results of distinct assessments. *Applied Measurement in Education* 6(1): 83–102.

Linn, R.L. & D. Harnisch (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measuremen*t 18: 109–118.

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

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

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

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

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

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

Novick, M.R. & P.H. Jackson (1974). *Statistical Methods for Educational and Psychological Research.* New York: McGraw-Hill.

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

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

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

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

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

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

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

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

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

Yen, W.M. (1997). The technical quality of performance assessments: Standard errors of percents of students reaching standards. *Educational Measurement: Issues and Practice*: 5–15.

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

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

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

Yen, W.M., R.C. Sykes, K. Ito & M. Julian (1997). *A Bayesian/IRT index of objective performance for tests with mixed-item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago: March.

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

Appendix A: ELA and Mathematics Test Configuration and Testing Times

Table A1. ELA Test Configuration

Grade	Day	Book	Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	24	6	0	0	30
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		**31**	**6**	**10**	**0**	**47**
4	1	1	24	6	0	0	30
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		**31**	**6**	**10**	**0**	**47**
5	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		**42**	**7**	**10**	**0**	**59**
6	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		**42**	**7**	**10**	**0**	**59**
7	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		**42**	**7**	**10**	**0**	**59**
8	1	1	35	7	0	0	42
	2	2	7	0	4	0	11
	3	3	0	0	6	0	6
	Total		**42**	**7**	**10**	**0**	**59**

Table A2. Mathematics Test Configuration

Grade	Day	Book	Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	20	4	0	0	24
	2	2	21	3	0	0	24
	3	3	0	0	8	0	8
	Total		**41**	**7**	**8**	**0**	**56**
4	1	1	20	4	0	0	24
	2	2	22	3	0	0	25
	3	3	0	0	8	0	8
	Total		**42**	**7**	**8**	**0**	**57**
5	1	1	20	4	0	0	24
	2	2	22	3	0	0	25
	3	3	0	0	8	0	8
	Total		**42**	**7**	**8**	**0**	**57**
6	1	1	24	4	0	0	28
	2	2	24	3	0	0	27
	3	3	0	0	8	0	8
	Total		**48**	**7**	**8**	**0**	**63**
7	1	1	24	4	0	0	28
	2	2	24	3	0	0	27
	3	3	0	0	8	0	8
	Total		**48**	**7**	**8**	**0**	**63**
8	1	1	24	4	0	0	28
	2	2	24	3	0	0	27
	3	3	0	0	8	0	8
	Total		**48**	**7**	**8**	**0**	**63**

Table A3. ELA Testing Times

Grade	Day	Book	Estimated Time on Task (min)	Session Time (min)
3	1	1	50	70
	2	2	50	70
	3	3	50	70
	Total		**150**	**210**
4	1	1	50	70
	2	2	50	70
	3	3	50	70
	Total		**150**	**210**
5	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		**180**	**270**
6	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		**180**	**270**
7	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		**180**	**270**
8	1	1	70	90
	2	2	60	90
	3	3	50	90
	Total		**180**	**270**

Source: *2015 Common Core ELA and Mathematics Test Guides.*

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

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

Table A4. Math Testing Times

Grade	Day	Book	Estimated Time on Task (min)	Session Time (min)
3	1	1	40	60
	2	2	40	60
	3	3	50	70
	Total		**130**	**190**
4	1	1	40	60
	2	2	40	60
	3	3	70	90
	Total		**150**	**210**
5	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		**150**	**250**
6	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		**150**	**250**
7	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		**150**	**250**
8	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		**150**	**250**

Source: *2015 Common Core ELA and Mathematics Test Guides.*

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

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

The testing times listed above do not include approximately 10 minutes reserved for preparation at the beginning of each session for handing out materials and reading directions. Additional details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the 2015 *Teacher's Directions* and the *School Administrator's Manual*, which are accessible online:

- *2015 Common Core ELA Teacher's Directions*
  - Grades 3–5: http://www.p12.nysed.gov/assessment/ei/2015/tdela35-15.pdf
  - Grades 6–8: http://www.p12.nysed.gov/assessment/ei/2015/tdela68-15cr.pdf

- *2015 Common Core Mathematics Teacher's Directions*
  - Grades 3–5: http://www.p12.nysed.gov/assessment/ei/2015/tdmath35-15.pdf
  - Grades 6–8: http://www.p12.nysed.gov/assessment/ei/2015/tdmath68-15.pdf
- *2015 Common Core ELA and Mathematics Tests School Administrator's Manual*
  - http://www.p12.nysed.gov/assessment/sam/ei/eisam15rev.pdf
- *2015 Common Core ELA and Mathematics Test Guides*
  - https://www.engageny.org/resource/test-guides-for-english-language-arts-and-mathematics

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Standard	Point Range		% of Test	
			Target	Actual	Target	Actual
3	55	Literature	18–44	31	33%–80%	56%
		Information	18–44	22	33%–80%	40%
		Language	1–4	2	2%–7%	4%
4	55	Literature	18–44	24	33%–80%	44%
		Information	18–44	27	33%–80%	49%
		Language	1–4	4	2%–7%	7%
5	66	Literature	18–51	28	27%–77%	42%
		Information	18–51	35	27%–77%	53%
		Language	1–4	3	2%–6%	5%
6	65	Literature	11–44	22	17%–67%	33%
		Information	25–58	42	38%–88%	64%
		Language	1–4	2	2%–6%	3%
7	66	Literature	11–44	23	17%–67%	35%
		Information	25–58	42	38%–88%	64%
		Language	1–4	1	2%–6%	2%
8	66	Literature	11–44	24	17%–67%	36%
		Information	25–58	41	38%–88%	62%
		Language	1–4	1	2%–6%	2%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Standard	Point Range		% of Test	
			Target	Actual	Target	Actual
3	60	Operations and Algebraic Thinking	23–31	27	38%–52%	45%
		Number and Operations in Base Ten	3–5	4	5%–8%	7%
		Number and Operations – Fractions	10–14	12	17%–23%	20%
		Measurement and Data	12–18	15	20%–30%	25%
		Geometry*	1–3	2	2%–5%	3%
4	66	Operations and Algebraic Thinking	11–15	13	17%–23%	20%
		Number and Operations in Base Ten	14–20	17	21%–30%	26%
		Number and Operations – Fractions	15–21	18	23%–32%	27%
		Measurement and Data	9–15	12	14%–23%	18%
		Geometry	5–7	6	8%–11%	9%
5	66	Operations and Algebraic Thinking	3–5	4	5%–8%	6%
		Number and Operations in Base Ten	15–21	18	23%–32%	27%
		Number and Operations – Fractions	22–28	25	33%–42%	38%
		Measurement and Data	13–21	17	20%–32%	26%
		Geometry*	1–3	2	2%–5%	3%
6	72	Ratios and Proportional Relationships	16–20	18	22%–28%	25%
		The Number System	13–19	16	18%–26%	22%
		Expressions and Equations	23–33	28	32%–46%	39%
		Geometry	8–12	10	11%–17%	14%
7	72	Ratios and Proportional Relationships	18–22	20	25%–31%	28%
		The Number System	12–16	14	17%–22%	19%
		Expressions and Equations	19–25	22	26%–35%	31%
		Geometry	3–7	5	4%–10%	7%
		Statistics and Probability	8–14	11	11%–19%	15%
8	72	Expressions and Equations	26–34	30	36%–47%	42%
		Functions	14–20	19	19%–28%	26%
		Geometry	16–22	17	22%–31%	24%
		Statistics and Probability	5–7	6	7%–10%	8%

* There is a slight difference between the "Target% of Test" shown in these tables and the tables presented in the Guides to the 2015 Common Core 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 Common Core mathematics standards.

Appendix C: Passage Selection Guidelines for Assessing ELA

General Guidelines

Along with instructional materials and teacher training, assessment development is essential to the successful implementation of the CCSS. While many of the expectations outlined in the CCSS align with previous versions of the NYS Learning Standards for ELA, the CCSS do represent some shifts in emphasis with direct implications for assessment development. In particular, the CCSS devote considerable attention to the types and nature of texts used in instruction and assessment. The foundation for preparing students for the linguistic rigors of college and of the workplace lies in the texts with which they interact. By the time that they graduate, students should be prepared to successfully read and analyze the types of complex texts that they will encounter after high school. Selecting passages of appropriate type and complexity for use in assessment is integral to this preparation.

One of the major shifts of the CCSS is an emphasis on developing skills for comprehending and analyzing informational texts. Increased exposure to informational texts better prepares students for the various types of texts that they will encounter in college and in the workplace. The array of passages selected for assessment from K–12 should support the development of the necessary skills to handle this range of informational texts.

Another shift is an increased emphasis on the analysis across multiple texts, often of varied genres and media. Several standards, especially for reading literature, require intertextual and multi-media analysis. These expectations require special attention to the selection of related passages, chosen specifically to support the assessment of the full range of expectations. It will also require careful consideration of which standards are appropriate for large-scale assessment formats, and how these assessments might be modified to include passages of a variety of media.

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

Text Complexity
Text Types
Text Suitability for Specific Standards

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

Appendix D: Universal Design Item Checklist

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

Universal Design Item Checklist	
√	The item is free from unnecessary cultural references.
√	The item is free from religious references.
F.	**Geographic Bias**
Definition	All groups of society should be portrayed accurately and fairly regarding geographic setting. A particular geographic setting shouldn't be used repeatedly, and urban, suburban, and rural settings should be represented across items.
√	The item is free of content that might offend a geographic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a geographic subgroup.
G.	**Disability Bias**
Definition	All groups of society should be portrayed accurately and fairly regarding disability. Stereotypes related to any particular disability should be avoided. No undue restrictions should exist in the item that would interfere with the ability of a student to comprehend or respond to the item.
√	The item is free of content that might offend a disability subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a disability subgroup.
√	A graphic representation is used in the items, as appropriate. The complexity of the graphic is appropriate to the purpose—simpler is better.
√	The item avoids content that depends on sensory knowledge (such as references to movement, sound, smell, etc.) unless this is crucial to the overall item.
√	The item could be put into Braille.
√	The item avoids using both O and Q.
√	Letter pairs can be easily distinguished when read. (S and T are okay; S and X are not).
H.	**Art Supports Text**
Definition	The art is related to the item and supports the reader when possible. The item text and art are legible and accessible, and the art is appropriately placed in the item to support the reader. The art does not distract the test taker, but instead provides a scaffold to overall comprehension.
√	All pictures relate to items.
√	The item is free from pictorial clutter: All pictures are needed to answer the item.
√	Graphics are clear and non-fuzzy.
√	Any symbols used are highly distinguishable.
√	Visual load requirements are reasonable for the grade level.
√	Multi-dimensional graphics and complex shading are avoided.
√	Tables have replaced any cluttered graphs.
√	Labels read clockwise (as is easier for Braille readers).
I.	**Special Populations Considerations**
Definition	Consideration must be given for maximum accessibility to all students including, but not limited to, English language learners, 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 NYS 3–8 CCSS test questions; however, these criteria are not a substitute for the full, detailed criteria documents, which are available online at the following links:

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

For Multiple-Choice Items:

Check that the content of each item

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

Check that the format of each item

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

Also check that

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

For Constructed-Response Items:

Check that the content of each item is

- designed to assess the targeted performance indicator
- appropriate for the grade level being tested
- presented at a reading level suitable for the grade level being tested
- appropriate in context
- written so that a student possessing 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 2015 Common Core Operational Test. The psychometrics department will provide support, as necessary, and will review the final item selection. The psychometrics department will provide data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2015 item selection included 2013 and 2014 embedded and stand-alone field test items and items field-tested in New York State in 2013 and 2014.

Here are the general guidelines for item selection:

- Satisfy the content specifications in terms of objective coverage and the number and percentage of MC and CR items on the test. An often-used criterion for objective coverage is within 5% of the percentages of score points and items per objective.
- To the extent possible, select both easy and difficult items to provide good measurement information at both ends of the performance scale.
- Avoid selecting items with too high/low p-values, items with flagged point biserials, and poorly fitting items.
- Minimize the number of items flagged for DIF (gender, ethnic, and High/Low Needs schools). Flagged items should be reviewed for content again. It needs to be remembered that some items may be flagged for DIF by chance only, and their content may not necessarily be biased against any of the analyzed subgroups. The psychometrics department will provide DIF information for each item. It is also possible to get "significant" DIF, but not bias, if the content is a necessary part of the construct that is measured. That is, some items may be flagged for DIF not out of chance and still not represent bias.
- Provide the NYSED with the following summary information:
    - Overview of the statistical properties of the tests
    - Blueprint comparison between the test build and the target. The focus is on the total number of points on the test
    - Raw score proportion correct comparison between the test build and the reference (i.e., Spring 2014 test)
    - Vertical linked average difficulty parameter (MC items only) across all grades
    - Vertically linked TCC based on the constructed test
    - TCC, Test Information Curves and Conditional SEM Curves for each subject and grade, again using the Spring 2014 operational test as a reference.

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2015 NYSTP Grades 3–8 Common Core ELA and Mathematics Tests. External linking and field-test items (i.e., those not contributing to students' scores) have been omitted. Additional detail on the standards to which these items align may be found at: http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards.

Table G1. English Language Arts Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.3.4
2	MC	1	CCSS.ELA-Literacy.RL.3.1
3	MC	1	CCSS.ELA-Literacy.RL.3.3
4	MC	1	CCSS.ELA-Literacy.RL.3.1
5	MC	1	CCSS.ELA-Literacy.RL.3.1
6	MC	1	CCSS.ELA-Literacy.RL.3.2
7	MC	1	CCSS.ELA-Literacy.RI.3.3
8	MC	1	CCSS.ELA-Literacy.RI.3.1
9	MC	1	CCSS.ELA-Literacy.RI.3.4
10	MC	1	CCSS.ELA-Literacy.RI.3.8
11	MC	1	CCSS.ELA-Literacy.RI.3.1
12	MC	1	CCSS.ELA-Literacy.RI.3.8
13	MC	1	CCSS.ELA-Literacy.L.3.4a
14	MC	1	CCSS.ELA-Literacy.RL.3.3
15	MC	1	CCSS.ELA-Literacy.RL.3.3
16	MC	1	CCSS.ELA-Literacy.RL.3.3
17	MC	1	CCSS.ELA-Literacy.RL.3.1
18	MC	1	CCSS.ELA-Literacy.RL.3.1
19	MC	1	CCSS.ELA-Literacy.RL.3.2
20	MC	1	CCSS.ELA-Literacy.RL.3.5
21	MC	1	CCSS.ELA-Literacy.RL.3.4
22	MC	1	CCSS.ELA-Literacy.RL.3.1
23	MC	1	CCSS.ELA-Literacy.RL.3.3
24	MC	1	CCSS.ELA-Literacy.RL.3.2
25	MC	1	CCSS.ELA-Literacy.RI.3.1
26	MC	1	CCSS.ELA-Literacy.L.3.4a
27	MC	1	CCSS.ELA-Literacy.RI.3.4
28	MC	1	CCSS.ELA-Literacy.RI.3.2
29	MC	1	CCSS.ELA-Literacy.RI.3.3
30	MC	1	CCSS.ELA-Literacy.RI.3.1
31	MC	1	CCSS.ELA-Literacy.RI.3.2

Table G1. English Language Arts Grade 3 Operational Item Map (cont.)

Item	Type	Points	Standard
32	CR	2	CCSS.ELA-Literacy.RL.3.3
33	CR	2	CCSS.ELA-Literacy.RL.3.5
34	CR	2	CCSS.ELA-Literacy.RI.3.5
35	CR	4	CCSS.ELA-Literacy.RI.3.3
36	CR	2	CCSS.ELA-Literacy.RL.3.7
37	CR	2	CCSS.ELA-Literacy.RL.3.2
38	CR	2	CCSS.ELA-Literacy.RI.3.8
39	CR	2	CCSS.ELA-Literacy.RI.3.2
40	CR	2	CCSS.ELA-Literacy.RL.3.4
41	CR	4	CCSS.ELA-Literacy.RL.3.3

Table G2. English Language Arts Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.4.3
2	MC	1	CCSS.ELA-Literacy.L.4.5b
3	MC	1	CCSS.ELA-Literacy.RL.4.3
4	MC	1	CCSS.ELA-Literacy.RL.4.1
5	MC	1	CCSS.ELA-Literacy.L.4.4a
6	MC	1	CCSS.ELA-Literacy.RL.4.2
7	MC	1	CCSS.ELA-Literacy.RI.4.5
8	MC	1	CCSS.ELA-Literacy.RI.4.3
9	MC	1	CCSS.ELA-Literacy.RI.4.1
10	MC	1	CCSS.ELA-Literacy.RI.4.1
11	MC	1	CCSS.ELA-Literacy.RI.4.4
12	MC	1	CCSS.ELA-Literacy.RI.4.3
13	MC	1	CCSS.ELA-Literacy.RL.4.3
14	MC	1	CCSS.ELA-Literacy.RL.4.4
15	MC	1	CCSS.ELA-Literacy.RL.4.1
16	MC	1	CCSS.ELA-Literacy.RL.4.2
17	MC	1	CCSS.ELA-Literacy.RL.4.3
18	MC	1	CCSS.ELA-Literacy.RL.4.2
19	MC	1	CCSS.ELA-Literacy.RI.4.1
20	MC	1	CCSS.ELA-Literacy.RI.4.2
21	MC	1	CCSS.ELA-Literacy.L.4.4a
22	MC	1	CCSS.ELA-Literacy.RI.4.3
23	MC	1	CCSS.ELA-Literacy.RI.4.5
24	MC	1	CCSS.ELA-Literacy.RI.4.2
25	MC	1	CCSS.ELA-Literacy.RI.4.5
26	MC	1	CCSS.ELA-Literacy.RI.4.8

Table G2. English Language Arts Grade 4 Operational Item Map (cont.)

Item	Type	Points	Standard
27	MC	1	CCSS.ELA-Literacy.RI.4.2
28	MC	1	CCSS.ELA-Literacy.RI.4.3
29	MC	1	CCSS.ELA-Literacy.RI.4.3
30	MC	1	CCSS.ELA-Literacy.L.4.4a
31	MC	1	CCSS.ELA-Literacy.RI.4.1
32	CR	2	CCSS.ELA-Literacy.RL.4.3
33	CR	2	CCSS.ELA-Literacy.RL.4.7
34	CR	2	CCSS.ELA-Literacy.RI.4.1
35	CR	4	CCSS.ELA-Literacy.RI.4.2
36	CR	2	CCSS.ELA-Literacy.RI.4.3
37	CR	2	CCSS.ELA-Literacy.RI.4.1
38	CR	2	CCSS.ELA-Literacy.RL.4.3
39	CR	2	CCSS.ELA-Literacy.RL.4.3
40	CR	2	CCSS.ELA-Literacy.RL.4.3
41	CR	4	CCSS.ELA-Literacy.RL.4.9

Table G3. English Language Arts Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.5.4
2	MC	1	CCSS.ELA-Literacy.RL.5.3
3	MC	1	CCSS.ELA-Literacy.RL.5.1
4	MC	1	CCSS.ELA-Literacy.RL.5.5
5	MC	1	CCSS.ELA-Literacy.RL.5.1
6	MC	1	CCSS.ELA-Literacy.RL.5.7
7	MC	1	CCSS.ELA-Literacy.RL.5.6
8	MC	1	CCSS.ELA-Literacy.RL.5.2
9	MC	1	CCSS.ELA-Literacy.RI.5.5
10	MC	1	CCSS.ELA-Literacy.RI.5.2
11	MC	1	CCSS.ELA-Literacy.RI.5.7
12	MC	1	CCSS.ELA-Literacy.RI.5.1
13	MC	1	CCSS.ELA-Literacy.RI.5.1
14	MC	1	CCSS.ELA-Literacy.L.5.4a
15	MC	1	CCSS.ELA-Literacy.RI.5.2
16	MC	1	CCSS.ELA-Literacy.RI.5.4
17	MC	1	CCSS.ELA-Literacy.RI.5.2
18	MC	1	CCSS.ELA-Literacy.RI.5.2
19	MC	1	CCSS.ELA-Literacy.RI.5.8
20	MC	1	CCSS.ELA-Literacy.RI.5.1
21	MC	1	CCSS.ELA-Literacy.RI.5.3

Table G3. English Language Arts Grade 5 Operational Item Map (cont.)

Item	Type	Points	Standard
22	MC	1	CCSS.ELA-Literacy.RI.5.3
23	MC	1	CCSS.ELA-Literacy.RI.5.1
24	MC	1	CCSS.ELA-Literacy.L.5.4a
25	MC	1	CCSS.ELA-Literacy.RI.5.5
26	MC	1	CCSS.ELA-Literacy.RI.5.8
27	MC	1	CCSS.ELA-Literacy.RI.5.3
28	MC	1	CCSS.ELA-Literacy.RI.5.2
29	MC	1	CCSS.ELA-Literacy.RI.5.2
30	MC	1	CCSS.ELA-Literacy.RL.5.5
31	MC	1	CCSS.ELA-Literacy.RL.5.1
32	MC	1	CCSS.ELA-Literacy.RL.5.3
33	MC	1	CCSS.ELA-Literacy.RL.5.2
34	MC	1	CCSS.ELA-Literacy.RL.5.1
35	MC	1	CCSS.ELA-Literacy.RL.5.2
36	MC	1	CCSS.ELA-Literacy.RI.5.1
37	MC	1	CCSS.ELA-Literacy.RI.5.8
38	MC	1	CCSS.ELA-Literacy.RI.5.8
39	MC	1	CCSS.ELA-Literacy.RI.5.3
40	MC	1	CCSS.ELA-Literacy.L.5.4a
41	MC	1	CCSS.ELA-Literacy.RI.5.1
42	MC	1	CCSS.ELA-Literacy.RI.5.2
43	CR	2	CCSS.ELA-Literacy.RL.5.3
44	CR	2	CCSS.ELA-Literacy.RL.5.4
45	CR	2	CCSS.ELA-Literacy.RI.5.7
46	CR	4	CCSS.ELA-Literacy.RI.5.2
47	CR	2	CCSS.ELA-Literacy.RI.5.1
48	CR	2	CCSS.ELA-Literacy.RI.5.3
49	CR	2	CCSS.ELA-Literacy.RL.5.3
50	CR	2	CCSS.ELA-Literacy.RL.5.3
51	CR	2	CCSS.ELA-Literacy.RL.5.1
52	CR	4	CCSS.ELA-Literacy.RL.5.9

Table G4. English Language Arts Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.6.5
2	MC	1	CCSS.ELA-Literacy.RI.6.4
3	MC	1	CCSS.ELA-Literacy.RI.6.1
4	MC	1	CCSS.ELA-Literacy.RI.6.5
5	MC	1	CCSS.ELA-Literacy.RI.6.1
6	MC	1	CCSS.ELA-Literacy.RI.6.2
7	MC	1	CCSS.ELA-Literacy.RI.6.1
8	MC	1	CCSS.ELA-Literacy.RI.6.5
9	MC	1	CCSS.ELA-Literacy.RI.6.4
10	MC	1	CCSS.ELA-Literacy.RI.6.8
11	MC	1	CCSS.ELA-Literacy.RI.6.8
12	MC	1	CCSS.ELA-Literacy.RI.6.3
13	MC	1	CCSS.ELA-Literacy.RI.6.4
14	MC	1	CCSS.ELA-Literacy.RI.6.6
15	MC	1	CCSS.ELA-Literacy.RL.6.2
16	MC	1	CCSS.ELA-Literacy.RL.6.4
17	MC	1	CCSS.ELA-Literacy.L.6.4c
18	MC	1	CCSS.ELA-Literacy.RL.6.1
19	MC	1	CCSS.ELA-Literacy.RL.6.3
20	MC	1	CCSS.ELA-Literacy.RL.6.2
21	MC	1	CCSS.ELA-Literacy.RL.6.1
22	MC	1	CCSS.ELA-Literacy.RI.6.3
23	MC	1	CCSS.ELA-Literacy.RI.6.4
24	MC	1	CCSS.ELA-Literacy.RI.6.1
25	MC	1	CCSS.ELA-Literacy.RI.6.3
26	MC	1	CCSS.ELA-Literacy.RI.6.1
27	MC	1	CCSS.ELA-Literacy.RI.6.2
28	MC	1	CCSS.ELA-Literacy.RI.6.6
29	MC	1	CCSS.ELA-Literacy.RI.6.8
30	MC	1	CCSS.ELA-Literacy.RI.6.1
31	MC	1	CCSS.ELA-Literacy.RI.6.3
32	MC	1	CCSS.ELA-Literacy.RI.6.5
33	MC	1	CCSS.ELA-Literacy.RI.6.3
34	MC	1	CCSS.ELA-Literacy.RI.6.6
35	MC	1	CCSS.ELA-Literacy.RI.6.2
36	MC	1	CCSS.ELA-Literacy.RL.6.3
37	MC	1	CCSS.ELA-Literacy.RL.6.5
38	MC	1	CCSS.ELA-Literacy.RL.6.3
39	MC	1	CCSS.ELA-Literacy.RL.6.4
40	MC	1	CCSS.ELA-Literacy.RL.6.6

Table G4. English Language Arts Grade 6 Operational Item Map (cont.)

Item	Type	Points	Standard
41	MC	1	CCSS.ELA-Literacy.RL.6.1
42	MC	1	CCSS.ELA-Literacy.L.6.4a
43	CR	2	CCSS.ELA-Literacy.RI.6.5
44	CR	2	CCSS.ELA-Literacy.RI.6.7
45	CR	2	CCSS.ELA-Literacy.RI.6.5
46	CR	4	CCSS.ELA-Literacy.RI.6.3
47	CR	2	CCSS.ELA-Literacy.RI.6.3
48	CR	2	CCSS.ELA-Literacy.RI.6.7
49	CR	2	CCSS.ELA-Literacy.RL.6.1
50	CR	2	CCSS.ELA-Literacy.RL.6.4
51	CR	2	CCSS.ELA-Literacy.RL.6.2
52	CR	4	CCSS.ELA-Literacy.RL.6.3

Table G5. English Language Arts Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.7.1
2	MC	1	CCSS.ELA-Literacy.RI.7.4
3	MC	1	CCSS.ELA-Literacy.RI.7.7
4	MC	1	CCSS.ELA-Literacy.RI.7.2
5	MC	1	CCSS.ELA-Literacy.RI.7.1
6	MC	1	CCSS.ELA-Literacy.RI.7.5
7	MC	1	CCSS.ELA-Literacy.RI.7.2
8	MC	1	CCSS.ELA-Literacy.RL.7.5
9	MC	1	CCSS.ELA-Literacy.RL.7.1
10	MC	1	CCSS.ELA-Literacy.RL.7.3
11	MC	1	CCSS.ELA-Literacy.L.7.4a
12	MC	1	CCSS.ELA-Literacy.RL.7.4
13	MC	1	CCSS.ELA-Literacy.RL.7.4
14	MC	1	CCSS.ELA-Literacy.RL.7.2
15	MC	1	CCSS.ELA-Literacy.RI.7.5
16	MC	1	CCSS.ELA-Literacy.RI.7.1
17	MC	1	CCSS.ELA-Literacy.RI.7.5
18	MC	1	CCSS.ELA-Literacy.RI.7.3
19	MC	1	CCSS.ELA-Literacy.RI.7.1
20	MC	1	CCSS.ELA-Literacy.RI.7.8
21	MC	1	CCSS.ELA-Literacy.RI.7.2
22	MC	1	CCSS.ELA-Literacy.RL.7.4
23	MC	1	CCSS.ELA-Literacy.RL.7.5
24	MC	1	CCSS.ELA-Literacy.RL.7.1

Table G5. English Language Arts Grade 7 Operational Item Map (cont.)

Item	Type	Points	Standard
25	MC	1	CCSS.ELA-Literacy.RL.7.3
26	MC	1	CCSS.ELA-Literacy.RL.7.4
27	MC	1	CCSS.ELA-Literacy.RL.7.3
28	MC	1	CCSS.ELA-Literacy.RL.7.6
29	MC	1	CCSS.ELA-Literacy.RI.7.5
30	MC	1	CCSS.ELA-Literacy.RI.7.8
31	MC	1	CCSS.ELA-Literacy.RI.7.3
32	MC	1	CCSS.ELA-Literacy.RI.7.4
33	MC	1	CCSS.ELA-Literacy.RI.7.2
34	MC	1	CCSS.ELA-Literacy.RI.7.3
35	MC	1	CCSS.ELA-Literacy.RI.7.2
36	MC	1	CCSS.ELA-Literacy.RI.7.4
37	MC	1	CCSS.ELA-Literacy.RI.7.5
38	MC	1	CCSS.ELA-Literacy.RI.7.1
39	MC	1	CCSS.ELA-Literacy.RI.7.1
40	MC	1	CCSS.ELA-Literacy.RI.7.6
41	MC	1	CCSS.ELA-Literacy.RI.7.8
42	MC	1	CCSS.ELA-Literacy.RI.7.7
43	CR	2	CCSS.ELA-Literacy.RI.7.1
44	CR	2	CCSS.ELA-Literacy.RI.7.2
45	CR	2	CCSS.ELA-Literacy.RL.7.5
46	CR	4	CCSS.ELA-Literacy.RL.7.3
47	CR	2	CCSS.ELA-Literacy.RL.7.3
48	CR	2	CCSS.ELA-Literacy.RL.7.1
49	CR	2	CCSS.ELA-Literacy.RI.7.5
50	CR	2	CCSS.ELA-Literacy.RI.7.2
51	CR	2	CCSS.ELA-Literacy.RI.7.1
52	CR	4	CCSS.ELA-Literacy.RI.7.9

Table G6. English Language Arts Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.8.4
2	MC	1	CCSS.ELA-Literacy.RI.8.5
3	MC	1	CCSS.ELA-Literacy.RI.8.3
4	MC	1	CCSS.ELA-Literacy.RI.8.2
5	MC	1	CCSS.ELA-Literacy.RI.8.8
6	MC	1	CCSS.ELA-Literacy.RI.8.5
7	MC	1	CCSS.ELA-Literacy.RI.8.6
8	MC	1	CCSS.ELA-Literacy.RL.8.3
9	MC	1	CCSS.ELA-Literacy.RL.8.1
10	MC	1	CCSS.ELA-Literacy.RL.8.2
11	MC	1	CCSS.ELA-Literacy.RL.8.6
12	MC	1	CCSS.ELA-Literacy.RL.8.4
13	MC	1	CCSS.ELA-Literacy.RL.8.3
14	MC	1	CCSS.ELA-Literacy.RL.8.2
15	MC	1	CCSS.ELA-Literacy.RI.8.1
16	MC	1	CCSS.ELA-Literacy.RI.8.3
17	MC	1	CCSS.ELA-Literacy.RI.8.1
18	MC	1	CCSS.ELA-Literacy.RI.8.4
19	MC	1	CCSS.ELA-Literacy.RI.8.3
20	MC	1	CCSS.ELA-Literacy.L.8.4a
21	MC	1	CCSS.ELA-Literacy.RI.8.8
22	MC	1	CCSS.ELA-Literacy.RI.8.1
23	MC	1	CCSS.ELA-Literacy.RI.8.6
24	MC	1	CCSS.ELA-Literacy.RI.8.6
25	MC	1	CCSS.ELA-Literacy.RI.8.3
26	MC	1	CCSS.ELA-Literacy.RI.8.8
27	MC	1	CCSS.ELA-Literacy.RI.8.4
28	MC	1	CCSS.ELA-Literacy.RI.8.1
29	MC	1	CCSS.ELA-Literacy.RL.8.3
30	MC	1	CCSS.ELA-Literacy.RL.8.4
31	MC	1	CCSS.ELA-Literacy.RL.8.1
32	MC	1	CCSS.ELA-Literacy.RL.8.3
33	MC	1	CCSS.ELA-Literacy.RL.8.6
34	MC	1	CCSS.ELA-Literacy.RL.8.6
35	MC	1	CCSS.ELA-Literacy.RL.8.2
36	MC	1	CCSS.ELA-Literacy.RI.8.5
37	MC	1	CCSS.ELA-Literacy.RI.8.1
38	MC	1	CCSS.ELA-Literacy.RI.8.3
39	MC	1	CCSS.ELA-Literacy.RI.8.5
40	MC	1	CCSS.ELA-Literacy.RI.8.1

Table G6. English Language Arts Grade 8 Operational Item Map (cont.)

Item	Type	Points	Standard
41	MC	1	CCSS.ELA-Literacy.RI.8.2
42	MC	1	CCSS.ELA-Literacy.RI.8.2
43	CR	2	CCSS.ELA-Literacy.RI.8.3
44	CR	2	CCSS.ELA-Literacy.RI.8.4
45	CR	2	CCSS.ELA-Literacy.RL.8.1
46	CR	4	CCSS.ELA-Literacy.RL.8.2
47	CR	2	CCSS.ELA-Literacy.RL.8.4
48	CR	2	CCSS.ELA-Literacy.RL.8.3
49	CR	2	CCSS.ELA-Literacy.RI.8.1
50	CR	2	CCSS.ELA-Literacy.RI.8.5
51	CR	2	CCSS.ELA-Literacy.RI.8.1
52	CR	4	CCSS.ELA-Literacy.RI.8.1

Table G7. Mathematics Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.3.OA.A.1
2	MC	1	CCSS.Math.Content.3.NF.A.1
3	MC	1	CCSS.Math.Content.3.OA.D.9
4	MC	1	CCSS.Math.Content.3.G.A.2
5	MC	1	CCSS.Math.Content.3.MD.C.5b
6	MC	1	CCSS.Math.Content.3.OA.A.4
7	MC	1	CCSS.Math.Content.3.NBT.A.3
8	MC	1	CCSS.Math.Content.3.OA.D.8
9	MC	1	CCSS.Math.Content.3.OA.A.3
10	MC	1	CCSS.Math.Content.3.MD.B.3
11	MC	1	CCSS.Math.Content.3.OA.B.5
12	MC	1	CCSS.Math.Content.3.NBT.A.1
13	MC	1	CCSS.Math.Content.3.NF.A.1
14	MC	1	CCSS.Math.Content.3.MD.A.1
15	MC	1	CCSS.Math.Content.3.MD.C.7d
16	MC	1	CCSS.Math.Content.3.MD.A.2
17	MC	1	CCSS.Math.Content.3.OA.A.2
18	MC	1	CCSS.Math.Content.3.MD.B.3
19	MC	1	CCSS.Math.Content.3.NF.A.3a
20	MC	1	CCSS.Math.Content.3.NF.A.1
21	MC	1	CCSS.Math.Content.3.OA.B.5
22	MC	1	CCSS.Math.Content.3.OA.A.1
23	MC	1	CCSS.Math.Content.3.OA.A.2
24	MC	1	CCSS.Math.Content.3.NF.A.3b

Table G7. Mathematics Grade 3 Operational Item Map (cont.)

Item	Type	Points	Standard
25	MC	1	CCSS.Math.Content.3.OA.B.6
26	MC	1	CCSS.Math.Content.3.NBT.A.3
27	MC	1	CCSS.Math.Content.3.OA.A.3
28	MC	1	CCSS.Math.Content.3.MD.A.1
29	MC	1	CCSS.Math.Content.3.MD.A.2
30	MC	1	CCSS.Math.Content.3.OA.D.8
31	MC	1	CCSS.Math.Content.3.NBT.A.1
32	MC	1	CCSS.Math.Content.3.MD.C.7c
33	MC	1	CCSS.Math.Content.3.NF.A.2b
34	MC	1	CCSS.Math.Content.3.OA.A.4
35	MC	1	CCSS.Math.Content.3.OA.A.3
36	MC	1	CCSS.Math.Content.3.G.A.2
37	MC	1	CCSS.Math.Content.3.MD.B.3
38	MC	1	CCSS.Math.Content.3.OA.D.9
39	MC	1	CCSS.Math.Content.3.OA.A.3
40	MC	1	CCSS.Math.Content.3.OA.B.6
41	MC	1	CCSS.Math.Content.3.NF.A.3b
42	CR	2	CCSS.Math.Content.3.NF.A.2a
43	CR	2	CCSS.Math.Content.3.OA.A.3
44	CR	2	CCSS.Math.Content.3.OA.A.3
45	CR	2	CCSS.Math.Content.3.OA.D.8
46	CR	2	CCSS.Math.Content.3.MD.C.6
47	CR	3	CCSS.Math.Content.3.OA.D.8
48	CR	3	CCSS.Math.Content.3.MD.C.7c
49	CR	3	CCSS.Math.Content.3.NF.A.2b

Table G8. Mathematics Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.4.NBT.A.2
2	MC	1	CCSS.Math.Content.3.MD.B.4
3	MC	1	CCSS.Math.Content.4.NF.A.1
4	MC	1	CCSS.Math.Content.4.MD.C.6
5	MC	1	CCSS.Math.Content.4.OA.A.1
6	MC	1	CCSS.Math.Content.4.NBT.A.1
7	MC	1	CCSS.Math.Content.4.NBT.A.3
8	MC	1	CCSS.Math.Content.4.OA.A.3
9	MC	1	CCSS.Math.Content.4.NF.A.1
10	MC	1	CCSS.Math.Content.3.MD.B.4
11	MC	1	CCSS.Math.Content.4.MD.C.5a

Table G8. Mathematics Grade 4 Operational Item Map (cont.)

Item	Type	Points	Standard
12	MC	1	CCSS.Math.Content.4.G.A.3
13	MC	1	CCSS.Math.Content.4.NBT.B.5
14	MC	1	CCSS.Math.Content.4.NF.B.4a
15	MC	1	CCSS.Math.Content.4.OA.A.2
16	MC	1	CCSS.Math.Content.4.NBT.B.6
17	MC	1	CCSS.Math.Content.3.MD.D.8
18	MC	1	CCSS.Math.Content.4.NBT.B.5
19	MC	1	CCSS.Math.Content.4.OA.A.2
20	MC	1	CCSS.Math.Content.4.NF.B.3a
21	MC	1	CCSS.Math.Content.4.NF.A.2
22	MC	1	CCSS.Math.Content.4.OA.C.5
23	MC	1	CCSS.Math.Content.4.NF.B.3a
24	MC	1	CCSS.Math.Content.4.NF.B.4c
25	MC	1	CCSS.Math.Content.4.MD.A.3
26	MC	1	CCSS.Math.Content.4.NBT.B.6
27	MC	1	CCSS.Math.Content.4.NBT.B.5
28	MC	1	CCSS.Math.Content.4.G.A.2
29	MC	1	CCSS.Math.Content.4.NF.B.4b
30	MC	1	CCSS.Math.Content.4.OA.B.4
31	MC	1	CCSS.Math.Content.4.OA.A.2
32	MC	1	CCSS.Math.Content.4.NF.A.1
33	MC	1	CCSS.Math.Content.4.NBT.A.1
34	MC	1	CCSS.Math.Content.3.G.A.1
35	MC	1	CCSS.Math.Content.4.NBT.B.5
36	MC	1	CCSS.Math.Content.4.MD.A.3
37	MC	1	CCSS.Math.Content.4.MD.C.5b
38	MC	1	CCSS.Math.Content.4.MD.A.3
39	MC	1	CCSS.Math.Content.4.NF.A.1
40	MC	1	CCSS.Math.Content.4.NBT.A.1
41	MC	1	CCSS.Math.Content.4.G.A.1
42	MC	1	CCSS.Math.Content.4.MD.C.6
43	CR	2	CCSS.Math.Content.4.MD.C.7
44	CR	2	CCSS.Math.Content.4.NBT.B.6
45	CR	2	CCSS.Math.Content.4.NBT.B.5
46	CR	2	CCSS.Math.Content.4.G.A.1
47	CR	2	CCSS.Math.Content.4.NBT.A.2
48	CR	2	CCSS.Math.Content.4.NF.B.4c
49	CR	3	CCSS.Math.Content.4.NF.B.3c
50	CR	3	CCSS.Math.Content.4.OA.A.3
51	CR	3	CCSS.Math.Content.4.OA.A.2
52	CR	3	CCSS.Math.Content.4.NF.A.2

Table G9. Mathematics Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.5.OA.A.2
2	MC	1	CCSS.Math.Content.5.MD.C.5b
3	MC	1	CCSS.Math.Content.5.NBT.B.6
4	MC	1	CCSS.Math.Content.4.NF.C.7
5	MC	1	CCSS.Math.Content.5.G.B.4
6	MC	1	CCSS.Math.Content.5.MD.C.5b
7	MC	1	CCSS.Math.Content.5.NBT.B.7
8	MC	1	CCSS.Math.Content.5.NF.B.6
9	MC	1	CCSS.Math.Content.5.OA.A.1
10	MC	1	CCSS.Math.Content.5.NF.B.6
11	MC	1	CCSS.Math.Content.5.NBT.B.6
12	MC	1	CCSS.Math.Content.4.NF.C.5
13	MC	1	CCSS.Math.Content.5.NF.B.4a
14	MC	1	CCSS.Math.Content.5.MD.A.1
15	MC	1	CCSS.Math.Content.5.NF.B.3
16	MC	1	CCSS.Math.Content.5.MD.C.3b
17	MC	1	CCSS.Math.Content.5.MD.C.3b
18	MC	1	CCSS.Math.Content.5.OA.A.2
19	MC	1	CCSS.Math.Content.5.NF.B.6
20	MC	1	CCSS.Math.Content.5.NBT.A.1
21	MC	1	CCSS.Math.Content.5.NF.B.5a
22	MC	1	CCSS.Math.Content.5.NBT.A.1
23	MC	1	CCSS.Math.Content.4.NF.C.6
24	MC	1	CCSS.Math.Content.4.MD.A.2
25	MC	1	CCSS.Math.Content.5.MD.C.4
26	MC	1	CCSS.Math.Content.5.NF.A.1
27	MC	1	CCSS.Math.Content.5.NBT.A.2
28	MC	1	CCSS.Math.Content.5.NF.B.5b
29	MC	1	CCSS.Math.Content.5.MD.C.4
30	MC	1	CCSS.Math.Content.5.G.B.3
31	MC	1	CCSS.Math.Content.5.NF.B.7a
32	MC	1	CCSS.Math.Content.5.NF.B.3
33	MC	1	CCSS.Math.Content.5.NBT.A.4
34	MC	1	CCSS.Math.Content.5.MD.A.1
35	MC	1	CCSS.Math.Content.5.MD.C.5b
36	MC	1	CCSS.Math.Content.5.NBT.B.7
37	MC	1	CCSS.Math.Content.5.MD.A.1
38	MC	1	CCSS.Math.Content.5.MD.A.1
39	MC	1	CCSS.Math.Content.5.NF.A.2
40	MC	1	CCSS.Math.Content.5.NF.B.5a

Table G9. Mathematics Grade 5 Operational Item Map (cont.)

Item	Type	Points	Standard
41	MC	1	CCSS.Math.Content.5.NBT.B.7
42	MC	1	CCSS.Math.Content.5.OA.A.1
43	CR	2	CCSS.Math.Content.5.NBT.A.3a
44	CR	2	CCSS.Math.Content.5.NF.B.4b
45	CR	2	CCSS.Math.Content.5.NF.B.7a
46	CR	2	CCSS.Math.Content.5.MD.B.2
47	CR	2	CCSS.Math.Content.5.NBT.B.7
48	CR	2	CCSS.Math.Content.5.NF.A.2
49	CR	3	CCSS.Math.Content.5.NF.B.6
50	CR	3	CCSS.Math.Content.5.NBT.B.6
51	CR	3	CCSS.Math.Content.5.MD.C.5b
52	CR	3	CCSS.Math.Content.5.NF.A.2

Table G10. Mathematics Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.6.EE.A.2c
2	MC	1	CCSS.Math.Content.6.NS.C.6c
3	MC	1	CCSS.Math.Content.6.RP.A.1
4	MC	1	CCSS.Math.Content.6.G.A.4
5	MC	1	CCSS.Math.Content.6.RP.A.3b
6	MC	1	CCSS.Math.Content.6.EE.B.6
7	MC	1	CCSS.Math.Content.5.OA.B.3
8	MC	1	CCSS.Math.Content.6.EE.B.8
9	MC	1	CCSS.Math.Content.5.OA.B.3
10	MC	1	CCSS.Math.Content.6.RP.A.2
11	MC	1	CCSS.Math.Content.6.EE.A.4
12	MC	1	CCSS.Math.Content.6.NS.C.6c
13	MC	1	CCSS.Math.Content.6.NS.A.1
14	MC	1	CCSS.Math.Content.6.EE.B.7
15	MC	1	CCSS.Math.Content.6.NS.A.1
16	MC	1	CCSS.Math.Content.6.RP.A.2
17	MC	1	CCSS.Math.Content.6.RP.A.3a
18	MC	1	CCSS.Math.Content.6.G.A.1
19	MC	1	CCSS.Math.Content.6.RP.A.2
20	MC	1	CCSS.Math.Content.6.NS.A.1
21	MC	1	CCSS.Math.Content.6.NS.A.1
22	MC	1	CCSS.Math.Content.5.G.A.2
23	MC	1	CCSS.Math.Content.6.EE.B.7
24	MC	1	CCSS.Math.Content.6.RP.A.3b

Table G10. Mathematics Grade 6 Operational Item Map (cont.)

Item	Type	Points	Standard
25	MC	1	CCSS.Math.Content.6.G.A.1
26	MC	1	CCSS.Math.Content.5.OA.B.3
27	MC	1	CCSS.Math.Content.6.EE.A.4
28	MC	1	CCSS.Math.Content.6.EE.B.6
29	MC	1	CCSS.Math.Content.6.RP.A.3a
30	MC	1	CCSS.Math.Content.6.NS.B.4
31	MC	1	CCSS.Math.Content.6.G.A.2
32	MC	1	CCSS.Math.Content.6.RP.A.1
33	MC	1	CCSS.Math.Content.6.NS.C.5
34	MC	1	CCSS.Math.Content.6.G.A.2
35	MC	1	CCSS.Math.Content.6.RP.A.2
36	MC	1	CCSS.Math.Content.6.RP.A.3b
37	MC	1	CCSS.Math.Content.6.EE.A.4
38	MC	1	CCSS.Math.Content.5.OA.B.3
39	MC	1	CCSS.Math.Content.6.EE.B.5
40	MC	1	CCSS.Math.Content.6.RP.A.3c
41	MC	1	CCSS.Math.Content.6.G.A.4
42	MC	1	CCSS.Math.Content.6.G.A.1
43	MC	1	CCSS.Math.Content.6.EE.B.5
44	MC	1	CCSS.Math.Content.6.EE.A.3
45	MC	1	CCSS.Math.Content.6.RP.A.3b
46	MC	1	CCSS.Math.Content.6.NS.B.4
47	MC	1	CCSS.Math.Content.6.NS.C.7a
48	MC	1	CCSS.Math.Content.6.G.A.2
49	CR	2	CCSS.Math.Content.6.RP.A.3b
50	CR	2	CCSS.Math.Content.6.EE.A.1
51	CR	2	CCSS.Math.Content.6.NS.B.4
52	CR	2	CCSS.Math.Content.6.NS.C.8
53	CR	2	CCSS.Math.Content.6.EE.A.2a
54	CR	2	CCSS.Math.Content.6.G.A.3
55	CR	3	CCSS.Math.Content.6.EE.C.9
56	CR	3	CCSS.Math.Content.6.EE.A.3
57	CR	3	CCSS.Math.Content.6.EE.B.7
58	CR	3	CCSS.Math.Content.6.RP.A.3a

Table G11. Mathematics Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.7.NS.A.1d
2	MC	1	CCSS.Math.Content.7.EE.B.4a
3	MC	1	CCSS.Math.Content.7.NS.A.2d
4	MC	1	CCSS.Math.Content.6.SP.B.4
5	MC	1	CCSS.Math.Content.7.NS.A.1a
6	MC	1	CCSS.Math.Content.7.SP.C.7b
7	MC	1	CCSS.Math.Content.7.RP.A.3
8	MC	1	CCSS.Math.Content.7.EE.A.1
9	MC	1	CCSS.Math.Content.7.RP.A.2d
10	MC	1	CCSS.Math.Content.7.NS.A.3
11	MC	1	CCSS.Math.Content.7.SP.C.6
12	MC	1	CCSS.Math.Content.7.EE.A.2
13	MC	1	CCSS.Math.Content.7.NS.A.3
14	MC	1	CCSS.Math.Content.7.RP.A.1
15	MC	1	CCSS.Math.Content.7.RP.A.3
16	MC	1	CCSS.Math.Content.7.NS.A.1b
17	MC	1	CCSS.Math.Content.7.NS.A.2a
18	MC	1	CCSS.Math.Content.7.EE.B.3
19	MC	1	CCSS.Math.Content.7.RP.A.1
20	MC	1	CCSS.Math.Content.7.NS.A.2d
21	MC	1	CCSS.Math.Content.7.EE.A.1
22	MC	1	CCSS.Math.Content.7.NS.A.3
23	MC	1	CCSS.Math.Content.7.EE.A.1
24	MC	1	CCSS.Math.Content.7.SP.C.6
25	MC	1	CCSS.Math.Content.7.G.B.4
26	MC	1	CCSS.Math.Content.7.SP.A.1
27	MC	1	CCSS.Math.Content.7.SP.B.4
28	MC	1	CCSS.Math.Content.7.RP.A.3
29	MC	1	CCSS.Math.Content.7.EE.A.2
30	MC	1	CCSS.Math.Content.7.SP.C.6
31	MC	1	CCSS.Math.Content.7.NS.A.3
32	MC	1	CCSS.Math.Content.7.G.B.4
33	MC	1	CCSS.Math.Content.7.RP.A.3
34	MC	1	CCSS.Math.Content.7.EE.A.1
35	MC	1	CCSS.Math.Content.7.NS.A.3
36	MC	1	CCSS.Math.Content.7.EE.B.4b
37	MC	1	CCSS.Math.Content.7.RP.A.1
38	MC	1	CCSS.Math.Content.7.SP.B.4
39	MC	1	CCSS.Math.Content.7.EE.A.1
40	MC	1	CCSS.Math.Content.7.RP.A.2c

Table G11. Mathematics Grade 7 Operational Item Map (cont.)

Item	Type	Points	Standard
41	MC	1	CCSS.Math.Content.7.SP.C.5
42	MC	1	CCSS.Math.Content.7.RP.A.3
43	MC	1	CCSS.Math.Content.7.EE.A.1
44	MC	1	CCSS.Math.Content.7.G.B.4
45	MC	1	CCSS.Math.Content.7.RP.A.3
46	MC	1	CCSS.Math.Content.7.EE.B.3
47	MC	1	CCSS.Math.Content.7.NS.A.3
48	MC	1	CCSS.Math.Content.7.RP.A.3
49	CR	2	CCSS.Math.Content.7.G.A.1
50	CR	2	CCSS.Math.Content.7.EE.B.4b
51	CR	2	CCSS.Math.Content.7.SP.A.2
52	CR	2	CCSS.Math.Content.7.RP.A.2b
53	CR	2	CCSS.Math.Content.7.EE.B.4a
54	CR	2	CCSS.Math.Content.7.NS.A.3
55	CR	3	CCSS.Math.Content.7.RP.A.2b
56	CR	3	CCSS.Math.Content.7.EE.B.4a
57	CR	3	CCSS.Math.Content.7.RP.A.3
58	CR	3	CCSS.Math.Content.7.EE.B.3

Table G12. Mathematics Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.8.SP.A.2
2	MC	1	CCSS.Math.Content.8.EE.A.3
3	MC	1	CCSS.Math.Content.8.F.A.3
4	MC	1	CCSS.Math.Content.8.G.A.2
5	MC	1	CCSS.Math.Content.8.EE.C.8b
6	MC	1	CCSS.Math.Content.8.F.A.3
7	MC	1	CCSS.Math.Content.8.EE.A.3
8	MC	1	CCSS.Math.Content.8.EE.A.1
9	MC	1	CCSS.Math.Content.8.F.B.5
10	MC	1	CCSS.Math.Content.8.EE.A.1
11	MC	1	CCSS.Math.Content.8.EE.C.8c
12	MC	1	CCSS.Math.Content.8.F.A.2
13	MC	1	CCSS.Math.Content.8.G.A.1a
14	MC	1	CCSS.Math.Content.7.G.B.5
15	MC	1	CCSS.Math.Content.8.F.B.4
16	MC	1	CCSS.Math.Content.8.EE.A.4
17	MC	1	CCSS.Math.Content.8.EE.A.3
18	MC	1	CCSS.Math.Content.8.SP.A.3

Table G12. Mathematics Grade 8 Operational Item Map (cont.)

Item	Type	Points	Standard
19	MC	1	CCSS.Math.Content.8.F.A.2
20	MC	1	CCSS.Math.Content.8.G.A.4
21	MC	1	CCSS.Math.Content.8.SP.A.1
22	MC	1	CCSS.Math.Content.8.G.A.5
23	MC	1	CCSS.Math.Content.8.G.A.2
24	MC	1	CCSS.Math.Content.8.EE.A.1
25	MC	1	CCSS.Math.Content.8.EE.B.5
26	MC	1	CCSS.Math.Content.8.G.C.9
27	MC	1	CCSS.Math.Content.8.EE.C.8c
28	MC	1	CCSS.Math.Content.8.F.A.2
29	MC	1	CCSS.Math.Content.8.SP.A.4
30	MC	1	CCSS.Math.Content.8.EE.B.6
31	MC	1	CCSS.Math.Content.8.EE.C.8b
32	MC	1	CCSS.Math.Content.8.F.A.3
33	MC	1	CCSS.Math.Content.8.EE.B.6
34	MC	1	CCSS.Math.Content.8.EE.A.4
35	MC	1	CCSS.Math.Content.8.F.B.5
36	MC	1	CCSS.Math.Content.7.G.A.3
37	MC	1	CCSS.Math.Content.8.SP.A.3
38	MC	1	CCSS.Math.Content.8.G.C.9
39	MC	1	CCSS.Math.Content.8.F.A.2
40	MC	1	CCSS.Math.Content.8.G.C.9
41	MC	1	CCSS.Math.Content.8.G.A.2
42	MC	1	CCSS.Math.Content.8.EE.B.6
43	MC	1	CCSS.Math.Content.8.EE.C.8a
44	MC	1	CCSS.Math.Content.7.G.B.5
45	MC	1	CCSS.Math.Content.8.EE.C.8b
46	MC	1	CCSS.Math.Content.8.SP.A.3
47	MC	1	CCSS.Math.Content.8.G.C.9
48	MC	1	CCSS.Math.Content.8.EE.B.5
49	CR	2	CCSS.Math.Content.8.EE.C.7a
50	CR	2	CCSS.Math.Content.8.F.A.3
51	CR	2	CCSS.Math.Content.8.EE.C.8b
52	CR	2	CCSS.Math.Content.8.F.B.4
53	CR	2	CCSS.Math.Content.8.EE.C.7a
54	CR	2	CCSS.Math.Content.8.F.A.1
55	CR	3	CCSS.Math.Content.8.F.B.4
56	CR	3	CCSS.Math.Content.8.G.A.3
57	CR	3	CCSS.Math.Content.8.G.A.4
58	CR	3	CCSS.Math.Content.8.EE.B.5

Appendix H: ELA Short-Response Rubric

2-Point Rubric–Short Response

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

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

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

Appendix I: ELA Extended-Response Rubric

New York State Grade 3 Expository Writing Evaluation Rubric

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

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

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

New York State Grade 4-5 Expository Writing Evaluation Rubric

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

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

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

New York State Grade 6-8 Expository Writing Evaluation Rubric

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

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

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

Appendix J: Mathematics Short-Response Rubric

2-Point Holistic Rubric

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

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

Appendix K: Mathematics Extended-Response Rubric

3-Point Holistic Rubric

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

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

Appendix L: Factor Analysis Results for Select Subgroups

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

Table L1. English Language Arts Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**6.34**	**15.47**	**15.47**
		2	1.54	3.76	19.23
		3	1.16	2.82	22.05
		4	1.07	2.62	24.67
		5	1.05	2.57	27.24
		6	1.04	2.53	29.77
		7	1.03	2.50	32.27
		8	1.02	2.48	34.75
		9	1.01	2.46	37.21
		10	1.00	2.45	39.66
SWD	All Codes	**1**	**7.39**	**18.03**	**18.03**
		2	1.61	3.93	21.96
		3	1.17	2.86	24.82
		4	1.06	2.59	27.41
		5	1.04	2.55	29.96
		6	1.01	2.46	32.42
		7	1.00	2.45	34.87
SUA	All Codes	**1**	**7.58**	**18.48**	**18.48**
		2	1.60	3.89	22.37
		3	1.16	2.83	25.20
		4	1.05	2.55	27.75
		5	1.03	2.52	30.27
		6	1.00	2.45	32.72

Table L1. English Language Arts Grade 3 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD/ SUA	SUA=504 plan codes	**1**	**7.03**	**17.14**	**17.14**
		2	1.62	3.95	21.09
		3	1.17	2.86	23.95
		4	1.07	2.62	26.57
		5	1.06	2.58	29.15
		6	1.02	2.49	31.64
		7	1.01	2.47	34.11
ELL/ SUA	SUA & ELL codes	**1**	**6.33**	**15.44**	**15.44**
		2	1.54	3.75	19.19
		3	1.16	2.83	22.02
		4	1.07	2.62	24.64
		5	1.06	2.58	27.22
		6	1.04	2.52	29.74
		7	1.02	2.50	32.24
		8	1.02	2.48	34.72
		9	1.01	2.47	37.19
		10	1.01	2.46	39.65

Table L2. English Language Arts Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**6.12**	**14.92**	**14.92**
		2	1.35	3.28	18.20
		3	1.18	2.87	21.07
		4	1.09	2.65	23.72
		5	1.08	2.63	26.35
		6	1.07	2.60	28.95
		7	1.05	2.57	31.52
		8	1.03	2.51	34.03
		9	1.02	2.49	36.52
		10	1.01	2.45	38.97

Table L2. English Language Arts Grade 4 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD	All Codes	**1**	**7.20**	**17.57**	**17.57**
		2	1.42	3.47	21.04
		3	1.20	2.93	23.97
		4	1.07	2.62	26.59
		5	1.04	2.54	29.13
		6	1.02	2.50	31.63
		7	1.00	2.45	34.08
SUA	All Codes	**1**	**7.49**	**18.26**	**18.26**
		2	1.40	3.42	21.68
		3	1.18	2.87	24.55
		4	1.05	2.56	27.11
		5	1.03	2.51	29.62
		6	1.02	2.48	32.10
SWD/ SUA	SUA=504 plan codes	**1**	**6.92**	**16.87**	**16.87**
		2	1.42	3.46	20.33
		3	1.20	2.94	23.27
		4	1.08	2.63	25.90
		5	1.05	2.55	28.45
		6	1.03	2.51	30.96
		7	1.01	2.46	33.42
ELL/ SUA	SUA & ELL codes	**1**	**6.09**	**14.86**	**14.86**
		2	1.35	3.29	18.15
		3	1.18	2.88	21.03
		4	1.09	2.66	23.69
		5	1.08	2.63	26.32
		6	1.07	2.61	28.93
		7	1.05	2.57	31.50
		8	1.03	2.51	34.01
		9	1.02	2.49	36.50
		10	1.01	2.45	38.95

Table L3. English Language Arts Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**7.63**	**14.67**	**14.67**
		2	1.61	3.10	17.77
		3	1.23	2.36	20.13
		4	1.14	2.20	22.33
		5	1.12	2.16	24.49
		6	1.07	2.05	26.54
		7	1.05	2.03	28.57
		8	1.05	2.02	30.59
		9	1.04	2.00	32.59
		10	1.03	1.98	34.57
SWD	All Codes	**1**	**8.86**	**17.05**	**17.05**
		2	1.65	3.17	20.22
		3	1.24	2.39	22.61
		4	1.11	2.13	24.74
		5	1.09	2.09	26.83
		6	1.05	2.02	28.85
		7	1.02	1.97	30.82
		8	1.02	1.96	32.78
		9	1.01	1.94	34.72
SUA	All Codes	**1**	**9.39**	**18.05**	**18.05**
		2	1.64	3.15	21.20
		3	1.24	2.39	23.59
		4	1.10	2.12	25.71
		5	1.07	2.06	27.77
		6	1.04	1.99	29.76
		7	1.02	1.96	31.72
SWD/ SUA	SUA=504 plan codes	**1**	**8.63**	**16.60**	**16.60**
		2	1.64	3.16	19.76
		3	1.24	2.38	22.14
		4	1.11	2.13	24.27
		5	1.10	2.11	26.38
		6	1.05	2.03	28.41
		7	1.03	1.98	30.39
		8	1.02	1.96	32.35
		9	1.01	1.95	34.30

Table L3. English Language Arts Grade 5 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/ SUA	SUA & ELL codes	**1**	**7.60**	**14.62**	**14.62**
		2	1.62	3.11	17.73
		3	1.22	2.36	20.09
		4	1.15	2.21	22.30
		5	1.12	2.16	24.46
		6	1.07	2.05	26.51
		7	1.06	2.03	28.54
		8	1.05	2.01	30.55
		9	1.04	2.00	32.55
		10	1.03	1.98	34.53

Table L4. English Language Arts Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**6.59**	**12.66**	**12.66**
		2	1.54	2.96	15.62
		3	1.22	2.35	17.97
		4	1.14	2.20	20.17
		5	1.13	2.17	22.34
		6	1.10	2.12	24.46
		7	1.09	2.09	26.55
		8	1.08	2.07	28.62
		9	1.07	2.05	30.67
		10	1.06	2.04	32.71
SWD	All Codes	**1**	**7.87**	**15.13**	**15.13**
		2	1.67	3.21	18.34
		3	1.21	2.33	20.67
		4	1.17	2.25	22.92
		5	1.09	2.09	25.01
		6	1.07	2.05	27.06
		7	1.04	2.01	29.07
		8	1.04	2.01	31.08
		9	1.03	1.99	33.07
		10	1.02	1.96	35.03

Table L4. English Language Arts Grade 6 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	**1**	**8.40**	**16.15**	**16.15**
		2	1.65	3.17	19.32
		3	1.20	2.31	21.63
		4	1.16	2.23	23.86
		5	1.06	2.03	25.89
		6	1.04	2.00	27.89
		7	1.03	1.98	29.87
		8	1.03	1.98	31.85
		9	1.02	1.96	33.81
SWD/ SUA	SUA=504 plan codes	**1**	**7.63**	**14.68**	**14.68**
		2	1.67	3.22	17.90
		3	1.21	2.32	20.22
		4	1.16	2.24	22.46
		5	1.09	2.10	24.56
		6	1.07	2.06	26.62
		7	1.05	2.02	28.64
		8	1.05	2.02	30.66
		9	1.03	1.98	32.64
		10	1.03	1.98	34.62
ELL/ SUA	SUA & ELL codes	**1**	**6.56**	**12.62**	**12.62**
		2	1.54	2.96	15.58
		3	1.22	2.35	17.93
		4	1.14	2.20	20.13
		5	1.13	2.17	22.30
		6	1.10	2.12	24.42
		7	1.09	2.09	26.51
		8	1.08	2.07	28.58
		9	1.07	2.05	30.63
		10	1.06	2.05	32.68

Table L5. English Language Arts Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**6.14**	**11.80**	**11.80**
		2	1.64	3.15	14.95
		3	1.25	2.41	17.36
		4	1.17	2.24	19.60
		5	1.14	2.19	21.79
		6	1.13	2.18	23.97
		7	1.11	2.14	26.11
		8	1.10	2.11	28.22
		9	1.09	2.09	30.31
		10	1.08	2.08	32.39
SWD	All Codes	**1**	**7.21**	**13.86**	**13.86**
		2	1.98	3.82	17.68
		3	1.27	2.45	20.13
		4	1.13	2.17	22.30
		5	1.09	2.11	24.41
		6	1.07	2.06	26.47
		7	1.06	2.04	28.51
		8	1.04	1.99	30.50
		9	1.03	1.98	32.48
		10	1.01	1.95	34.43
SUA	All Codes	**1**	**7.72**	**14.84**	**14.84**
		2	1.99	3.83	18.67
		3	1.26	2.41	21.08
		4	1.11	2.13	23.21
		5	1.07	2.06	25.27
		6	1.05	2.03	27.30
		7	1.04	2.00	29.30
		8	1.03	1.98	31.28
		9	1.01	1.94	33.22
		10	1.00	1.93	35.15

Table L5. English Language Arts Grade 7 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD/ SUA	SUA=504 plan codes	**1**	**7.04**	**13.54**	**13.54**
		2	1.95	3.76	17.30
		3	1.27	2.44	19.74
		4	1.13	2.18	21.92
		5	1.09	2.10	24.02
		6	1.08	2.07	26.09
		7	1.06	2.04	28.13
		8	1.04	2.01	30.14
		9	1.03	1.98	32.12
		10	1.02	1.96	34.08
ELL/ SUA	SUA & ELL codes	**1**	**6.13**	**11.79**	**11.79**
		2	1.62	3.12	14.91
		3	1.25	2.41	17.32
		4	1.16	2.23	19.55
		5	1.15	2.20	21.75
		6	1.14	2.19	23.94
		7	1.12	2.15	26.09
		8	1.10	2.11	28.20
		9	1.08	2.09	30.29
		10	1.08	2.08	32.37

Table L6. English Language Arts Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**6.79**	**13.06**	**13.06**
		2	1.76	3.39	16.45
		3	1.22	2.34	18.79
		4	1.18	2.27	21.06
		5	1.13	2.17	23.23
		6	1.12	2.16	25.39
		7	1.12	2.15	27.54
		8	1.09	2.09	29.63
		9	1.08	2.07	31.70
		10	1.07	2.05	33.75

Table L6. English Language Arts Grade 8 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD	All Codes	**1**	**8.21**	**15.79**	**15.79**
		2	2.03	3.91	19.70
		3	1.24	2.38	22.08
		4	1.12	2.15	24.23
		5	1.06	2.05	26.28
		6	1.06	2.03	28.31
		7	1.05	2.01	30.32
		8	1.04	1.99	32.31
		9	1.01	1.95	34.26
SUA	All Codes	**1**	**8.77**	**16.87**	**16.87**
		2	1.99	3.84	20.71
		3	1.22	2.35	23.06
		4	1.10	2.12	25.18
		5	1.05	2.03	27.21
		6	1.05	2.01	29.22
		7	1.03	1.99	31.21
		8	1.02	1.97	33.18
		9	1.00	1.93	35.11
SWD/ SUA	SUA=504 plan codes	**1**	**8.05**	**15.47**	**15.47**
		2	2.01	3.87	19.34
		3	1.23	2.37	21.71
		4	1.12	2.15	23.86
		5	1.07	2.05	25.91
		6	1.06	2.04	27.95
		7	1.05	2.02	29.97
		8	1.04	2.00	31.97
		9	1.01	1.94	33.91
		10	1.00	1.93	35.84
ELL/ SUA	SUA & ELL codes	**1**	**6.82**	**13.11**	**13.11**
		2	1.77	3.40	16.51
		3	1.23	2.36	18.87
		4	1.18	2.27	21.14
		5	1.13	2.17	23.31
		6	1.12	2.16	25.47
		7	1.12	2.16	27.63
		8	1.09	2.09	29.72
		9	1.08	2.08	31.80
		10	1.06	2.04	33.84

Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**11.07**	**22.59**	**22.59**
		2	1.89	3.86	26.45
		3	1.25	2.55	29.00
		4	1.17	2.38	31.38
		5	1.01	2.07	33.45
SWD	All Codes	**1**	**11.51**	**23.50**	**23.50**
		2	1.77	3.61	27.11
		3	1.30	2.65	29.76
		4	1.12	2.28	32.04
		5	1.02	2.08	34.12
SUA	All Codes	**1**	**11.80**	**24.08**	**24.08**
		2	1.83	3.73	27.81
		3	1.27	2.59	30.40
		4	1.12	2.29	32.69
SWD/ SUA	SUA=504 plan codes	**1**	**10.98**	**22.41**	**22.41**
		2	1.77	3.60	26.01
		3	1.32	2.70	28.71
		4	1.12	2.29	31.00
		5	1.03	2.11	33.11
		6	1.01	2.06	35.17
ELL/ SUA	SUA & ELL codes	**1**	**11.03**	**22.51**	**22.51**
		2	1.89	3.86	26.37
		3	1.25	2.56	28.93
		4	1.17	2.39	31.32
		5	1.02	2.07	33.39

Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**12.50**	**24.04**	**24.04**
		2	1.56	3.00	27.04
		3	1.30	2.50	29.54
		4	1.20	2.31	31.85
		5	1.12	2.16	34.01
		6	1.02	1.96	35.97
SWD	All Codes	**1**	**12.67**	**24.37**	**24.37**
		2	1.47	2.83	27.20
		3	1.39	2.67	29.87
		4	1.16	2.23	32.10
		5	1.11	2.13	34.23
		6	1.00	1.92	36.15
SUA	All Codes	**1**	**13.59**	**26.13**	**26.13**
		2	1.52	2.92	29.05
		3	1.35	2.60	31.65
		4	1.17	2.26	33.91
		5	1.09	2.10	36.01
SWD/ SUA	SUA=504 plan codes	**1**	**12.12**	**23.31**	**23.31**
		2	1.47	2.82	26.13
		3	1.40	2.70	28.83
		4	1.17	2.25	31.08
		5	1.13	2.18	33.26
		6	1.01	1.95	35.21
ELL/ SUA	SUA & ELL codes	**1**	**12.44**	**23.93**	**23.93**
		2	1.55	2.98	26.91
		3	1.30	2.50	29.41
		4	1.20	2.31	31.72
		5	1.13	2.17	33.89
		6	1.02	1.96	35.85

Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**10.24**	**19.70**	**19.70**
		2	2.19	4.22	23.92
		3	1.18	2.28	26.20
		4	1.08	2.07	28.27
		5	1.06	2.05	30.32
		6	1.02	1.97	32.29
		7	1.01	1.93	34.22
SWD	All Codes	**1**	**10.33**	**19.87**	**19.87**
		2	2.04	3.92	23.79
		3	1.19	2.28	26.07
		4	1.07	2.06	28.13
		5	1.04	1.99	30.12
		6	1.00	1.93	32.05
SUA	All Codes	**1**	**11.27**	**21.67**	**21.67**
		2	2.13	4.10	25.77
		3	1.17	2.25	28.02
		4	1.05	2.02	30.04
		5	1.02	1.96	32.00
SWD/ SUA	SUA=504 plan codes	**1**	**9.99**	**19.22**	**19.22**
		2	2.01	3.87	23.09
		3	1.19	2.29	25.38
		4	1.07	2.06	27.44
		5	1.05	2.01	29.45
		6	1.01	1.94	31.39
		7	1.00	1.93	33.32
ELL/ SUA	SUA & ELL codes	**1**	**10.26**	**19.74**	**19.74**
		2	2.19	4.22	23.96
		3	1.19	2.28	26.24
		4	1.08	2.07	28.31
		5	1.07	2.05	30.36
		6	1.02	1.97	32.33
		7	1.01	1.93	34.26

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**11.23**	**19.36**	**19.36**
		2	1.60	2.76	22.12
		3	1.28	2.21	24.33
		4	1.11	1.91	26.24
		5	1.07	1.84	28.08
		6	1.06	1.83	29.91
		7	1.01	1.74	31.65
		8	1.01	1.74	33.39
SWD	All Codes	**1**	**10.55**	**18.19**	**18.19**
		2	1.54	2.65	20.84
		3	1.39	2.39	23.23
		4	1.11	1.91	25.14
		5	1.08	1.86	27.00
		6	1.05	1.81	28.81
		7	1.02	1.76	30.57
		8	1.01	1.74	32.31
SUA	All Codes	**1**	**11.99**	**20.68**	**20.68**
		2	1.60	2.76	23.44
		3	1.37	2.36	25.80
		4	1.09	1.88	27.68
		5	1.06	1.83	29.51
		6	1.02	1.75	31.26
		7	1.00	1.73	32.99
SWD/ SUA	SUA=504 plan codes	**1**	**10.06**	**17.35**	**17.35**
		2	1.52	2.62	19.97
		3	1.40	2.41	22.38
		4	1.11	1.92	24.30
		5	1.09	1.87	26.17
		6	1.06	1.82	27.99
		7	1.03	1.78	29.77
		8	1.02	1.76	31.53
		9	1.00	1.73	33.26

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/ SUA	SUA & ELL codes	**1**	**11.27**	**19.43**	**19.43**
		2	1.60	2.76	22.19
		3	1.29	2.22	24.41
		4	1.11	1.91	26.32
		5	1.07	1.84	28.16
		6	1.07	1.84	30.00
		7	1.01	1.75	31.75
		8	1.01	1.74	33.49

Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**10.13**	**17.46**	**17.46**
		2	1.93	3.32	20.78
		3	1.25	2.15	22.93
		4	1.12	1.93	24.86
		5	1.11	1.92	26.78
		6	1.07	1.84	28.62
		7	1.05	1.81	30.43
		8	1.02	1.76	32.19
		9	1.02	1.75	33.94
		10	1.00	1.72	35.66
SWD	All Codes	**1**	**8.97**	**15.46**	**15.46**
		2	1.80	3.10	18.56
		3	1.30	2.24	20.80
		4	1.13	1.94	22.74
		5	1.12	1.93	24.67
		6	1.07	1.85	26.52
		7	1.04	1.80	28.32
		8	1.04	1.78	30.10
		9	1.02	1.77	31.87
		10	1.02	1.75	33.62

Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	**1**	**10.60**	**18.28**	**18.28**
		2	1.92	3.32	21.60
		3	1.28	2.21	23.81
		4	1.11	1.91	25.72
		5	1.08	1.86	27.58
		6	1.03	1.77	29.35
		7	1.01	1.75	31.10
		8	1.00	1.72	32.82
SWD/ SUA	SUA=504 plan codes	**1**	**8.56**	**14.76**	**14.76**
		2	1.74	3.01	17.77
		3	1.30	2.25	20.02
		4	1.13	1.95	21.97
		5	1.13	1.94	23.91
		6	1.08	1.87	25.78
		7	1.05	1.81	27.59
		8	1.04	1.79	29.38
		9	1.03	1.78	31.16
		10	1.03	1.77	32.93
ELL/ SUA	SUA & ELL codes	**1**	**10.24**	**17.65**	**17.65**
		2	1.94	3.34	20.99
		3	1.25	2.16	23.15
		4	1.12	1.94	25.09
		5	1.11	1.92	27.01
		6	1.07	1.85	28.86
		7	1.06	1.82	30.68
		8	1.02	1.76	32.44
		9	1.02	1.76	34.20

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	**1**	**9.50**	**16.38**	**16.38**
		2	1.67	2.88	19.26
		3	1.23	2.12	21.38
		4	1.13	1.95	23.33
		5	1.09	1.88	25.21
		6	1.06	1.83	27.04
		7	1.06	1.83	28.87
		8	1.05	1.81	30.68
		9	1.03	1.78	32.46
		10	1.03	1.77	34.23
SWD	All Codes	**1**	**9.02**	**15.54**	**15.54**
		2	1.60	2.77	18.31
		3	1.27	2.18	20.49
		4	1.10	1.89	22.38
		5	1.08	1.86	24.24
		6	1.06	1.82	26.06
		7	1.05	1.81	27.87
		8	1.04	1.79	29.66
		9	1.03	1.77	31.43
		10	1.02	1.75	33.18
SUA	All Codes	**1**	**10.20**	**17.59**	**17.59**
		2	1.67	2.87	20.46
		3	1.26	2.17	22.63
		4	1.08	1.86	24.49
		5	1.06	1.83	26.32
		6	1.03	1.77	28.09
		7	1.01	1.74	29.83
		8	1.01	1.74	31.57
SWD/ SUA	SUA=504 plan codes	**1**	**8.79**	**15.15**	**15.15**
		2	1.59	2.75	17.90
		3	1.26	2.18	20.08
		4	1.10	1.90	21.98
		5	1.08	1.87	23.85
		6	1.06	1.83	25.68
		7	1.05	1.82	27.50
		8	1.04	1.80	29.30
		9	1.03	1.77	31.07
		10	1.02	1.76	32.83

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup (cont.)

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/ SUA	SUA & ELL codes	**1**	**9.54**	**16.45**	**16.45**
		2	1.68	2.89	19.34
		3	1.23	2.13	21.47
		4	1.14	1.96	23.43
		5	1.09	1.87	25.30
		6	1.06	1.83	27.13
		7	1.06	1.83	28.96
		8	1.04	1.80	30.76
		9	1.04	1.79	32.55
		10	1.03	1.77	34.32

Appendix M: Classical Test Theory Statistics

These tables support the classical test theory analyses described in Section 5, "Operational Test Data Collection and Classical Analysis." They include item type, sample size, p-value, percent of omitted responses and the point-biserial of the key. External linking and field test items (i.e., those not contributing to students' scores) have been omitted.

Table M1. English Language Arts Grade 3 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	156,425	.78	0.04	.45
2	MC	156,370	.80	0.07	.33
3	MC	156,295	.49	0.12	.26
4	MC	156,320	.83	0.10	.40
5	MC	156,314	.70	0.11	.41
6	MC	156,221	.48	0.17	.32
7	MC	155,591	.47	0.57	.14
8	MC	155,607	.68	0.56	.50
9	MC	155,466	.66	0.65	.39
10	MC	155,245	.58	0.79	.31
11	MC	155,142	.59	0.86	.50
12	MC	154,956	.46	0.97	.32
13	MC	153,643	.54	1.81	.36
14	MC	153,240	.49	2.07	.44
15	MC	152,794	.58	2.36	.31
16	MC	152,344	.53	2.64	.32
17	MC	151,798	.60	2.99	.49
18	MC	149,038	.45	4.76	.30
19	MC	148,712	.53	4.96	.45
20	MC	148,019	.51	5.41	.36
21	MC	147,356	.42	5.83	.36
22	MC	146,796	.44	6.19	.39
23	MC	146,269	.53	6.53	.38
24	MC	145,890	.53	6.77	.40
25	MC	156,416	.56	0.04	.48
26	MC	156,394	.77	0.05	.44
27	MC	156,343	.55	0.09	.38
28	MC	156,361	.55	0.08	.27
29	MC	156,368	.76	0.07	.40
30	MC	156,326	.48	0.10	.35
31	MC	156,139	.53	0.22	.32

Table M1. English Language Arts Grade 3 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
32	CR	154,939	.48	0.98	
33	CR	153,904	.52	1.65	
34	CR	151,682	.42	3.07	
35	CR	148,280	.40	5.24	
36	CR	155,963	.56	0.33	
37	CR	152,047	.41	2.83	
38	CR	153,273	.45	2.05	
39	CR	152,151	.48	2.77	
40	CR	149,260	.43	4.61	
41	CR	147,434	.40	5.78	

Table M2. English Language Arts Grade 4 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	151,797	.73	0.05	.28
2	MC	151,782	.66	0.06	.44
3	MC	151,791	.71	0.06	.39
4	MC	151,777	.78	0.07	.30
5	MC	151,709	.33	0.11	.35
6	MC	151,763	.59	0.08	.33
7	MC	151,765	.74	0.08	.41
8	MC	151,759	.55	0.08	.47
9	MC	151,755	.55	0.08	.43
10	MC	151,764	.73	0.08	.35
11	MC	151,710	.45	0.11	.33
12	MC	151,707	.47	0.11	.35
13	MC	151,438	.50	0.29	.26
14	MC	151,398	.67	0.32	.47
15	MC	151,289	.60	0.39	.34
16	MC	151,217	.61	0.44	.40
17	MC	151,177	.47	0.46	.27
18	MC	151,093	.58	0.52	.33
19	MC	150,448	.48	0.94	.41
20	MC	150,262	.66	1.07	.38
21	MC	150,123	.57	1.16	.39
22	MC	149,852	.46	1.34	.46
23	MC	149,687	.81	1.44	.42
24	MC	149,402	.42	1.63	.33
25	MC	151,831	.59	0.03	.15
26	MC	151,821	.54	0.04	.33

Table M2. English Language Arts Grade 4 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
27	MC	151,778	.50	0.07	.33
28	MC	151,764	.49	0.08	.43
29	MC	151,795	.26	0.06	.12
30	MC	151,772	.59	0.07	.45
31	MC	151,641	.74	0.16	.46
32	CR	151,157	.68	0.48	
33	CR	150,026	.62	1.22	
34	CR	149,731	.59	1.41	
35	CR	147,760	.45	2.71	
36	CR	151,620	.57	0.17	
37	CR	150,841	.64	0.68	
38	CR	150,666	.67	0.80	
39	CR	148,155	.36	2.45	
40	CR	146,664	.59	3.43	
41	CR	144,870	.36	4.62	

Table M3. English Language Arts Grade 5 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	149,459	.80	0.01	.41
2	MC	149,403	.78	0.05	.36
3	MC	149,422	.65	0.04	.40
4	MC	149,410	.54	0.05	.37
5	MC	149,401	.53	0.05	.38
6	MC	149,421	.77	0.04	.54
7	MC	149,410	.86	0.05	.46
8	MC	149,414	.78	0.04	.45
9	MC	149,342	.58	0.09	.33
10	MC	149,360	.55	0.08	.24
11	MC	149,359	.49	0.08	.22
12	MC	149,330	.63	0.10	.41
13	MC	149,344	.63	0.09	.38
14	MC	149,321	.62	0.11	.34
15	MC	149,369	.62	0.07	.33
16	MC	149,130	.81	0.23	.44
17	MC	149,033	.46	0.30	.28
18	MC	148,930	.58	0.37	.39
19	MC	148,834	.74	0.43	.48
20	MC	148,887	.72	0.40	.40
21	MC	148,744	.28	0.49	.21

Table M3. English Language Arts Grade 5 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
22	MC	148,707	.46	0.52	.43
23	MC	148,115	.58	0.91	.26
24	MC	148,058	.73	0.95	.50
25	MC	147,878	.45	1.07	.46
26	MC	147,761	.58	1.15	.37
27	MC	147,644	.59	1.23	.39
28	MC	147,410	.51	1.38	.40
29	MC	147,244	.52	1.50	.40
30	MC	146,256	.43	2.16	.27
31	MC	146,079	.61	2.28	.39
32	MC	145,798	.42	2.46	.32
33	MC	145,617	.75	2.58	.49
34	MC	145,434	.48	2.71	.38
35	MC	145,293	.57	2.80	.46
36	MC	149,447	.76	0.02	.46
37	MC	149,452	.66	0.02	.32
38	MC	149,449	.62	0.02	.42
39	MC	149,413	.70	0.04	.37
40	MC	149,444	.60	0.02	.42
41	MC	149,433	.72	0.03	.40
42	MC	149,369	.75	0.07	.33
43	CR	149,172	.81	0.21	
44	CR	148,842	.64	0.43	
45	CR	148,612	.71	0.58	
46	CR	147,912	.49	1.05	
47	CR	149,321	.73	0.11	
48	CR	148,722	.65	0.51	
49	CR	148,652	.71	0.55	
50	CR	148,342	.71	0.76	
51	CR	147,811	.67	1.12	
52	CR	146,177	.46	2.21	

Table M4. English Language Arts Grade 6 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	148,688	.64	0.08	.53
2	MC	148,725	.86	0.05	.41
3	MC	148,734	.89	0.04	.37
4	MC	148,672	.52	0.09	.37
5	MC	148,684	.44	0.08	.34
6	MC	148,701	.72	0.07	.48
7	MC	148,628	.45	0.12	.19
8	MC	148,594	.63	0.14	.33
9	MC	148,606	.54	0.13	.41
10	MC	148,555	.71	0.16	.42
11	MC	148,481	.54	0.21	.35
12	MC	148,379	.58	0.28	.44
13	MC	148,509	.75	0.20	.43
14	MC	148,454	.70	0.23	.43
15	MC	148,215	.58	0.39	.36
16	MC	148,273	.47	0.35	.31
17	MC	148,215	.68	0.39	.44
18	MC	148,019	.44	0.52	.35
19	MC	147,925	.43	0.59	.31
20	MC	147,978	.46	0.55	.28
21	MC	147,939	.49	0.58	.37
22	MC	147,542	.65	0.85	.41
23	MC	147,495	.36	0.88	.34
24	MC	147,337	.53	0.98	.32
25	MC	147,189	.37	1.08	.23
26	MC	147,036	.58	1.19	.41
27	MC	146,969	.52	1.23	.28
28	MC	146,818	.60	1.33	.26
29	MC	145,899	.47	1.95	.33
30	MC	145,527	.43	2.20	.20
31	MC	145,224	.53	2.40	.40
32	MC	144,933	.44	2.60	.40
33	MC	144,742	.38	2.73	.15
34	MC	144,671	.55	2.77	.42
35	MC	144,604	.64	2.82	.46
36	MC	148,736	.62	0.04	.42
37	MC	148,755	.69	0.03	.42
38	MC	148,731	.55	0.05	.33
39	MC	148,732	.69	0.05	.38
40	MC	148,708	.56	0.06	.47

Table M4. English Language Arts Grade 6 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
41	MC	148,713	.69	0.06	.38
42	MC	148,689	.71	0.07	.37
43	CR	148,210	.77	0.40	
44	CR	147,411	.66	0.93	
45	CR	147,264	.67	1.03	
46	CR	145,992	.56	1.89	
47	CR	148,070	.72	0.49	
48	CR	147,967	.70	0.56	
49	CR	147,593	.68	0.81	
50	CR	145,988	.62	1.89	
51	CR	144,032	.56	3.20	
52	CR	144,246	.51	3.06	

Table M5. English Language Arts Grade 7 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	137,484	.73	0.13	.34
2	MC	137,603	.87	0.04	.37
3	MC	137,544	.63	0.08	.28
4	MC	137,479	.59	0.13	.41
5	MC	137,555	.67	0.08	.49
6	MC	137,431	.40	0.17	.39
7	MC	137,530	.79	0.09	.36
8	MC	137,446	.50	0.16	.29
9	MC	137,541	.71	0.09	.32
10	MC	137,498	.69	0.12	.28
11	MC	137,475	.60	0.13	.24
12	MC	137,463	.57	0.14	.21
13	MC	137,446	.45	0.16	.23
14	MC	137,506	.58	0.11	.29
15	MC	137,506	.41	0.11	.45
16	MC	137,235	.42	0.31	.27
17	MC	137,396	.59	0.19	.46
18	MC	137,242	.39	0.30	.38
19	MC	137,379	.56	0.20	.33
20	MC	137,330	.60	0.24	.48
21	MC	137,316	.67	0.25	.42
22	MC	137,130	.53	0.39	.31
23	MC	137,101	.51	0.41	.44
24	MC	137,096	.47	0.41	.31

Table M5. English Language Arts Grade 7 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
25	MC	136,862	.42	0.58	.28
26	MC	137,042	.70	0.45	.54
27	MC	136,870	.54	0.57	.36
28	MC	136,822	.48	0.61	.40
29	MC	134,928	.51	1.98	.40
30	MC	134,586	.32	2.23	.20
31	MC	134,370	.53	2.39	.46
32	MC	134,010	.44	2.65	.36
33	MC	133,770	.44	2.83	.37
34	MC	133,507	.52	3.02	.36
35	MC	133,432	.49	3.07	.26
36	MC	137,611	.73	0.04	.38
37	MC	137,607	.69	0.04	.45
38	MC	137,528	.58	0.10	.40
39	MC	137,568	.50	0.07	.40
40	MC	137,594	.62	0.05	.35
41	MC	137,542	.51	0.09	.27
42	MC	137,506	.46	0.11	.40
43	CR	136,925	.71	0.53	
44	CR	135,637	.69	1.47	
45	CR	132,941	.58	3.43	
46	CR	134,130	.55	2.56	
47	CR	136,917	.71	0.54	
48	CR	136,224	.72	1.04	
49	CR	136,712	.74	0.69	
50	CR	135,407	.74	1.64	
51	CR	132,559	.71	3.71	
52	CR	133,763	.56	2.83	

Table M6. English Language Arts Grade 8 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	133,120	.89	0.05	.25
2	MC	133,084	.57	0.07	.46
3	MC	133,121	.77	0.04	.46
4	MC	133,084	.57	0.07	.35
5	MC	133,028	.45	0.11	.44
6	MC	133,033	.60	0.11	.34
7	MC	133,046	.72	0.10	.47
8	MC	133,029	.57	0.11	.45
9	MC	133,035	.58	0.11	.25
10	MC	133,015	.56	0.12	.29
11	MC	133,010	.72	0.13	.31
12	MC	133,077	.45	0.08	.13
13	MC	132,999	.39	0.14	.34
14	MC	133,049	.77	0.10	.43
15	MC	132,974	.40	0.15	.28
16	MC	132,838	.61	0.26	.50
17	MC	132,894	.29	0.21	.22
18	MC	132,979	.59	0.15	.24
19	MC	132,934	.64	0.18	.41
20	MC	132,898	.45	0.21	.40
21	MC	132,851	.55	0.25	.29
22	MC	132,735	.57	0.33	.35
23	MC	132,764	.57	0.31	.31
24	MC	132,755	.55	0.32	.52
25	MC	132,632	.40	0.41	.32
26	MC	132,473	.45	0.53	.46
27	MC	132,619	.68	0.42	.50
28	MC	132,556	.59	0.47	.45
29	MC	132,289	.68	0.67	.48
30	MC	132,252	.64	0.70	.51
31	MC	132,135	.54	0.78	.27
32	MC	131,981	.48	0.90	.47
33	MC	131,790	.61	1.04	.50
34	MC	131,720	.69	1.10	.50
35	MC	131,650	.56	1.15	.41
36	MC	133,123	.65	0.04	.56
37	MC	133,118	.62	0.05	.32
38	MC	133,120	.72	0.05	.39
39	MC	133,119	.49	0.05	.23
40	MC	133,122	.73	0.04	.37

Table M6. English Language Arts Grade 8 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
41	MC	133,099	.69	0.06	.48
42	MC	133,070	.77	0.08	.47
43	CR	132,565	.76	0.46	
44	CR	131,435	.74	1.31	
45	CR	132,048	.80	0.85	
46	CR	130,473	.60	2.03	
47	CR	131,795	.69	1.04	
48	CR	131,058	.73	1.59	
49	CR	132,537	.77	0.48	
50	CR	131,025	.73	1.62	
51	CR	129,401	.73	2.84	
52	CR	128,948	.60	3.18	

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	155,148	.84	0.03	.48
2	MC	155,132	.92	0.04	.37
3	MC	154,968	.58	0.15	.50
4	MC	155,067	.92	0.09	.34
5	MC	155,066	.86	0.09	.39
6	MC	155,008	.81	0.12	.46
7	MC	154,995	.83	0.13	.38
8	MC	154,870	.77	0.21	.54
9	MC	155,002	.70	0.13	.51
10	MC	155,023	.53	0.11	.58
11	MC	154,839	.63	0.23	.45
12	MC	154,990	.73	0.14	.49
13	MC	154,935	.80	0.17	.53
14	MC	154,950	.84	0.16	.47
15	MC	154,853	.66	0.22	.44
16	MC	154,848	.43	0.23	.35
17	MC	154,752	.62	0.29	.47
18	MC	154,797	.76	0.26	.57
19	MC	154,440	.51	0.49	.46
20	MC	153,556	.73	1.06	.48
21	MC	155,123	.74	0.05	.41
22	MC	155,115	.60	0.05	.40
23	MC	155,031	.70	0.11	.46
24	MC	154,935	.41	0.17	.40

Table M7. Mathematics Grade 3 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
25	MC	155,048	.74	0.10	.35
26	MC	154,892	.33	0.20	.35
27	MC	155,016	.71	0.12	.59
28	MC	154,892	.48	0.20	.52
29	MC	155,007	.64	0.12	.45
30	MC	154,993	.47	0.13	.49
31	MC	154,987	.65	0.14	.45
32	MC	154,888	.77	0.20	.44
33	MC	154,895	.67	0.20	.48
34	MC	154,851	.83	0.22	.40
35	MC	154,777	.82	0.27	.51
36	MC	154,851	.82	0.22	.39
37	MC	154,870	.63	0.21	.62
38	MC	154,721	.60	0.31	.51
39	MC	154,667	.88	0.34	.47
40	MC	154,403	.78	0.51	.56
41	MC	154,040	.60	0.75	.53
42	CR	154,993	.46	0.13	
43	CR	154,915	.59	0.18	
44	CR	154,854	.50	0.22	
45	CR	154,863	.56	0.22	
46	CR	154,871	.44	0.21	
47	CR	154,844	.49	0.23	
48	CR	154,911	.64	0.19	
49	CR	154,606	.47	0.38	

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	148,773	.89	0.02	.38
2	MC	148,747	.75	0.04	.48
3	MC	148,705	.76	0.06	.59
4	MC	148,696	.75	0.07	.49
5	MC	148,697	.88	0.07	.36
6	MC	148,658	.65	0.10	.46
7	MC	148,651	.66	0.10	.51
8	MC	148,591	.73	0.14	.59
9	MC	148,682	.68	0.08	.54
10	MC	148,603	.52	0.13	.49
11	MC	148,556	.57	0.16	.39

Table M8. Mathematics Grade 4 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
12	MC	148,625	.44	0.12	.41
13	MC	148,469	.74	0.22	.52
14	MC	148,608	.50	0.13	.43
15	MC	148,483	.66	0.21	.56
16	MC	148,556	.79	0.16	.46
17	MC	148,465	.63	0.23	.52
18	MC	148,337	.74	0.31	.43
19	MC	148,386	.85	0.28	.47
20	MC	147,858	.69	0.63	.48
21	MC	148,754	.59	0.03	.50
22	MC	148,687	.52	0.08	.58
23	MC	148,721	.85	0.05	.49
24	MC	148,700	.68	0.07	.50
25	MC	148,644	.59	0.10	.61
26	MC	148,448	.65	0.24	.56
27	MC	148,518	.68	0.19	.57
28	MC	148,606	.45	0.13	.33
29	MC	148,670	.44	0.09	.47
30	MC	148,654	.65	0.10	.51
31	MC	148,628	.45	0.12	.55
32	MC	148,650	.65	0.10	.64
33	MC	148,593	.51	0.14	.49
34	MC	148,618	.71	0.12	.46
35	MC	148,371	.67	0.29	.48
36	MC	148,437	.58	0.24	.61
37	MC	148,446	.75	0.24	.42
38	MC	148,290	.62	0.34	.49
39	MC	148,390	.75	0.28	.58
40	MC	148,271	.56	0.36	.62
41	MC	148,216	.62	0.39	.49
42	MC	147,954	.65	0.57	.44
43	CR	148,648	.37	0.10	
44	CR	148,391	.57	0.27	
45	CR	148,522	.38	0.19	
46	CR	148,502	.74	0.20	
47	CR	148,640	.80	0.11	
48	CR	148,523	.37	0.19	
49	CR	148,332	.36	0.31	
50	CR	148,460	.52	0.23	
51	CR	148,384	.52	0.28	
52	CR	148,437	.49	0.24	

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	143,903	.91	0.01	.33
2	MC	143,836	.73	0.06	.45
3	MC	143,742	.84	0.12	.46
4	MC	143,842	.78	0.05	.42
5	MC	143,834	.69	0.06	.48
6	MC	143,799	.87	0.08	.48
7	MC	143,867	.88	0.04	.35
8	MC	143,557	.35	0.25	.37
9	MC	143,826	.79	0.07	.48
10	MC	143,517	.30	0.28	.41
11	MC	143,740	.85	0.13	.51
12	MC	143,752	.49	0.12	.59
13	MC	143,793	.90	0.09	.29
14	MC	143,723	.52	0.14	.56
15	MC	143,681	.57	0.17	.49
16	MC	143,749	.75	0.12	.46
17	MC	143,705	.54	0.15	.44
18	MC	143,719	.81	0.14	.48
19	MC	143,291	.45	0.44	.41
20	MC	143,592	.60	0.23	.43
21	MC	143,875	.63	0.03	.36
22	MC	143,884	.78	0.03	.41
23	MC	143,870	.86	0.03	.48
24	MC	143,831	.78	0.06	.51
25	MC	143,805	.54	0.08	.53
26	MC	143,827	.85	0.06	.40
27	MC	143,856	.84	0.04	.36
28	MC	143,784	.60	0.09	.40
29	MC	143,835	.91	0.06	.36
30	MC	143,785	.52	0.09	.34
31	MC	143,824	.64	0.07	.39
32	MC	143,667	.48	0.18	.44
33	MC	143,829	.68	0.06	.52
34	MC	143,790	.52	0.09	.56
35	MC	143,707	.75	0.15	.52
36	MC	143,806	.92	0.08	.36
37	MC	143,649	.42	0.19	.42
38	MC	143,723	.44	0.14	.59
39	MC	143,696	.53	0.16	.50
40	MC	143,609	.38	0.19	.34

Table M9. Mathematics Grade 5 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
41	MC	143,640	.72	0.19	.52
42	MC	143,443	.75	0.33	.48
43	CR	143,697	.48	0.15	
44	CR	143,642	.27	0.19	
45	CR	143,605	.51	0.22	
46	CR	143,363	.46	0.39	
47	CR	143,437	.34	0.34	
48	CR	143,648	.45	0.19	
49	CR	143,569	.26	0.24	
50	CR	143,375	.28	0.38	
51	CR	143,582	.62	0.23	
52	CR	143,458	.48	0.32	

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	141,796	.87	0.06	.46
2	MC	141,828	.77	0.04	.48
3	MC	141,672	.58	0.15	.59
4	MC	141,814	.76	0.05	.39
5	MC	141,738	.72	0.10	.45
6	MC	141,772	.58	0.08	.54
7	MC	141,740	.63	0.10	.45
8	MC	141,780	.58	0.07	.33
9	MC	141,744	.53	0.10	.43
10	MC	141,753	.54	0.09	.33
11	MC	141,784	.53	0.07	.49
12	MC	141,753	.77	0.09	.50
13	MC	141,607	.56	0.19	.47
14	MC	141,764	.73	0.08	.54
15	MC	141,565	.50	0.22	.52
16	MC	141,745	.40	0.10	.46
17	MC	141,636	.60	0.17	.46
18	MC	141,684	.40	0.14	.35
19	MC	141,669	.62	0.15	.53
20	MC	141,521	.38	0.25	.52
21	MC	141,617	.65	0.19	.49
22	MC	141,559	.64	0.23	.52
23	MC	141,516	.48	0.26	.48
24	MC	141,082	.59	0.56	.58

Table M10. Mathematics Grade 6 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
25	MC	141,740	.35	0.10	.46
26	MC	141,832	.46	0.03	.51
27	MC	141,754	.56	0.09	.45
28	MC	141,822	.75	0.04	.52
29	MC	141,627	.59	0.18	.42
30	MC	141,759	.65	0.09	.49
31	MC	141,686	.78	0.14	.48
32	MC	141,808	.80	0.05	.52
33	MC	141,696	.56	0.13	.43
34	MC	141,635	.27	0.17	.45
35	MC	141,806	.77	0.05	.51
36	MC	141,657	.52	0.16	.32
37	MC	141,678	.51	0.14	.54
38	MC	141,758	.27	0.09	.29
39	MC	141,762	.76	0.08	.51
40	MC	141,587	.60	0.21	.55
41	MC	141,774	.89	0.07	.32
42	MC	141,591	.54	0.20	.52
43	MC	141,724	.86	0.11	.43
44	MC	141,746	.46	0.09	.45
45	MC	141,504	.52	0.27	.60
46	MC	141,619	.70	0.18	.32
47	MC	141,623	.55	0.18	.50
48	MC	141,128	.47	0.53	.41
49	CR	141,570	.60	0.22	
50	CR	141,211	.46	0.47	
51	CR	141,246	.71	0.45	
52	CR	141,541	.46	0.24	
53	CR	140,552	.49	0.94	
54	CR	140,840	.32	0.73	
55	CR	140,262	.42	1.14	
56	CR	140,154	.40	1.22	
57	CR	140,894	.37	0.69	
58	CR	141,106	.43	0.55	

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	126,211	.84	0.04	.31
2	MC	126,086	.84	0.14	.43
3	MC	126,040	.70	0.17	.52
4	MC	126,225	.96	0.03	.24
5	MC	126,179	.73	0.06	.54
6	MC	126,137	.59	0.10	.55
7	MC	126,029	.41	0.18	.33
8	MC	126,058	.46	0.16	.46
9	MC	126,161	.73	0.08	.45
10	MC	125,886	.39	0.30	.35
11	MC	125,980	.60	0.22	.53
12	MC	126,113	.19	0.12	.42
13	MC	125,964	.44	0.23	.52
14	MC	126,044	.48	0.17	.43
15	MC	125,902	.31	0.28	.33
16	MC	126,102	.81	0.13	.48
17	MC	126,028	.68	0.18	.42
18	MC	125,860	.38	0.32	.43
19	MC	125,842	.38	0.33	.46
20	MC	125,974	.56	0.23	.62
21	MC	125,647	.32	0.49	.31
22	MC	125,759	.39	0.40	.43
23	MC	125,790	.53	0.37	.34
24	MC	125,681	.54	0.46	.19
25	MC	126,079	.69	0.14	.46
26	MC	126,168	.44	0.07	.39
27	MC	126,183	.73	0.06	.36
28	MC	125,669	.40	0.47	.32
29	MC	126,006	.37	0.20	.33
30	MC	126,017	.53	0.19	.55
31	MC	125,961	.49	0.24	.51
32	MC	125,914	.40	0.27	.53
33	MC	125,962	.40	0.24	.35
34	MC	126,110	.54	0.12	.45
35	MC	125,969	.49	0.23	.33
36	MC	126,059	.55	0.16	.43
37	MC	125,957	.61	0.24	.55
38	MC	125,986	.65	0.22	.42
39	MC	126,057	.67	0.16	.48
40	MC	126,022	.72	0.19	.48

Table M11. Mathematics Grade 7 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
41	MC	126,049	.61	0.17	.48
42	MC	125,919	.55	0.27	.54
43	MC	125,864	.52	0.31	.44
44	MC	125,784	.49	0.38	.47
45	MC	125,676	.52	0.46	.51
46	MC	125,782	.72	0.38	.53
47	MC	125,542	.32	0.57	.47
48	MC	125,409	.38	0.67	.50
49	CR	125,976	.62	0.22	
50	CR	125,750	.51	0.40	
51	CR	125,580	.65	0.54	
52	CR	125,633	.55	0.50	
53	CR	125,593	.61	0.53	
54	CR	125,336	.54	0.73	
55	CR	125,107	.40	0.91	
56	CR	125,477	.46	0.62	
57	CR	124,333	.33	1.53	
58	CR	124,086	.34	1.72	

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	P-value	% Omit	Pbis Key
1	MC	95,254	.91	0.03	.24
2	MC	95,185	.74	0.10	.41
3	MC	95,240	.82	0.04	.38
4	MC	95,223	.61	0.06	.40
5	MC	95,192	.81	0.09	.43
6	MC	95,105	.51	0.18	.41
7	MC	95,111	.47	0.18	.48
8	MC	95,240	.34	0.04	.45
9	MC	95,204	.54	0.08	.48
10	MC	95,168	.56	0.12	.46
11	MC	95,156	.61	0.13	.51
12	MC	95,142	.33	0.14	.29
13	MC	95,128	.57	0.16	.45
14	MC	95,104	.48	0.18	.33
15	MC	95,204	.77	0.08	.52
16	MC	95,184	.68	0.10	.43
17	MC	95,177	.57	0.11	.53
18	MC	95,174	.38	0.11	.43

Table M12. Mathematics Grade 8 Classical Item Analysis (cont.)

Item	Type	N-Count	P-value	% Omit	Pbis Key
19	MC	95,143	.29	0.14	.23
20	MC	95,126	.49	0.16	.43
21	MC	95,176	.59	0.11	.27
22	MC	95,007	.42	0.29	.34
23	MC	95,158	.62	0.13	.46
24	MC	95,066	.28	0.22	.37
25	MC	95,240	.66	0.04	.54
26	MC	95,187	.70	0.10	.40
27	MC	95,126	.47	0.16	.38
28	MC	95,083	.44	0.21	.41
29	MC	95,157	.48	0.13	.49
30	MC	95,149	.62	0.14	.45
31	MC	95,116	.42	0.17	.43
32	MC	95,200	.66	0.08	.40
33	MC	95,097	.43	0.19	.47
34	MC	95,208	.59	0.08	.50
35	MC	95,187	.59	0.10	.45
36	MC	95,210	.69	0.07	.37
37	MC	95,209	.64	0.07	.38
38	MC	94,672	.45	0.08	.38
39	MC	95,146	.40	0.14	.35
40	MC	95,065	.41	0.23	.42
41	MC	95,179	.75	0.11	.45
42	MC	95,078	.41	0.21	.43
43	MC	95,163	.72	0.12	.51
44	MC	95,137	.34	0.15	.41
45	MC	95,093	.32	0.20	.28
46	MC	95,070	.40	0.22	.49
47	MC	94,991	.49	0.30	.45
48	MC	95,004	.66	0.29	.49
49	CR	93,997	.52	1.35	
50	CR	92,357	.44	3.07	
51	CR	92,808	.28	2.59	
52	CR	94,521	.40	0.80	
53	CR	92,657	.42	2.75	
54	CR	93,825	.36	1.53	
55	CR	93,783	.29	1.57	
56	CR	93,968	.45	1.38	
57	CR	90,321	.51	5.20	
58	CR	90,978	.33	4.52	

Appendix N: Items Flagged for DIF

These tables support the DIF information in Section 5, "Operational Test Data Collection and Classical Analysis." They include item numbers, focal group, and directions of DIF and DIF statistics. Tables N1–N3 show items flagged by the SMD, or Mantel-Haenszel methods. No mathematics constructed-response items were flagged for DIF, so that table has been omitted. Note that positive values of SMD and Delta in Tables N1–N3 indicate DIF in favor of a focal group, and negative values of SMD and Delta indicate DIF against a focal group. External linking and field test items (i.e., those not contributing to students' scores) have been omitted.

Table N1. ELA MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
4	21	Female	Against	1.57	1,541.8	-1.06
4	30	Black	Against	1.57	799.0	-1.06
4	30	Hispanic	Against	1.64	1,150.0	-1.16
5	16	Female	Against	1.83	1,583.3	-1.43
7	2	Female	Against	1.72	908.9	-1.28
7	26	Hispanic	Against	1.54	604.4	-1.02
7	32	Female	Against	1.55	1,295.5	-1.03
7	36	Hispanic	Against	1.58	784.2	-1.07
8	2	Female	Against	1.62	1,407.2	-1.13

Table N2. ELA CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect Size
3	40	Black	In Favor	0.1096	0.176
3	40	Hispanic	In Favor	0.1077	0.175
5	47	Black	In Favor	0.1154	0.202
5	47	Hispanic	In Favor	0.0977	0.172
5	48	High Needs	In Favor	0.1244	0.193
5	51	High Needs	In Favor	0.1170	0.182
7	43	Black	In Favor	0.1147	0.189
7	43	Hispanic	In Favor	0.1194	0.198
7	43	High Needs	In Favor	0.1328	0.220
7	44	Black	In Favor	0.1148	0.173
7	44	Hispanic	In Favor	0.1231	0.186
7	44	High Needs	In Favor	0.1317	0.199
7	46	Female	In Favor	0.2271	0.226
7	50	Black	In Favor	0.1187	0.189
7	50	Hispanic	In Favor	0.1126	0.180
7	50	High Needs	In Favor	0.1117	0.179
8	44	High Needs	In Favor	0.1048	0.176

Table N3. Mathematics MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	9	Black	Against	1.68	864.1	-1.22
3	9	Hispanic	Against	1.55	751.6	-1.04
3	39	Black	In Favor	0.64	300.4	1.06
4	2	Black	Against	1.56	554.9	-1.04
4	5	Hispanic	Against	1.60	460.8	-1.11
4	7	Female	Against	1.54	1,133.4	-1.01
4	42	Black	Against	1.58	742.4	-1.07
5	38	Female	Against	1.56	1,031.6	-1.04
6	10	Female	Against	1.58	1,591.3	-1.07
6	16	Female	Against	1.84	2,406.6	-1.44
7	1	Female	Against	1.56	708.9	-1.04
7	4	Female	In Favor	0.57	321.0	1.33
7	5	Female	Against	1.75	1,235.8	-1.31
8	25	Female	Against	1.69	964.2	-1.24
8	48	Female	Against	1.54	716.6	-1.01

Appendix O: Item Response Theory Statistics

External linking and field test items (i.e., those not contributing to students' scores) have been omitted.

Table O1. English Language Arts Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	448.32	7	117.95	186.62	Y
2	3PL	215.27	7	55.66	186.56	Y
3	3PL	56.50	7	13.23	186.45	Y
4	3PL	221.01	7	57.20	186.46	Y
5	3PL	200.04	7	51.59	186.46	Y
6	3PL	139.55	7	35.42	186.34	Y
7	3PL	93.54	7	23.13	185.63	Y
8	3PL	276.00	7	71.89	185.63	Y
9	3PL	240.04	7	62.28	185.43	Y
10	3PL	87.51	7	21.52	185.18	Y
11	3PL	271.12	7	70.59	185.04	Y
12	3PL	82.09	7	20.07	184.83	Y
13	3PL	164.19	7	42.01	183.34	Y
14	3PL	293.34	7	76.53	182.89	Y
15	3PL	77.04	7	18.72	182.39	Y
16	3PL	267.41	7	69.60	181.74	Y
17	3PL	216.45	7	55.98	181.16	Y
18	3PL	96.66	7	23.96	177.79	Y
19	3PL	236.11	7	61.23	177.34	Y
20	3PL	131.99	7	33.41	176.51	Y
21	3PL	188.39	7	48.48	175.65	Y
22	3PL	282.38	7	73.60	175.05	Y
23	3PL	104.60	7	26.08	174.38	Y
24	3PL	182.32	7	46.86	173.88	Y
25	3PL	208.21	7	53.78	186.59	Y
26	3PL	118.76	7	29.87	186.55	Y
27	3PL	106.67	7	26.64	186.47	Y
28	3PL	53.40	7	12.40	186.51	Y
29	3PL	107.15	7	26.77	186.51	Y
30	3PL	174.45	7	44.75	186.48	Y
31	3PL	59.24	7	13.96	186.24	Y
32	2PPC	604.12	16	103.97	184.79	Y
33	2PPC	650.08	16	112.09	183.54	Y
34	2PPC	562.83	16	96.67	180.87	Y
35	2PPC	1156.79	34	136.16	176.97	Y

Table O1. English Language Arts Grade 3 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
36	2PPC	576.43	16	99.07	186.09	Y
37	2PPC	628.89	16	108.34	181.29	Y
38	2PPC	372.91	16	63.09	182.83	Y
39	2PPC	403.02	16	68.42	181.52	Y
40	2PPC	403.11	16	68.43	178.06	Y
41	2PPC	1527.23	34	181.08	175.84	N

Table O2. English Language Arts Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	109.35	7	27.35	186.57	Y
2	3PL	121.32	7	30.55	186.53	Y
3	3PL	161.73	7	41.35	186.55	Y
4	3PL	212.88	7	55.02	186.52	Y
5	3PL	309.83	7	80.94	186.43	Y
6	3PL	246.62	7	64.04	186.50	Y
7	3PL	202.56	7	52.26	186.50	Y
8	3PL	185.49	7	47.70	186.51	Y
9	3PL	255.63	7	66.45	186.49	Y
10	3PL	83.36	7	20.41	186.53	Y
11	3PL	91.98	7	22.71	186.45	Y
12	3PL	143.85	7	36.57	186.47	Y
13	3PL	116.01	7	29.13	186.17	Y
14	3PL	154.18	7	39.33	186.10	Y
15	3PL	99.08	7	24.61	185.97	Y
16	3PL	172.08	7	44.12	185.87	Y
17	3PL	45.32	7	10.24	185.84	Y
18	3PL	83.33	7	20.40	185.72	Y
19	3PL	139.94	7	35.53	184.94	Y
20	3PL	126.94	7	32.06	184.71	Y
21	3PL	92.03	7	22.73	184.60	Y
22	3PL	305.55	7	79.79	184.21	Y
23	3PL	130.98	7	33.14	184.06	Y
24	3PL	174.15	7	44.67	183.67	Y
25	3PL	76.55	7	18.59	186.60	Y
26	3PL	93.25	7	23.05	186.58	Y
27	3PL	97.37	7	24.15	186.54	Y
28	3PL	187.87	7	48.34	186.51	Y
29	3PL	29.54	7	6.02	186.57	Y
30	3PL	169.70	7	43.48	186.53	Y

Table O2. English Language Arts Grade 4 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
31	3PL	153.48	7	39.15	186.37	Y
32	2PPC	544.60	16	93.45	185.79	Y
33	2PPC	387.01	16	65.59	184.39	Y
34	2PPC	817.04	16	141.60	184.06	Y
35	2PPC	892.72	34	104.13	181.58	Y
36	2PPC	966.52	16	168.03	186.36	Y
37	2PPC	468.35	16	79.97	185.39	Y
38	2PPC	719.29	16	124.32	185.16	Y
39	2PPC	445.91	16	76.00	182.13	Y
40	2PPC	649.23	16	111.94	180.27	Y
41	2PPC	1299.05	34	153.41	178.00	Y

Table O3. English Language Arts Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	67.88	7	16.27	186.64	Y
2	3PL	55.57	7	12.98	186.56	Y
3	3PL	82.74	7	20.24	186.60	Y
4	3PL	91.80	7	22.66	186.58	Y
5	3PL	135.52	7	34.35	186.57	Y
6	3PL	122.55	7	30.88	186.58	Y
7	3PL	100.80	7	25.07	186.57	Y
8	3PL	296.71	7	77.43	186.59	Y
9	3PL	97.54	7	24.20	186.51	Y
10	3PL	45.16	7	10.20	186.52	Y
11	3PL	61.19	7	14.48	186.53	Y
12	3PL	92.13	7	22.75	186.48	Y
13	3PL	72.62	7	17.54	186.49	Y
14	3PL	70.07	7	16.86	186.45	Y
15	3PL	56.51	7	13.23	186.55	Y
16	3PL	79.47	7	19.37	186.25	Y
17	3PL	107.57	7	26.88	186.12	Y
18	3PL	86.14	7	21.15	185.97	Y
19	3PL	223.41	7	57.84	185.85	Y
20	3PL	190.22	7	48.97	185.89	Y
21	3PL	194.94	7	50.23	185.74	Y
22	3PL	208.17	7	53.76	185.69	Y
23	3PL	41.83	7	9.31	185.00	Y
24	3PL	100.23	7	24.92	184.91	Y
25	3PL	332.62	7	87.02	184.70	Y

Table O3. English Language Arts Grade 5 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
26	3PL	167.82	7	42.98	184.55	Y
27	3PL	89.74	7	22.11	184.44	Y
28	3PL	140.79	7	35.76	184.17	Y
29	3PL	133.68	7	33.86	183.96	Y
30	3PL	110.78	7	27.74	182.73	Y
31	3PL	111.68	7	27.98	182.54	Y
32	3PL	170.16	7	43.61	182.22	Y
33	3PL	112.73	7	28.26	182.00	Y
34	3PL	216.53	7	56.00	181.81	Y
35	3PL	141.88	7	36.05	181.64	Y
36	3PL	112.75	7	28.26	186.63	Y
37	3PL	118.31	7	29.75	186.63	Y
38	3PL	181.34	7	46.59	186.63	Y
39	3PL	71.79	7	17.31	186.57	Y
40	3PL	140.96	7	35.80	186.63	Y
41	3PL	139.91	7	35.52	186.62	Y
42	3PL	132.38	7	33.51	186.52	Y
43	2PPC	241.09	16	39.79	186.25	Y
44	2PPC	573.14	16	98.49	185.84	Y
45	2PPC	297.89	16	49.83	185.54	Y
46	2PPC	642.51	34	73.79	184.67	Y
47	2PPC	467.38	16	79.79	186.45	Y
48	2PPC	322.99	16	54.27	185.67	Y
49	2PPC	400.15	16	67.91	185.65	Y
50	2PPC	309.61	16	51.90	185.30	Y
51	2PPC	426.74	16	72.61	184.50	Y
52	2PPC	802.25	34	93.16	182.42	Y

Table O4. English Language Arts Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	155.16	7	39.60	186.52	Y
2	3PL	91.10	7	22.48	186.57	Y
3	3PL	59.87	7	14.13	186.59	Y
4	3PL	143.23	7	36.41	186.49	Y
5	3PL	78.23	7	19.04	186.51	Y
6	3PL	92.11	7	22.75	186.53	Y
7	3PL	102.97	7	25.65	186.43	Y
8	3PL	76.49	7	18.57	186.42	Y
9	3PL	98.51	7	24.46	186.42	Y
10	3PL	74.71	7	18.10	186.36	Y
11	3PL	184.16	7	47.35	186.23	Y
12	3PL	103.24	7	25.72	186.09	Y
13	3PL	108.96	7	27.25	186.27	Y
14	3PL	153.59	7	39.18	186.23	Y
15	3PL	85.85	7	21.07	185.94	Y
16	3PL	70.66	7	17.01	186.00	Y
17	3PL	119.12	7	29.97	185.94	Y
18	3PL	161.06	7	41.17	185.70	Y
19	3PL	88.67	7	21.83	185.57	Y
20	3PL	146.38	7	37.25	185.67	Y
21	3PL	80.92	7	19.76	185.59	Y
22	3PL	104.15	7	25.96	185.05	Y
23	3PL	103.85	7	25.88	185.03	Y
24	3PL	86.06	7	21.13	184.81	Y
25	3PL	100.41	7	24.97	184.64	Y
26	3PL	281.89	7	73.47	184.44	Y
27	3PL	225.87	7	58.50	184.37	Y
28	3PL	404.31	7	106.19	184.18	Y
29	3PL	99.85	7	24.82	183.07	Y
30	3PL	26.75	7	5.28	182.57	Y
31	3PL	101.18	7	25.17	182.16	Y
32	3PL	123.54	7	31.15	181.79	Y
33	3PL	49.94	7	11.48	181.58	Y
34	3PL	289.79	7	75.58	181.49	Y
35	3PL	108.95	7	27.25	181.38	Y
36	3PL	102.78	7	25.60	186.59	Y
37	3PL	97.17	7	24.10	186.61	Y
38	3PL	57.18	7	13.41	186.57	Y
39	3PL	77.42	7	18.82	186.58	Y
40	3PL	136.43	7	34.59	186.54	Y

Table O4. English Language Arts Grade 6 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
41	3PL	126.93	7	32.05	186.54	Y
42	3PL	52.02	7	12.03	186.50	Y
43	2PPC	360.50	16	60.90	185.91	Y
44	2PPC	373.19	16	63.14	184.90	Y
45	2PPC	369.03	16	62.41	184.74	Y
46	2PPC	877.08	34	102.24	183.21	Y
47	2PPC	537.02	16	92.11	185.76	Y
48	2PPC	259.39	16	43.03	185.61	Y
49	2PPC	372.71	16	63.06	185.12	Y
50	2PPC	658.19	16	113.53	183.07	Y
51	2PPC	471.98	16	80.61	180.64	Y
52	2PPC	1183.60	34	139.41	181.02	Y

Table O5. English Language Arts Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	56.16	7	13.14	186.43	Y
2	3PL	94.12	7	23.29	186.59	Y
3	3PL	39.23	7	8.61	186.50	Y
4	3PL	128.05	7	32.35	186.41	Y
5	3PL	114.67	7	28.78	186.53	Y
6	3PL	144.19	7	36.66	186.36	Y
7	3PL	99.20	7	24.64	186.49	Y
8	3PL	39.53	7	8.70	186.37	Y
9	3PL	207.56	7	53.60	186.49	Y
10	3PL	140.99	7	35.81	186.44	Y
11	3PL	178.35	7	45.80	186.41	Y
12	3PL	94.67	7	23.43	186.39	Y
13	3PL	84.88	7	20.81	186.35	Y
14	3PL	51.88	7	12.00	186.45	Y
15	3PL	136.16	7	34.52	186.45	Y
16	3PL	79.50	7	19.38	186.05	Y
17	3PL	139.89	7	35.52	186.28	Y
18	3PL	214.87	7	55.55	186.04	Y
19	3PL	88.48	7	21.78	186.28	Y
20	3PL	136.17	7	34.52	186.18	Y
21	3PL	98.65	7	24.50	186.20	Y
22	3PL	166.34	7	42.59	185.90	Y
23	3PL	123.10	7	31.03	185.88	Y
24	3PL	46.13	7	10.46	185.85	Y

Table O5. English Language Arts Grade 7 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
25	3PL	63.09	7	14.99	185.57	Y
26	3PL	156.92	7	40.07	185.81	Y
27	3PL	310.27	7	81.05	185.54	Y
28	3PL	140.70	7	35.73	185.46	Y
29	3PL	119.32	7	30.02	182.95	Y
30	3PL	51.18	7	11.81	182.43	Y
31	3PL	169.40	7	43.40	182.18	Y
32	3PL	81.02	7	19.78	181.71	Y
33	3PL	167.20	7	42.82	181.37	Y
34	3PL	205.84	7	53.14	181.02	Y
35	3PL	62.71	7	14.89	180.90	Y
36	3PL	108.14	7	27.03	186.58	Y
37	3PL	91.18	7	22.50	186.59	Y
38	3PL	122.88	7	30.97	186.47	Y
39	3PL	136.18	7	34.53	186.56	Y
40	3PL	109.81	7	27.48	186.56	Y
41	3PL	52.39	7	12.13	186.50	Y
42	3PL	97.60	7	24.22	186.42	Y
43	2PPC	282.64	16	47.14	185.66	Y
44	2PPC	510.73	16	87.46	183.87	Y
45	2PPC	671.42	16	115.86	180.38	Y
46	2PPC	968.99	34	113.38	181.93	Y
47	2PPC	378.35	16	64.06	185.66	Y
48	2PPC	688.97	16	118.97	184.80	Y
49	2PPC	455.13	16	77.63	185.35	Y
50	2PPC	286.92	16	47.89	183.60	Y
51	2PPC	948.80	16	164.90	179.71	Y
52	2PPC	1186.64	34	139.78	181.46	Y

Table O6. English Language Arts Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	622.26	7	164.44	186.56	Y
2	3PL	139.42	7	35.39	186.54	Y
3	3PL	109.35	7	27.36	186.57	Y
4	3PL	62.79	7	14.91	186.53	Y
5	3PL	188.94	7	48.63	186.46	Y
6	3PL	73.29	7	17.72	186.45	Y
7	3PL	93.35	7	23.08	186.47	Y
8	3PL	98.90	7	24.56	186.46	Y
9	3PL	44.45	7	10.01	186.46	Y
10	3PL	50.78	7	11.70	186.43	Y
11	3PL	813.91	7	215.65	186.44	N
12	3PL	271.54	7	70.70	186.52	Y
13	3PL	125.41	7	31.65	186.42	Y
14	3PL	156.69	7	40.01	186.49	Y
15	3PL	201.07	7	51.87	186.39	Y
16	3PL	151.82	7	38.70	186.19	Y
17	3PL	169.24	7	43.36	186.29	Y
18	3PL	146.49	7	37.28	186.38	Y
19	3PL	125.14	7	31.57	186.31	Y
20	3PL	121.52	7	30.61	186.25	Y
21	3PL	49.05	7	11.24	186.19	Y
22	3PL	73.47	7	17.76	186.02	Y
23	3PL	75.31	7	18.26	186.09	Y
24	3PL	174.26	7	44.70	186.05	Y
25	3PL	146.38	7	37.25	185.91	Y
26	3PL	214.70	7	55.51	185.65	Y
27	3PL	103.81	7	25.87	185.86	Y
28	3PL	106.92	7	26.71	185.78	Y
29	3PL	114.10	7	28.62	185.41	Y
30	3PL	117.05	7	29.41	185.34	Y
31	3PL	53.75	7	12.49	185.21	Y
32	3PL	238.52	7	61.88	185.01	Y
33	3PL	185.11	7	47.60	184.69	Y
34	3PL	99.48	7	24.72	184.62	Y
35	3PL	556.05	7	146.74	184.52	Y
36	3PL	176.80	7	45.38	186.57	Y
37	3PL	55.80	7	13.04	186.59	Y
38	3PL	88.21	7	21.71	186.59	Y
39	3PL	69.36	7	16.67	186.58	Y
40	3PL	64.48	7	15.36	186.59	Y

Table O6. English Language Arts Grade 8 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
41	3PL	85.37	7	20.95	186.54	Y
42	3PL	77.23	7	18.77	186.51	Y
43	2PPC	339.81	16	57.24	185.77	Y
44	2PPC	464.87	16	79.35	184.29	Y
45	2PPC	280.51	16	46.76	185.07	Y
46	2PPC	748.21	34	86.61	182.88	Y
47	2PPC	446.98	16	76.19	184.75	Y
48	2PPC	480.90	16	82.18	183.81	Y
49	2PPC	417.02	16	70.89	185.79	Y
50	2PPC	431.61	16	73.47	183.63	Y
51	2PPC	349.84	16	59.02	181.51	Y
52	2PPC	1456.27	34	172.48	180.65	Y

Table O7. Mathematics Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	261.25	7	67.95	186.59	Y
2	3PL	139.34	7	35.37	186.58	Y
3	3PL	111.27	7	27.87	186.40	Y
4	3PL	63.04	7	14.98	186.52	Y
5	3PL	40.62	7	8.98	186.51	Y
6	3PL	197.13	7	50.82	186.44	Y
7	3PL	43.43	7	9.74	186.43	Y
8	3PL	226.33	7	58.62	186.25	Y
9	3PL	88.27	7	21.72	186.43	Y
10	3PL	152.80	7	38.97	186.43	Y
11	3PL	86.42	7	21.23	186.26	Y
12	3PL	76.19	7	18.49	186.45	Y
13	3PL	192.78	7	49.65	186.35	Y
14	3PL	100.06	7	24.87	186.38	Y
15	3PL	106.01	7	26.46	186.27	Y
16	3PL	67.39	7	16.14	186.25	Y
17	3PL	125.17	7	31.58	186.11	Y
18	3PL	114.71	7	28.79	186.21	Y
19	3PL	124.30	7	31.35	185.81	Y
20	3PL	72.50	7	17.51	184.74	Y
21	3PL	78.58	7	19.13	186.58	Y
22	3PL	149.01	7	37.95	186.56	Y
23	3PL	97.60	7	24.21	186.47	Y
24	3PL	141.18	7	35.86	186.34	Y
25	3PL	224.48	7	58.12	186.50	Y
26	3PL	534.80	7	141.06	186.30	Y
27	3PL	121.49	7	30.60	186.45	Y
28	3PL	182.74	7	46.97	186.25	Y
29	3PL	147.15	7	37.46	186.42	Y
30	3PL	449.63	7	118.30	186.40	Y
31	3PL	84.89	7	20.82	186.42	Y
32	3PL	92.68	7	22.90	186.24	Y
33	3PL	106.67	7	26.64	186.31	Y
34	3PL	255.56	7	66.43	186.25	Y
35	3PL	228.31	7	59.15	186.14	Y
36	3PL	49.90	7	11.47	186.21	Y
37	3PL	146.16	7	37.19	186.25	Y
38	3PL	115.46	7	28.99	186.05	Y
39	3PL	149.84	7	38.18	186.02	Y
40	3PL	162.02	7	41.43	185.69	Y

Table O7. Mathematics Grade 3 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
41	3PL	131.76	7	33.34	185.30	Y
42	2PPC	380.07	16	64.36	186.40	Y
43	2PPC	1680.11	16	294.18	186.32	N
44	2PPC	1488.32	16	260.27	186.24	N
45	2PPC	1031.89	16	179.59	186.26	Y
46	2PPC	924.14	16	160.54	186.26	Y
47	2PPC	473.51	25	63.43	186.23	Y
48	2PPC	1134.31	25	156.88	186.31	Y
49	2PPC	945.16	25	130.13	185.93	Y

Table O8. Mathematics Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	99.02	7	24.59	186.64	Y
2	3PL	132.16	7	33.45	186.62	Y
3	3PL	117.50	7	29.53	186.54	Y
4	3PL	87.22	7	21.44	186.55	Y
5	3PL	64.27	7	15.31	186.54	Y
6	3PL	90.33	7	22.27	186.51	Y
7	3PL	136.70	7	34.66	186.50	Y
8	3PL	111.24	7	27.86	186.41	Y
9	3PL	159.87	7	40.86	186.52	Y
10	3PL	101.98	7	25.39	186.43	Y
11	3PL	66.64	7	15.94	186.36	Y
12	3PL	97.20	7	24.11	186.43	Y
13	3PL	94.40	7	23.36	186.25	Y
14	3PL	102.07	7	25.41	186.42	Y
15	3PL	56.64	7	13.27	186.26	Y
16	3PL	60.34	7	14.25	186.36	Y
17	3PL	65.86	7	15.73	186.24	Y
18	3PL	65.89	7	15.74	186.07	Y
19	3PL	235.08	7	60.96	186.11	Y
20	3PL	74.70	7	18.09	185.43	Y
21	3PL	100.62	7	25.02	186.59	Y
22	3PL	92.81	7	22.93	186.52	Y
23	3PL	249.89	7	64.91	186.57	Y
24	3PL	165.14	7	42.26	186.54	Y
25	3PL	185.26	7	47.64	186.45	Y
26	3PL	105.56	7	26.34	186.21	Y
27	3PL	70.02	7	16.84	186.30	Y

Table O8. Mathematics Grade 4 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
28	3PL	142.01	7	36.08	186.42	Y
29	3PL	186.48	7	47.97	186.51	Y
30	3PL	81.76	7	19.98	186.51	Y
31	3PL	107.54	7	26.87	186.45	Y
32	3PL	145.46	7	37.01	186.49	Y
33	3PL	71.19	7	17.15	186.43	Y
34	3PL	47.01	7	10.69	186.45	Y
35	3PL	101.39	7	25.23	186.09	Y
36	3PL	118.20	7	29.72	186.14	Y
37	3PL	62.73	7	14.89	186.17	Y
38	3PL	142.63	7	36.25	185.97	Y
39	3PL	63.43	7	15.08	186.11	Y
40	3PL	86.21	7	21.17	185.96	Y
41	3PL	96.13	7	23.82	185.91	Y
42	3PL	101.41	7	25.23	185.58	Y
43	2PPC	444.19	16	75.69	186.48	Y
44	2PPC	354.72	16	59.88	186.18	Y
45	2PPC	288.00	16	48.08	186.31	Y
46	2PPC	338.75	16	57.06	186.33	Y
47	2PPC	462.36	16	78.91	186.51	Y
48	2PPC	240.32	16	39.66	186.34	Y
49	2PPC	824.39	25	113.05	186.09	Y
50	2PPC	385.46	25	50.98	186.24	Y
51	2PPC	925.93	25	127.41	186.17	Y
52	2PPC	568.34	25	76.84	186.25	Y

Table O9. Mathematics Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	52.89	7	12.26	186.64	Y
2	3PL	391.03	7	102.64	186.55	Y
3	3PL	241.20	7	62.59	186.42	Y
4	3PL	87.21	7	21.44	186.55	Y
5	3PL	56.83	7	13.32	186.53	Y
6	3PL	193.87	7	49.94	186.49	Y
7	3PL	169.51	7	43.43	186.60	Y
8	3PL	213.92	7	55.30	186.16	Y
9	3PL	308.38	7	80.55	186.55	Y
10	3PL	186.97	7	48.10	186.15	Y
11	3PL	199.95	7	51.57	186.45	Y

Table O9. Mathematics Grade 5 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
12	3PL	269.00	7	70.02	186.42	Y
13	3PL	75.29	7	18.25	186.48	Y
14	3PL	231.53	7	60.01	186.42	Y
15	3PL	73.01	7	17.64	186.33	Y
16	3PL	57.71	7	13.55	186.43	Y
17	3PL	87.28	7	21.46	186.39	Y
18	3PL	106.00	7	26.46	186.38	Y
19	3PL	158.32	7	40.44	185.84	Y
20	3PL	50.27	7	11.56	186.23	Y
21	3PL	93.27	7	23.06	186.61	Y
22	3PL	48.43	7	11.07	186.60	Y
23	3PL	138.27	7	35.08	186.61	Y
24	3PL	91.04	7	22.46	186.57	Y
25	3PL	140.31	7	35.63	186.54	Y
26	3PL	54.04	7	12.57	186.55	Y
27	3PL	243.71	7	63.26	186.58	Y
28	3PL	71.88	7	17.34	186.48	Y
29	3PL	69.40	7	16.68	186.55	Y
30	3PL	39.37	7	8.65	186.46	Y
31	3PL	65.51	7	15.64	186.55	Y
32	3PL	70.64	7	17.01	186.34	Y
33	3PL	71.65	7	17.28	186.55	Y
34	3PL	117.11	7	29.43	186.47	Y
35	3PL	285.93	7	74.55	186.38	Y
36	3PL	130.36	7	32.97	186.50	Y
37	3PL	221.99	7	57.46	186.32	Y
38	3PL	246.32	7	63.96	186.38	Y
39	3PL	195.79	7	50.46	186.38	Y
40	3PL	90.08	7	22.20	186.26	Y
41	3PL	133.40	7	33.78	186.29	Y
42	3PL	94.32	7	23.34	186.06	Y
43	2PPC	282.71	16	47.15	186.37	Y
44	2PPC	1174.50	16	204.80	186.26	N
45	2PPC	395.92	16	67.16	186.23	Y
46	2PPC	427.60	16	72.76	185.89	Y
47	2PPC	667.28	16	115.13	186.01	Y
48	2PPC	629.93	16	108.53	186.34	Y
49	2PPC	4472.62	25	628.99	186.22	N
50	2PPC	606.54	25	82.24	185.94	Y
51	2PPC	811.02	25	111.16	186.25	Y
52	2PPC	876.47	25	120.42	186.09	Y

Table O10. Mathematics Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	192.56	7	49.59	186.54	Y
2	3PL	74.26	7	17.97	186.60	Y
3	3PL	147.45	7	37.54	186.42	Y
4	3PL	90.92	7	22.43	186.59	Y
5	3PL	178.91	7	45.95	186.47	Y
6	3PL	133.36	7	33.77	186.52	Y
7	3PL	77.47	7	18.83	186.50	Y
8	3PL	122.42	7	30.85	186.53	Y
9	3PL	57.69	7	13.55	186.47	Y
10	3PL	123.87	7	31.23	186.53	Y
11	3PL	108.33	7	27.08	186.55	Y
12	3PL	111.15	7	27.84	186.52	Y
13	3PL	90.75	7	22.38	186.33	Y
14	3PL	161.16	7	41.20	186.53	Y
15	3PL	76.99	7	18.70	186.24	Y
16	3PL	164.00	7	41.96	186.51	Y
17	3PL	122.78	7	30.94	186.35	Y
18	3PL	69.25	7	16.64	186.42	Y
19	3PL	80.36	7	19.61	186.39	Y
20	3PL	72.47	7	17.50	186.21	Y
21	3PL	115.00	7	28.86	186.35	Y
22	3PL	73.44	7	17.76	186.26	Y
23	3PL	58.74	7	13.83	186.22	Y
24	3PL	58.98	7	13.89	185.63	Y
25	3PL	87.02	7	21.39	186.48	Y
26	3PL	69.39	7	16.67	186.62	Y
27	3PL	78.11	7	19.00	186.50	Y
28	3PL	113.30	7	28.41	186.58	Y
29	3PL	134.23	7	34.00	186.33	Y
30	3PL	76.00	7	18.44	186.51	Y
31	3PL	65.19	7	15.55	186.44	Y
32	3PL	128.98	7	32.60	186.56	Y
33	3PL	80.78	7	19.72	186.43	Y
34	3PL	139.12	7	35.31	186.38	Y
35	3PL	205.50	7	53.05	186.57	Y
36	3PL	173.44	7	44.48	186.38	Y
37	3PL	97.00	7	24.05	186.39	Y
38	3PL	104.86	7	26.16	186.53	Y
39	3PL	78.54	7	19.12	186.49	Y
40	3PL	78.37	7	19.07	186.25	Y

Table O10. Mathematics Grade 6 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
41	3PL	122.30	7	30.81	186.53	Y
42	3PL	61.63	7	14.60	186.29	Y
43	3PL	267.99	7	69.75	186.46	Y
44	3PL	60.24	7	14.23	186.47	Y
45	3PL	158.00	7	40.36	186.18	Y
46	3PL	408.32	7	107.26	186.34	Y
47	3PL	51.60	7	11.92	186.34	Y
48	3PL	76.22	7	18.50	185.65	Y
49	2PPC	474.08	16	80.98	186.25	Y
50	2PPC	594.78	16	102.32	185.79	Y
51	2PPC	1329.26	16	232.15	185.83	N
52	2PPC	474.97	16	81.13	186.25	Y
53	2PPC	1573.63	16	275.35	184.94	N
54	2PPC	346.05	16	58.34	185.35	Y
55	2PPC	1808.97	25	252.29	184.53	N
56	2PPC	658.56	25	89.60	184.40	Y
57	2PPC	692.18	25	94.35	185.41	Y
58	2PPC	972.32	25	133.97	185.65	Y

Table O11. Mathematics Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	304.69	7	79.56	186.59	Y
2	3PL	244.78	7	63.55	186.41	Y
3	3PL	57.50	7	13.50	186.31	Y
4	3PL	383.01	7	100.49	186.61	Y
5	3PL	183.07	7	47.06	186.56	Y
6	3PL	97.08	7	24.08	186.48	Y
7	3PL	87.42	7	21.49	186.31	Y
8	3PL	44.40	7	9.99	186.34	Y
9	3PL	216.80	7	56.07	186.51	Y
10	3PL	92.02	7	22.72	186.12	Y
11	3PL	81.00	7	19.78	186.22	Y
12	3PL	418.19	7	109.89	186.45	Y
13	3PL	84.33	7	20.67	186.20	Y
14	3PL	42.45	7	9.48	186.33	Y
15	3PL	320.50	7	83.79	186.14	Y
16	3PL	176.13	7	45.20	186.42	Y
17	3PL	140.41	7	35.66	186.29	Y
18	3PL	80.08	7	19.53	186.07	Y

Table O11. Mathematics Grade 7 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
19	3PL	73.78	7	17.85	186.01	Y
20	3PL	141.32	7	35.90	186.22	Y
21	3PL	304.06	7	79.39	185.77	Y
22	3PL	43.47	7	9.75	185.89	Y
23	3PL	52.90	7	12.27	185.94	Y
24	3PL	184.11	7	47.34	185.78	Y
25	3PL	61.33	7	14.52	186.39	Y
26	3PL	30.73	7	6.34	186.53	Y
27	3PL	35.40	7	7.59	186.56	Y
28	3PL	210.81	7	54.47	185.77	Y
29	3PL	61.64	7	14.60	186.27	Y
30	3PL	64.94	7	15.48	186.31	Y
31	3PL	131.47	7	33.27	186.26	Y
32	3PL	85.94	7	21.10	186.18	Y
33	3PL	98.17	7	24.37	186.22	Y
34	3PL	39.18	7	8.60	186.46	Y
35	3PL	102.98	7	25.65	186.23	Y
36	3PL	106.84	7	26.68	186.40	Y
37	3PL	41.63	7	9.26	186.23	Y
38	3PL	211.90	7	54.76	186.26	Y
39	3PL	62.83	7	14.92	186.38	Y
40	3PL	57.00	7	13.36	186.35	Y
41	3PL	74.24	7	17.97	186.36	Y
42	3PL	94.52	7	23.39	186.17	Y
43	3PL	54.36	7	12.66	186.09	Y
44	3PL	38.04	7	8.30	186.00	Y
45	3PL	45.10	7	10.18	185.82	Y
46	3PL	137.64	7	34.91	185.98	Y
47	3PL	65.48	7	15.63	185.62	Y
48	3PL	152.54	7	38.90	185.39	Y
49	2PPC	1112.62	16	193.86	186.25	N
50	2PPC	745.57	16	128.97	185.92	Y
51	2PPC	487.90	16	83.42	185.67	Y
52	2PPC	456.89	16	77.94	185.79	Y
53	2PPC	1583.70	16	277.13	185.75	N
54	2PPC	381.27	16	64.57	185.41	Y
55	2PPC	1117.48	25	154.50	185.06	Y
56	2PPC	759.18	25	103.83	185.51	Y
57	2PPC	699.67	25	95.41	183.85	Y
58	2PPC	744.45	25	101.75	183.40	Y

Table O12. Mathematics Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	122.13	7	30.77	186.62	Y
2	3PL	166.49	7	42.63	186.49	Y
3	3PL	360.72	7	94.54	186.60	Y
4	3PL	119.60	7	30.09	186.55	Y
5	3PL	445.23	7	117.12	186.51	Y
6	3PL	140.26	7	35.62	186.33	Y
7	3PL	61.79	7	14.64	186.34	Y
8	3PL	98.19	7	24.37	186.60	Y
9	3PL	49.04	7	11.24	186.51	Y
10	3PL	98.25	7	24.39	186.45	Y
11	3PL	67.21	7	16.09	186.41	Y
12	3PL	67.18	7	16.08	186.40	Y
13	3PL	42.16	7	9.40	186.39	Y
14	3PL	58.81	7	13.85	186.34	Y
15	3PL	185.70	7	47.76	186.53	Y
16	3PL	141.40	7	35.92	186.48	Y
17	3PL	133.39	7	33.78	186.48	Y
18	3PL	253.36	7	65.84	186.47	Y
19	3PL	92.08	7	22.74	186.42	Y
20	3PL	69.74	7	16.77	186.36	Y
21	3PL	90.04	7	22.19	186.48	Y
22	3PL	60.96	7	14.42	186.15	Y
23	3PL	83.20	7	20.37	186.45	Y
24	3PL	139.24	7	35.34	186.24	Y
25	3PL	278.48	7	72.56	186.58	Y
26	3PL	80.95	7	19.77	186.48	Y
27	3PL	500.68	7	131.94	186.40	Y
28	3PL	48.34	7	11.05	186.28	Y
29	3PL	75.31	7	18.26	186.43	Y
30	3PL	82.28	7	20.12	186.42	Y
31	3PL	64.08	7	15.25	186.37	Y
32	3PL	87.55	7	21.53	186.50	Y
33	3PL	107.44	7	26.84	186.33	Y
34	3PL	59.19	7	13.95	186.54	Y
35	3PL	79.37	7	19.34	186.49	Y
36	3PL	160.06	7	40.91	186.53	Y
37	3PL	98.92	7	24.57	186.53	Y
38	3PL	32.50	7	6.82	185.49	Y
39	3PL	45.83	7	10.38	186.43	Y
40	3PL	317.80	7	83.06	186.27	Y

Table O12. Mathematics Grade 8 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
41	3PL	269.45	7	70.14	186.50	Y
42	3PL	63.07	7	14.99	186.31	Y
43	3PL	159.09	7	40.65	186.47	Y
44	3PL	44.36	7	9.99	186.42	Y
45	3PL	55.93	7	13.08	186.32	Y
46	3PL	81.02	7	19.78	186.27	Y
47	3PL	61.87	7	14.67	186.10	Y
48	3PL	88.73	7	21.84	186.12	Y
49	2PPC	1103.46	16	192.24	184.13	N
50	2PPC	878.18	16	152.41	180.92	Y
51	2PPC	319.61	16	53.67	181.88	Y
52	2PPC	1114.16	16	194.13	185.14	N
53	2PPC	463.78	16	79.16	181.54	Y
54	2PPC	432.23	16	73.58	183.78	Y
55	2PPC	703.71	25	95.98	183.66	Y
56	2PPC	1661.00	25	231.37	184.05	N
57	2PPC	1609.11	25	224.03	176.95	N
58	2PPC	355.26	25	46.71	178.30	Y

Table O13. English Language Arts Grade 3 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
1	1	0.751	-1.289	0.040		
2	1	0.485	-1.962	0.006		
3	1	0.340	0.690	0.139		
4	1	0.685	-1.706	0.013		
5	1	0.593	-0.914	0.075		
6	1	0.487	0.638	0.157		
7	1	0.118	0.707	0.017		
8	1	1.084	-0.314	0.219		
9	1	0.499	-0.927	0.010		
10	1	0.415	-0.066	0.147		
11	1	1.076	0.073	0.198		
12	1	0.585	0.883	0.200		
13	1	0.456	-0.071	0.067		
14	1	1.061	0.557	0.214		
15	1	0.350	-0.559	0.021		
16	1	0.425	0.261	0.134		
17	1	1.039	0.077	0.214		
18	1	0.666	1.124	0.254		
19	1	0.792	0.249	0.161		
20	1	0.755	0.736	0.266		
21	1	0.994	1.002	0.228		
22	1	0.860	0.814	0.202		
23	1	0.687	0.484	0.217		
24	1	0.616	0.288	0.152		
25	1	0.756	-0.075	0.080		
26	1	0.916	-0.681	0.293		
27	1	0.480	-0.220	0.030		
28	1	0.287	-0.442	0.008		
29	1	0.584	-1.291	0.057		
30	1	0.571	0.611	0.165		
31	1	0.404	0.107	0.086		
32	2	1.248	-1.782	2.044		
33	2	1.162	-2.007	1.709		
34	2	1.183	-0.937	2.056		
35	4	1.234	-1.418	-0.176	1.498	2.923
36	2	1.246	-2.520	1.456		
37	2	1.371	-0.881	2.150		
38	2	1.380	-1.486	2.240		
39	2	1.175	-1.406	1.646		
40	2	1.047	-1.054	2.074		
41	4	1.264	-1.918	-0.258	2.027	3.363

Table O14. English Language Arts Grade 4 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
1	1	0.323	-1.928	0.009		
2	1	0.806	-0.137	0.255		
3	1	0.512	-1.194	0.004		
4	1	0.380	-2.079	0.006		
5	1	1.139	1.248	0.176		
6	1	0.375	-0.487	0.035		
7	1	0.573	-1.233	0.034		
8	1	1.049	0.315	0.221		
9	1	0.902	0.403	0.234		
10	1	0.452	-1.324	0.046		
11	1	0.518	0.811	0.150		
12	1	0.451	0.430	0.067		
13	1	0.267	0.141	0.021		
14	1	0.954	-0.134	0.269		
15	1	0.419	-0.365	0.077		
16	1	0.523	-0.350	0.089		
17	1	0.360	0.911	0.146		
18	1	0.362	-0.530	0.008		
19	1	0.735	0.555	0.156		
20	1	0.474	-0.874	0.024		
21	1	0.676	0.285	0.222		
22	1	1.038	0.622	0.170		
23	1	0.699	-1.324	0.130		
24	1	0.933	1.125	0.247		
25	1	0.117	-1.671	0.024		
26	1	0.435	0.151	0.124		
27	1	0.472	0.576	0.152		
28	1	0.814	0.523	0.173		
29	1	0.492	3.601	0.215		
30	1	0.842	0.119	0.203		
31	1	0.801	-0.674	0.199		
32	2	1.268	-2.480	0.197		
33	2	1.202	-2.012	0.616		
34	2	1.469	-1.928	0.859		
35	4	1.312	-2.306	-0.382	1.322	2.941
36	2	1.262	-2.168	1.192		
37	2	1.196	-2.580	0.659		
38	2	1.574	-2.632	0.221		
39	2	0.980	-0.123	1.653		
40	2	1.345	-2.007	0.922		
41	4	1.417	-1.128	0.256	1.725	3.420

Table O15. English Language Arts Grade 5 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
1	1	0.608	-1.615	0.028		
2	1	0.490	-1.731	0.035		
3	1	0.521	-0.752	0.050		
4	1	0.679	0.387	0.222		
5	1	0.851	0.462	0.247		
6	1	1.225	-0.754	0.218		
7	1	0.921	-1.586	0.135		
8	1	0.675	-1.414	0.004		
9	1	0.642	0.440	0.302		
10	1	0.313	0.231	0.160		
11	1	0.229	0.253	0.030		
12	1	0.604	-0.384	0.130		
13	1	0.547	-0.368	0.146		
14	1	0.475	-0.339	0.147		
15	1	0.441	-0.389	0.141		
16	1	0.754	-1.301	0.166		
17	1	0.734	1.119	0.288		
18	1	0.582	-0.069	0.151		
19	1	0.812	-0.838	0.142		
20	1	0.526	-1.248	0.013		
21	1	0.842	1.854	0.192		
22	1	1.149	0.591	0.196		
23	1	0.273	-0.695	0.028		
24	1	0.998	-0.561	0.231		
25	1	1.210	0.526	0.171		
26	1	0.486	-0.218	0.103		
27	1	0.655	0.055	0.216		
28	1	0.782	0.431	0.201		
29	1	0.751	0.354	0.190		
30	1	0.814	1.250	0.284		
31	1	0.688	0.050	0.239		
32	1	0.706	0.985	0.197		
33	1	1.011	-0.649	0.258		
34	1	0.778	0.603	0.193		
35	1	0.895	0.104	0.188		
36	1	0.960	-0.635	0.299		
37	1	0.366	-1.182	0.004		
38	1	0.532	-0.657	0.003		
39	1	0.651	-0.361	0.292		
40	1	0.756	-0.004	0.211		

Table O15. English Language Arts Grade 5 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
41	1	0.543	-1.190	0.046		
42	1	0.414	-1.768	0.007		
43	2	1.410	-3.425	-1.099		
44	2	1.428	-2.311	0.284		
45	2	1.295	-2.772	-0.139		
46	4	1.116	-2.523	-0.891	1.005	2.440
47	2	1.202	-3.717	-0.041		
48	2	1.178	-2.519	0.395		
49	2	1.596	-3.126	-0.173		
50	2	1.652	-3.247	-0.168		
51	2	1.270	-2.690	0.276		
52	4	1.288	-1.967	-0.480	1.037	2.362

Table O16. English Language Arts Grade 6 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
1	1	0.880	-0.438	0.053		
2	1	0.782	-1.641	0.097		
3	1	0.788	-1.911	0.103		
4	1	0.674	0.449	0.193		
5	1	0.559	0.770	0.134		
6	1	0.991	-0.467	0.246		
7	1	0.712	1.717	0.341		
8	1	0.439	-0.568	0.102		
9	1	0.668	0.200	0.150		
10	1	0.677	-0.723	0.149		
11	1	0.432	-0.136	0.039		
12	1	0.901	0.156	0.225		
13	1	0.675	-1.096	0.079		
14	1	0.629	-0.920	0.040		
15	1	0.605	0.182	0.220		
16	1	0.619	0.885	0.224		
17	1	0.817	-0.328	0.231		
18	1	1.047	0.944	0.254		
19	1	0.748	1.089	0.227		
20	1	0.322	0.405	0.029		
21	1	0.672	0.567	0.184		
22	1	0.911	0.095	0.330		
23	1	0.866	1.140	0.175		
24	1	0.518	0.477	0.204		

Table O16. English Language Arts Grade 6 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
25	1	0.700	1.694	0.239		
26	1	0.557	-0.227	0.070		
27	1	0.298	-0.103	0.007		
28	1	0.275	-0.882	0.004		
29	1	0.776	0.916	0.255		
30	1	0.225	1.239	0.074		
31	1	0.896	0.494	0.248		
32	1	1.092	0.789	0.215		
33	1	0.149	2.106	0.019		
34	1	0.644	0.117	0.131		
35	1	0.763	-0.302	0.150		
36	1	0.711	-0.134	0.191		
37	1	0.603	-0.891	0.037		
38	1	0.422	-0.046	0.085		
39	1	0.580	-0.661	0.170		
40	1	1.092	0.246	0.225		
41	1	0.509	-1.047	0.006		
42	1	0.531	-0.977	0.095		
43	2	1.347	-3.123	-0.559		
44	2	1.523	-2.360	0.215		
45	2	1.568	-2.598	0.157		
46	4	1.471	-2.728	-1.377	0.467	2.171
47	2	1.441	-2.606	-0.270		
48	2	1.361	-3.035	0.102		
49	2	1.397	-2.347	0.073		
50	2	1.419	-1.338	0.024		
51	2	1.428	-1.331	0.696		
52	4	1.444	-2.382	-0.805	0.744	2.277

Table O17. English Language Arts Grade 7 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
1	1	0.500	-1.079	0.176		
2	1	0.813	-1.488	0.354		
3	1	0.318	-1.060	0.030		
4	1	1.083	0.248	0.329		
5	1	0.889	-0.536	0.156		
6	1	0.917	0.748	0.174		
7	1	0.528	-1.822	0.014		
8	1	0.457	0.638	0.203		
9	1	0.413	-1.545	0.005		
10	1	0.333	-1.595	0.008		
11	1	0.258	-1.084	0.006		
12	1	0.212	-0.950	0.009		
13	1	0.712	1.407	0.325		
14	1	0.486	0.228	0.263		
15	1	0.975	0.511	0.125		
16	1	0.450	1.080	0.173		
17	1	0.869	-0.075	0.195		
18	1	1.187	0.839	0.204		
19	1	0.452	-0.172	0.116		
20	1	0.968	-0.093	0.210		
21	1	0.650	-0.630	0.154		
22	1	0.352	-0.299	0.022		
23	1	0.827	0.197	0.160		
24	1	0.582	0.756	0.213		
25	1	0.562	1.094	0.202		
26	1	1.370	-0.419	0.264		
27	1	0.449	-0.218	0.048		
28	1	0.873	0.488	0.214		
29	1	0.920	0.446	0.243		
30	1	0.642	1.983	0.220		
31	1	1.185	0.303	0.243		
32	1	0.816	0.741	0.213		
33	1	1.178	0.756	0.255		
34	1	0.501	0.040	0.102		
35	1	0.289	0.081	0.034		
36	1	0.701	-0.578	0.319		
37	1	0.763	-0.657	0.180		
38	1	0.720	0.031	0.221		
39	1	1.021	0.496	0.251		
40	1	0.699	0.079	0.313		

Table O17. English Language Arts Grade 7 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
41	1	0.389	0.473	0.164		
42	1	0.862	0.528	0.188		
43	2	1.437	-3.558	-0.148		
44	2	1.564	-2.923	-0.106		
45	2	1.378	-1.691	0.399		
46	4	1.480	-3.533	-1.398	0.353	2.135
47	2	1.413	-2.944	-0.348		
48	2	1.398	-2.991	-0.420		
49	2	1.613	-3.441	-0.617		
50	2	1.530	-3.367	-0.556		
51	2	1.734	-3.145	-0.401		
52	4	1.462	-3.263	-1.622	0.309	1.989

Table O18. English Language Arts Grade 8 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
1	1	0.427	-3.255	0.007		
2	1	1.072	0.257	0.237		
3	1	0.907	-0.722	0.270		
4	1	0.538	0.133	0.188		
5	1	1.041	0.585	0.169		
6	1	0.528	0.062	0.220		
7	1	0.848	-0.564	0.211		
8	1	0.755	0.009	0.133		
9	1	0.337	0.238	0.206		
10	1	0.318	-0.392	0.032		
11	1	0.358	-1.659	0.003		
12	1	0.110	1.189	0.011		
13	1	0.654	1.035	0.147		
14	1	0.634	-1.310	0.044		
15	1	0.315	0.813	0.002		
16	1	1.183	0.053	0.239		
17	1	1.013	1.695	0.192		
18	1	0.253	-0.872	0.007		
19	1	0.813	0.025	0.280		
20	1	0.886	0.676	0.176		
21	1	0.371	0.078	0.117		
22	1	0.598	0.314	0.234		
23	1	0.407	-0.089	0.125		
24	1	1.279	0.190	0.184		

Table O18. English Language Arts Grade 8 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
25	1	0.734	1.058	0.188		
26	1	1.110	0.544	0.159		
27	1	1.029	-0.297	0.230		
28	1	1.030	0.204	0.250		
29	1	1.068	-0.160	0.281		
30	1	0.954	-0.257	0.167		
31	1	0.302	-0.125	0.049		
32	1	1.330	0.518	0.196		
33	1	1.277	0.103	0.254		
34	1	1.115	-0.267	0.257		
35	1	0.520	-0.260	0.036		
36	1	1.179	-0.282	0.152		
37	1	0.403	-0.516	0.096		
38	1	0.742	-0.366	0.311		
39	1	0.464	1.295	0.279		
40	1	0.536	-1.010	0.147		
41	1	0.836	-0.531	0.157		
42	1	0.876	-0.852	0.215		
43	2	1.542	-3.541	-0.530		
44	2	1.290	-3.354	-0.181		
45	2	1.467	-3.483	-0.979		
46	4	1.400	-2.643	-1.491	0.067	1.516
47	2	1.345	-2.486	-0.064		
48	2	1.493	-2.869	-0.385		
49	2	1.580	-3.778	-0.562		
50	2	1.325	-2.810	-0.233		
51	2	1.570	-2.993	-0.269		
52	4	1.201	-2.825	-1.360	0.188	1.369

Table O19. Mathematics Grade 3 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.946	-1.374	0.004	
2	1	0.876	-2.053	0.010	
3	1	0.981	0.177	0.175	
4	1	0.773	-2.236	0.012	
5	1	0.869	-0.966	0.391	
6	1	0.783	-1.296	0.004	
7	1	0.785	-0.878	0.356	
8	1	1.170	-0.595	0.197	
9	1	1.132	-0.145	0.272	
10	1	0.979	0.117	0.033	
11	1	1.067	0.288	0.316	
12	1	0.838	-0.617	0.131	
13	1	1.090	-0.939	0.105	
14	1	0.912	-1.334	0.044	
15	1	0.703	-0.206	0.184	
16	1	0.526	0.992	0.121	
17	1	0.861	0.059	0.207	
18	1	1.253	-0.609	0.139	
19	1	1.087	0.629	0.213	
20	1	0.850	-0.547	0.168	
21	1	1.112	0.056	0.453	
22	1	0.808	0.411	0.292	
23	1	0.751	-0.544	0.129	
24	1	0.889	1.023	0.170	
25	1	0.458	-1.328	0.009	
26	1	1.508	1.322	0.181	
27	1	1.271	-0.399	0.135	
28	1	0.990	0.493	0.112	
29	1	0.644	-0.376	0.070	
30	1	0.839	0.487	0.095	
31	1	0.771	-0.108	0.209	
32	1	0.802	-0.721	0.229	
33	1	0.805	-0.291	0.156	
34	1	0.671	-1.571	0.003	
35	1	1.030	-1.132	0.048	
36	1	0.632	-1.494	0.068	
37	1	1.175	-0.237	0.038	
38	1	0.872	0.032	0.136	
39	1	1.108	-1.456	0.067	
40	1	1.172	-0.777	0.111	

Table O19. Mathematics Grade 3 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
41	1	1.232	0.194	0.221	
42	2	1.280	-0.740	1.632	
43	2	1.119	-1.506	0.828	
44	2	1.201	0.275	0.077	
45	2	1.368	-0.626	0.389	
46	2	0.914	-0.068	0.887	
47	3	1.288	0.088	-0.107	0.912
48	3	0.938	-0.961	-1.667	1.300
49	3	1.071	-0.912	0.691	0.995

Table O20. Mathematics Grade 4 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.865	-1.425	0.330	
2	1	0.937	-0.497	0.270	
3	1	1.181	-0.751	0.089	
4	1	0.788	-0.819	0.127	
5	1	0.729	-1.420	0.347	
6	1	0.906	0.070	0.273	
7	1	0.969	-0.085	0.210	
8	1	1.274	-0.504	0.155	
9	1	0.943	-0.343	0.137	
10	1	1.171	0.590	0.227	
11	1	0.840	0.670	0.304	
12	1	0.523	0.629	0.044	
13	1	0.959	-0.594	0.178	
14	1	1.072	0.814	0.246	
15	1	1.067	-0.139	0.175	
16	1	0.858	-0.842	0.235	
17	1	0.853	-0.114	0.141	
18	1	0.832	-0.307	0.332	
19	1	0.874	-1.514	0.019	
20	1	0.834	-0.294	0.201	
21	1	0.901	0.156	0.183	
22	1	0.991	0.212	0.064	
23	1	0.974	-1.437	0.010	
24	1	0.796	-0.389	0.140	
25	1	1.196	0.017	0.095	
26	1	1.098	-0.097	0.179	

Table O20. Mathematics Grade 4 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
27	1	1.282	-0.168	0.211	
28	1	0.848	1.243	0.264	
29	1	0.929	0.785	0.142	
30	1	1.019	0.010	0.230	
31	1	1.037	0.576	0.085	
32	1	1.217	-0.275	0.058	
33	1	0.794	0.405	0.124	
34	1	0.843	-0.248	0.274	
35	1	0.785	-0.268	0.171	
36	1	1.170	0.024	0.080	
37	1	0.953	-0.102	0.408	
38	1	0.726	-0.199	0.104	
39	1	1.253	-0.561	0.175	
40	1	1.243	0.129	0.103	
41	1	0.965	0.195	0.240	
42	1	0.811	0.097	0.271	
43	2	1.060	2.399	-0.986	
44	2	1.260	0.934	-1.141	
45	2	1.411	0.503	1.344	
46	2	0.915	-0.961	-0.958	
47	2	1.166	-2.907	-0.794	
48	2	1.099	1.593	-0.122	
49	3	1.162	0.184	1.180	1.139
50	3	1.059	0.976	-0.345	-0.203
51	3	1.123	-0.054	-0.740	1.342
52	3	0.991	0.201	-0.125	0.562

Table O21. Mathematics Grade 5 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.858	-1.212	0.491	
2	1	0.645	-0.906	0.005	
3	1	0.909	-1.173	0.107	
4	1	1.165	-0.012	0.476	
5	1	0.927	-0.078	0.255	
6	1	1.143	-1.348	0.070	
7	1	0.645	-2.022	0.007	
8	1	1.723	1.416	0.211	
9	1	0.811	-1.094	0.006	
10	1	1.328	1.425	0.132	
11	1	1.258	-1.085	0.121	

Table O21. Mathematics Grade 5 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
12	1	1.430	0.536	0.119	
13	1	0.595	-1.864	0.342	
14	1	0.969	0.344	0.075	
15	1	1.164	0.528	0.246	
16	1	0.863	-0.378	0.267	
17	1	0.580	0.205	0.058	
18	1	1.171	-0.535	0.322	
19	1	1.234	1.143	0.252	
20	1	0.727	0.321	0.209	
21	1	0.668	0.434	0.317	
22	1	0.737	-0.675	0.259	
23	1	1.018	-1.242	0.101	
24	1	1.152	-0.489	0.269	
25	1	1.573	0.610	0.229	
26	1	1.105	-0.457	0.499	
27	1	0.568	-1.823	0.006	
28	1	0.608	0.240	0.196	
29	1	0.980	-1.412	0.385	
30	1	0.644	0.985	0.265	
31	1	0.766	0.462	0.336	
32	1	0.933	0.870	0.195	
33	1	1.354	0.141	0.295	
34	1	1.249	0.506	0.145	
35	1	0.910	-0.741	0.055	
36	1	0.889	-1.839	0.171	
37	1	1.231	1.171	0.219	
38	1	1.726	0.736	0.119	
39	1	1.398	0.705	0.246	
40	1	0.940	1.466	0.211	
41	1	0.973	-0.379	0.159	
42	1	0.942	-0.365	0.249	
43	2	1.129	0.097	0.743	
44	2	1.000	0.453	3.178	
45	2	0.714	0.804	-0.506	
46	2	1.326	-0.087	1.272	
47	2	1.698	1.289	1.643	
48	2	1.144	0.061	1.102	
49	3	0.875	-0.251	2.643	1.850
50	3	1.356	0.980	1.641	2.026
51	3	1.098	-0.733	1.054	-1.040
52	3	1.174	0.430	0.199	0.707

Table O22. Mathematics Grade 6 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	1.317	-1.165	0.255	
2	1	1.015	-0.553	0.268	
3	1	1.022	-0.016	0.071	
4	1	0.658	-0.668	0.263	
5	1	0.694	-0.707	0.107	
6	1	1.374	0.365	0.235	
7	1	0.987	0.304	0.309	
8	1	0.687	0.795	0.346	
9	1	1.268	0.793	0.304	
10	1	0.841	1.047	0.343	
11	1	1.104	0.576	0.230	
12	1	0.985	-0.737	0.179	
13	1	1.017	0.496	0.243	
14	1	0.959	-0.670	0.073	
15	1	0.898	0.440	0.111	
16	1	0.731	0.857	0.074	
17	1	0.955	0.342	0.269	
18	1	0.793	1.344	0.209	
19	1	1.041	0.068	0.196	
20	1	1.154	0.911	0.115	
21	1	0.838	-0.128	0.191	
22	1	0.898	-0.182	0.139	
23	1	1.139	0.764	0.209	
24	1	1.388	0.179	0.187	
25	1	1.188	1.153	0.143	
26	1	1.169	0.728	0.173	
27	1	1.471	0.710	0.333	
28	1	1.049	-0.599	0.174	
29	1	1.256	0.692	0.371	
30	1	0.767	-0.261	0.119	
31	1	1.091	-0.527	0.315	
32	1	1.300	-0.733	0.246	
33	1	1.003	0.633	0.296	
34	1	0.949	1.376	0.063	
35	1	0.955	-0.885	0.099	
36	1	1.083	1.176	0.369	
37	1	1.198	0.510	0.182	
38	1	0.750	1.975	0.135	
39	1	1.120	-0.509	0.241	
40	1	1.065	0.073	0.160	

Table O22. Mathematics Grade 6 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
41	1	0.698	-1.650	0.325	
42	1	1.129	0.462	0.201	
43	1	0.908	-1.517	0.085	
44	1	0.814	0.780	0.154	
45	1	1.730	0.452	0.177	
46	1	0.391	-1.296	0.006	
47	1	0.941	0.357	0.176	
48	1	1.340	1.042	0.275	
49	2	1.052	-0.023	-0.458	
50	2	1.075	0.131	0.590	
51	2	1.176	-0.359	-1.256	
52	2	0.836	-0.326	0.975	
53	2	1.312	-0.791	1.336	
54	2	1.214	0.225	2.366	
55	3	1.234	1.081	-0.190	0.988
56	3	0.811	1.299	0.371	-0.255
57	3	1.531	-0.631	1.516	2.345
58	3	1.137	0.068	0.266	1.392

Table O23. Mathematics Grade 7 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.524	-1.840	0.146	
2	1	0.974	-1.170	0.187	
3	1	1.285	-0.106	0.288	
4	1	0.874	-2.779	0.018	
5	1	1.096	-0.563	0.121	
6	1	1.360	0.305	0.228	
7	1	0.786	1.382	0.227	
8	1	1.278	0.931	0.238	
9	1	0.739	-0.735	0.103	
10	1	1.160	1.408	0.251	
11	1	1.207	0.272	0.233	
12	1	1.922	1.569	0.079	
13	1	1.475	0.858	0.188	
14	1	1.170	0.939	0.264	
15	1	1.807	1.530	0.197	
16	1	1.130	-0.909	0.191	
17	1	0.948	0.164	0.365	
18	1	1.372	1.195	0.198	

Table O23. Mathematics Grade 7 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
19	1	1.133	1.092	0.165	
20	1	1.319	0.162	0.111	
21	1	1.720	1.568	0.223	
22	1	1.445	1.173	0.208	
23	1	1.364	1.159	0.391	
24	1	0.383	1.728	0.364	
25	1	1.175	0.168	0.378	
26	1	0.946	1.134	0.231	
27	1	0.745	0.004	0.406	
28	1	2.116	1.390	0.293	
29	1	1.867	1.465	0.268	
30	1	1.701	0.602	0.243	
31	1	1.498	0.731	0.234	
32	1	1.560	0.933	0.162	
33	1	1.467	1.338	0.267	
34	1	0.818	0.500	0.208	
35	1	1.007	1.252	0.336	
36	1	1.171	0.727	0.315	
37	1	1.386	0.246	0.244	
38	1	0.628	-0.235	0.154	
39	1	1.080	0.095	0.306	
40	1	1.204	-0.089	0.340	
41	1	0.777	-0.026	0.148	
42	1	1.701	0.559	0.257	
43	1	1.222	0.808	0.286	
44	1	1.268	0.839	0.246	
45	1	1.469	0.684	0.252	
46	1	1.102	-0.440	0.179	
47	1	1.295	1.225	0.126	
48	1	1.086	0.965	0.123	
49	2	1.021	-0.211	-0.492	
50	2	1.728	-1.381	1.761	
51	2	0.872	-2.117	0.521	
52	2	1.531	0.378	-0.285	
53	2	1.571	-1.323	0.407	
54	2	1.114	0.774	-0.709	
55	3	1.272	0.291	0.925	0.919
56	3	1.160	0.085	0.250	0.835
57	3	1.538	1.217	0.823	1.839
58	3	1.305	1.584	0.395	1.120

Table O24. Mathematics Grade 8 Operational Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.653	-1.778	0.566	
2	1	0.733	-0.938	0.202	
3	1	0.772	-1.457	0.189	
4	1	0.614	-0.374	0.145	
5	1	0.924	-1.344	0.136	
6	1	0.773	0.289	0.202	
7	1	1.007	0.331	0.158	
8	1	0.977	0.822	0.116	
9	1	1.004	0.137	0.201	
10	1	0.940	0.081	0.223	
11	1	1.431	-0.001	0.284	
12	1	1.636	1.289	0.241	
13	1	0.954	0.118	0.246	
14	1	1.282	0.917	0.346	
15	1	1.473	-0.793	0.249	
16	1	0.794	-0.535	0.231	
17	1	0.930	-0.250	0.093	
18	1	1.318	0.806	0.196	
19	1	1.393	1.602	0.228	
20	1	0.722	0.235	0.140	
21	1	0.405	0.105	0.250	
22	1	1.102	0.998	0.273	
23	1	0.913	-0.190	0.224	
24	1	1.653	1.154	0.167	
25	1	1.038	-0.641	0.073	
26	1	0.956	-0.228	0.380	
27	1	1.579	0.752	0.305	
28	1	1.076	0.693	0.236	
29	1	1.365	0.397	0.220	
30	1	1.151	0.057	0.328	
31	1	1.025	0.684	0.195	
32	1	0.761	-0.248	0.293	
33	1	1.575	0.610	0.215	
34	1	1.053	-0.089	0.213	
35	1	0.876	-0.058	0.215	
36	1	0.526	-1.041	0.083	
37	1	0.630	-0.355	0.228	
38	1	0.769	0.662	0.207	
39	1	0.797	0.945	0.206	
40	1	1.660	0.770	0.242	

Table O24. Mathematics Grade 8 Operational Item Parameter Estimates (cont.)

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
41	1	0.854	-1.031	0.148	
42	1	1.074	0.712	0.194	
43	1	1.243	-0.636	0.232	
44	1	1.076	0.957	0.164	
45	1	1.291	1.383	0.231	
46	1	1.193	0.580	0.147	
47	1	0.884	0.322	0.180	
48	1	1.105	-0.335	0.242	
49	2	1.206	-0.353	-0.234	
50	2	1.112	0.456	-0.247	
51	2	1.204	1.519	0.105	
52	2	0.888	0.429	0.067	
53	2	1.155	0.722	-0.361	
54	2	0.534	1.644	-0.991	
55	3	1.249	0.354	1.017	1.119
56	3	0.955	0.206	-0.011	-0.059
57	3	0.853	0.222	-0.269	-0.397
58	3	1.250	0.911	0.432	0.424

Appendix P: Derivation and Estimation of Classification Consistency and Accuracy

Classification Consistency

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

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

where

$g(\theta)$ is the density of θ.

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

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

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

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

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

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

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

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

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

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

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

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

The agreement index (classification consistency) can be computed as

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

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

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

Then, Kappa (Cohen, 1960) is

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

Classification Accuracy

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

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

where

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

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

Estimating Classification Indices

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

Files needed:

- Raw-to-Scale score conversion file
    a. Contains the raw-to-scale score conversions
    b. This is used to provide both raw and scale score classification estimates, which is useful when the raw-to-scale score transformation is not one-to-one.
- Cut score file
    a. Contains the cut scores to be used
    b. Results are provided for all cut scores simultaneously (all performance levels), as well as the estimates based on each of the cut scores separately (Level 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 Q: Raw Score-to-Scale Score and Scale Score Frequency Tables

Tables Q1–Q12 show the raw score-to-scale score conversion tables, while Tables Q13–Q24 show the scale score distributions, by frequency (n-count), percent, cumulative frequency, and cumulative percent. The data in the tables include all students with valid scores.

Table Q1. ELA Grade 3 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	147	52	28	300	10
1	155	46	29	303	10
2	163	40	30	306	10
3	171	35	31	310	10
4	179	31	32	313	10
5	187	27	33	316	10
6	195	24	34	320	10
7	203	21	35	323	10
8	211	19	36	326	10
9	219	17	37	329	10
10	227	16	38	333	10
11	233	15	39	336	10
12	239	14	40	340	10
13	244	13	41	344	10
14	249	13	42	347	10
15	254	12	43	351	11
16	258	12	44	358	11
17	262	11	45	360	11
18	266	11	46	365	12
19	270	11	47	370	12
20	274	11	48	375	13
21	277	11	49	382	14
22	281	10	50	389	15
23	284	10	51	397	16
24	287	10	52	405	18
25	291	10	53	413	21
26	294	10	54	421	23
27	297	10	55	429	26

Table Q2. ELA Grade 4 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	138	52	28	295	10
1	146	46	29	298	9
2	154	41	30	301	9
3	162	36	31	304	9
4	170	32	32	307	9
5	178	28	33	310	9
6	186	25	34	313	9
7	194	22	35	316	9
8	202	20	36	320	9
9	210	18	37	322	9
10	218	16	38	325	9
11	226	15	39	329	9
12	232	14	40	332	10
13	237	13	41	335	10
14	242	13	42	339	10
15	247	12	43	343	10
16	252	12	44	347	11
17	256	11	45	351	11
18	260	11	46	356	11
19	264	11	47	361	12
20	268	11	48	367	13
21	271	11	49	374	14
22	275	10	50	382	16
23	278	10	51	391	18
24	282	10	52	399	20
25	287	10	53	407	22
26	288	10	54	415	25
27	291	10	55	423	28

Table Q3. ELA Grade 5 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	97	84
1	105	74
2	113	65
3	121	56
4	129	49
5	137	43
6	145	37
7	153	33
8	161	29
9	169	25
10	177	22
11	185	20
12	193	17
13	201	16
14	209	14
15	217	13
16	225	12
17	229	11
18	233	11
19	237	11
20	241	10
21	244	10
22	247	10
23	251	10
24	254	10
25	257	10
26	260	9
27	262	9
28	265	9
29	268	9
30	271	9
31	273	9
32	276	9
33	278	9
34	281	9
35	283	9
36	286	9
37	289	8
38	291	8
39	293	8
40	296	8
41	298	8
42	301	8
43	303	8
44	306	9
45	309	9
46	311	9
47	314	9
48	317	9
49	320	9
50	323	9
51	326	9
52	329	9
53	332	10
54	336	10
55	339	10
56	346	11
57	348	11
58	352	12
59	357	12
60	363	13
61	370	14
62	378	16
63	389	19
64	397	21
65	405	24
66	413	27

Table Q4. ELA Grade 6 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	117	85
1	125	74
2	133	64
3	141	56
4	149	48
5	157	41
6	165	35
7	173	30
8	181	26
9	190	22
10	198	19
11	206	16
12	214	15
13	222	13
14	227	12
15	232	12
16	236	11
17	240	11
18	243	10
19	247	10
20	250	10
21	253	10
22	257	9
23	259	9
24	262	9
25	265	9
26	268	9
27	270	9
28	273	9
29	276	9
30	278	8
31	281	8
32	283	8
33	285	8
34	288	8
35	290	8
36	293	8
37	295	8
38	297	8
39	300	8
40	302	8
41	304	8
42	307	8
43	309	8
44	312	8
45	314	8
46	317	8
47	320	8
48	322	9
49	325	9
50	328	9
51	331	9
52	334	9
53	338	9
54	340	10
55	344	10
56	348	10
57	352	11
58	357	12
59	362	12
60	368	13
61	376	15
62	385	17
63	397	21
64	405	24
65	413	27
66	421	31

Table Q5. ELA Grade 7 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	98	89
1	106	79
2	114	70
3	122	62
4	130	54
5	138	48
6	146	41
7	154	36
8	162	31
9	170	27
10	178	23
11	186	20
12	194	18
13	202	16
14	210	14
15	218	13
16	226	12
17	230	11
18	235	11
19	239	11
20	242	10
21	246	10
22	250	10
23	253	10
24	256	10
25	259	10
26	262	9
27	265	9
28	268	9
29	271	9
30	274	9
31	277	9
32	279	9
33	282	8
34	284	8
35	287	8
36	289	8
37	292	8
38	294	8
39	297	8
40	299	8
41	301	8
42	304	8
43	306	8
44	309	8
45	311	8
46	313	8
47	318	8
48	319	8
49	321	8
50	324	8
51	327	9
52	330	9
53	333	9
54	336	9
55	340	10
56	343	10
57	347	10
58	352	11
59	357	12
60	363	13
61	370	14
62	379	17
63	390	20
64	398	23
65	406	26
66	414	29

Table Q6. ELA Grade 8 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	100	80
1	107	72
2	115	64
3	123	57
4	131	50
5	139	44
6	147	38
7	155	33
8	163	29
9	171	25
10	179	22
11	187	19
12	194	17
13	202	15
14	210	14
15	218	13
16	226	12
17	231	11
18	235	11
19	239	11
20	243	10
21	246	10
22	250	10
23	253	10
24	256	10
25	259	9
26	262	9
27	265	9
28	268	9
29	270	9
30	273	9
31	275	9
32	278	8
33	280	8
34	284	8
35	285	8
36	287	8
37	290	8
38	292	8
39	294	8
40	296	8
41	299	8
42	301	7
43	303	7
44	305	7
45	308	8
46	310	8
47	312	8
48	316	8
49	317	8
50	320	8
51	323	8
52	325	8
53	328	9
54	332	9
55	335	9
56	339	10
57	343	10
58	348	11
59	353	12
60	359	13
61	367	15
62	376	17
63	389	22
64	397	25
65	405	28
66	412	32

Table Q7. Mathematics Grade 3 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	137	76
1	145	64
2	153	55
3	161	47
4	169	40
5	177	34
6	185	29
7	193	25
8	201	22
9	209	19
10	217	16
11	225	14
12	231	13
13	236	12
14	240	11
15	244	11
16	248	10
17	252	10
18	255	9
19	258	9
20	261	9
21	263	9
22	266	8
23	269	8
24	271	8
25	273	8
26	276	8
27	278	8
28	280	8
29	282	7
30	285	7
31	287	7
32	289	7
33	291	7
34	293	7
35	295	7
36	297	7
37	299	7
38	302	7
39	304	7
40	306	7
41	308	7
42	310	7
43	314	8
44	315	8
45	318	8
46	320	8
47	323	8
48	326	8
49	329	9
50	332	9
51	336	9
52	340	10
53	344	10
54	348	11
55	354	12
56	360	13
57	369	15
58	381	19
59	389	23
60	397	26

Table Q8. Mathematics Grade 4 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	137	95
1	145	83
2	153	72
3	161	62
4	169	54
5	177	46
6	186	39
7	194	34
8	202	29
9	210	25
10	218	22
11	226	19
12	234	16
13	240	15
14	246	13
15	250	12
16	254	11
17	258	10
18	261	10
19	264	9
20	267	9
21	269	9
22	272	8
23	274	8
24	276	8
25	278	8
26	280	7
27	283	7
28	284	7
29	286	7
30	288	7
31	290	7
32	291	7
33	293	7
34	295	6
35	296	6
36	298	6
37	300	6
38	301	6
39	303	6
40	305	6
41	306	6
42	308	6
43	310	6
44	311	6
45	314	7
46	315	7
47	317	7
48	319	7
49	321	7
50	323	7
51	325	7
52	327	7
53	329	7
54	331	8
55	334	8
56	337	8
57	341	9
58	343	9
59	346	10
60	351	10
61	355	11
62	361	13
63	368	14
64	379	18
65	397	26
66	405	31

Table Q9. Mathematics Grade 5 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	127	110
1	135	95
2	143	82
3	151	71
4	159	62
5	167	54
6	175	47
7	183	41
8	191	35
9	199	30
10	207	26
11	215	22
12	223	19
13	232	16
14	238	14
15	244	13
16	249	12
17	254	11
18	258	11
19	261	10
20	265	10
21	268	10
22	272	9
23	275	9
24	278	9
25	280	9
26	283	9
27	286	8
28	288	8
29	290	8
30	294	8
31	295	8
32	297	8
33	300	8
34	302	7
35	304	7
36	306	7
37	308	7
38	310	7
39	312	7
40	314	7
41	316	7
42	319	7
43	320	7
44	322	7
45	324	7
46	326	7
47	328	7
48	330	7
49	333	7
50	335	7
51	337	7
52	340	7
53	342	7
54	346	8
55	348	8
56	351	8
57	354	9
58	358	9
59	362	10
60	368	11
61	373	12
62	381	14
63	391	17
64	399	20
65	407	23
66	415	26

Table Q10. Mathematics Grade 6 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	125	137
1	133	122
2	141	109
3	150	96
4	158	84
5	166	74
6	174	65
7	182	56
8	190	48
9	198	41
10	206	34
11	214	29
12	223	24
13	233	19
14	241	16
15	247	14
16	252	12
17	256	12
18	260	11
19	263	10
20	267	10
21	270	9
22	272	9
23	275	9
24	278	8
25	280	8
26	282	8
27	284	8
28	286	8
29	289	7
30	291	7
31	292	7
32	294	7
33	296	7
34	298	7
35	300	7
36	301	7
37	303	7
38	305	7
39	306	6
40	308	6
41	310	6
42	311	6
43	313	6
44	315	6
45	316	6
46	318	6
47	319	6
48	321	6
49	323	6
50	324	6
51	326	6
52	328	6
53	330	6
54	332	7
55	334	7
56	336	7
57	338	7
58	340	7
59	342	7
60	345	8
61	347	8
62	350	8
63	353	8
64	357	9
65	360	9
66	365	10
67	370	11
68	376	13
69	384	15
70	395	18
71	403	21
72	411	25

Table Q11. Mathematics Grade 7 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	124	121
1	132	103
2	140	88
3	148	75
4	155	66
5	163	56
6	171	49
7	179	42
8	187	37
9	194	34
10	202	30
11	210	27
12	218	25
13	226	23
14	233	20
15	244	17
16	252	15
17	258	13
18	264	12
19	268	11
20	272	10
21	276	10
22	279	9
23	282	9
24	285	8
25	287	8
26	290	8
27	293	7
28	294	7
29	296	7
30	298	7
31	300	7
32	302	7
33	304	6
34	305	6
35	307	6
36	309	6
37	310	6
38	312	6
39	313	6
40	315	6
41	316	6
42	318	6
43	319	6
44	321	5
45	322	5
46	323	5
47	325	5
48	326	5
49	328	5
50	329	5
51	331	5
52	332	5
53	334	5
54	335	5
55	337	5
56	339	5
57	340	5
58	342	6
59	344	6
60	346	6
61	348	6
62	350	6
63	352	6
64	354	6
65	357	7
66	360	7
67	364	8
68	368	9
69	374	10
70	382	13
71	390	17
72	398	22

Table Q12. Mathematics Grade 8 Raw Score-to-Scale Score Table

Raw Score	Scale Score	Standard Error
0	124	191
1	132	168
2	140	146
3	148	127
4	156	110
5	164	94
6	172	81
7	179	70
8	187	59
9	195	50
10	203	42
11	211	35
12	219	29
13	227	24
14	237	19
15	244	17
16	250	15
17	255	13
18	259	12
19	263	11
20	267	10
21	270	10
22	273	9
23	275	9
24	278	9
25	280	8
26	282	8
27	284	8
28	287	8
29	288	8
30	290	7
31	292	7
32	294	7
33	296	7
34	297	7
35	299	7
36	301	7
37	302	7
38	304	7
39	306	6
40	307	6
41	309	6
42	310	6
43	312	6
44	313	6
45	315	6
46	316	6
47	318	6
48	319	6
49	321	6
50	322	6
51	324	6
52	326	6
53	327	6
54	329	6
55	331	6
56	332	6
57	334	6
58	336	6
59	338	7
60	340	7
61	343	7
62	345	7
63	349	8
64	351	8
65	354	8
66	357	9
67	362	10
68	367	11
69	374	13
70	384	17
71	392	20
72	400	25

Table Q13. English Language Arts Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
147	26	0.01%	26	0.01%
155	34	0.02%	60	0.03%
163	101	0.06%	161	0.09%
171	169	0.10%	330	0.19%
179	327	0.18%	657	0.37%
187	618	0.35%	1,275	0.72%
195	902	0.51%	2,177	1.23%
203	1,253	0.71%	3,430	1.93%
211	1,617	0.91%	5,047	2.84%
219	1,985	1.12%	7,032	3.96%
227	2,203	1.24%	9,235	5.20%
233	2,454	1.38%	11,689	6.58%
239	2,712	1.53%	14,401	8.11%
244	2,909	1.64%	17,310	9.75%
249	3,154	1.78%	20,464	11.53%
254	3,410	1.92%	23,874	13.45%
258	3,712	2.09%	27,586	15.54%
262	3,868	2.18%	31,454	17.72%
266	4,146	2.34%	35,600	20.05%
270	4,396	2.48%	39,996	22.53%
274	4,656	2.62%	44,652	25.15%
277	4,984	2.81%	49,636	27.96%
281	5,130	2.89%	54,766	30.85%
284	5,489	3.09%	60,255	33.94%
287	5,574	3.14%	65,829	37.08%
291	5,705	3.21%	71,534	40.30%
294	6,191	3.49%	77,725	43.78%
297	6,389	3.60%	84,114	47.38%
300	6,329	3.57%	90,443	50.95%
303	6,454	3.64%	96,897	54.58%
306	6,347	3.58%	103,244	58.16%
310	6,409	3.61%	109,653	61.77%
313	6,531	3.68%	116,184	65.45%
316	6,324	3.56%	122,508	69.01%
320	6,229	3.51%	128,737	72.52%
323	6,014	3.39%	134,751	75.91%
326	5,678	3.20%	140,429	79.11%
329	5,349	3.01%	145,778	82.12%
333	5,049	2.84%	150,827	84.96%

Table Q13. English Language Arts Grade 3 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
336	4,607	2.60%	155,434	87.56%
340	3,993	2.25%	159,427	89.81%
344	3,679	2.07%	163,106	91.88%
347	3,108	1.75%	166,214	93.63%
351	2,758	1.55%	168,972	95.19%
358	2,270	1.28%	171,242	96.46%
360	1,838	1.04%	173,080	97.50%
365	1,358	0.76%	174,438	98.26%
370	1,075	0.61%	175,513	98.87%
375	765	0.43%	176,278	99.30%
382	524	0.30%	176,802	99.60%
389	367	0.21%	177,169	99.80%
397	209	0.12%	177,378	99.92%
405	85	0.05%	177,463	99.97%
413	41	0.02%	177,504	99.99%
421	13	0.01%	177,517	100.00%
429	2	0.00%	177,519	100.00%

Table Q14. English Language Arts Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
138	5	0.00%	5	0.00%
146	19	0.01%	24	0.01%
154	37	0.02%	61	0.03%
162	68	0.04%	129	0.07%
170	162	0.09%	291	0.16%
178	277	0.16%	568	0.32%
186	547	0.31%	1,115	0.62%
194	784	0.44%	1,899	1.06%
202	1,078	0.60%	2,977	1.67%
210	1,330	0.75%	4,307	2.41%
218	1,607	0.90%	5,914	3.31%
226	1,838	1.03%	7,752	4.34%
232	2,106	1.18%	9,858	5.52%
237	2,396	1.34%	12,254	6.87%
242	2,623	1.47%	14,877	8.33%
247	2,827	1.58%	17,704	9.92%
252	3,102	1.74%	20,806	11.66%
256	3,385	1.90%	24,191	13.55%
260	3,549	1.99%	27,740	15.54%
264	3,849	2.16%	31,589	17.70%
268	4,097	2.30%	35,686	19.99%
271	4,296	2.41%	39,982	22.40%
275	4,582	2.57%	44,564	24.97%
278	4,933	2.76%	49,497	27.73%
282	5,060	2.83%	54,557	30.57%
287	5,231	2.93%	59,788	33.50%
288	5,599	3.14%	65,387	36.63%
291	5,531	3.10%	70,918	39.73%
295	5,720	3.20%	76,638	42.94%
298	5,956	3.34%	82,594	46.27%
301	6,049	3.39%	88,643	49.66%
304	6,042	3.39%	94,685	53.05%
307	6,141	3.44%	100,826	56.49%
310	6,308	3.53%	107,134	60.02%
313	6,235	3.49%	113,369	63.51%
316	6,278	3.52%	119,647	67.03%
320	6,248	3.50%	125,895	70.53%
322	5,950	3.33%	131,845	73.87%
325	5,669	3.18%	137,514	77.04%

Table Q14. English Language Arts Grade 4 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
329	5,663	3.17%	143,177	80.21%
332	5,316	2.98%	148,493	83.19%
335	4,925	2.76%	153,418	85.95%
339	4,696	2.63%	158,114	88.58%
343	4,257	2.38%	162,371	90.97%
347	3,750	2.10%	166,121	93.07%
351	3,262	1.83%	169,383	94.90%
356	2,701	1.51%	172,084	96.41%
361	2,107	1.18%	174,191	97.59%
367	1,606	0.90%	175,797	98.49%
374	1,118	0.63%	176,915	99.12%
382	762	0.43%	177,677	99.54%
391	465	0.26%	178,142	99.80%
399	217	0.12%	178,359	99.93%
407	107	0.06%	178,466	99.99%
415	23	0.01%	178,489	100.00%
423	3	0.00%	178,492	100.00%

Table Q15. English Language Arts Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
97	13	0.01%	13	0.01%
105	8	0.00%	21	0.01%
113	8	0.00%	29	0.02%
121	17	0.01%	46	0.03%
129	24	0.01%	70	0.04%
137	54	0.03%	124	0.07%
145	90	0.05%	214	0.13%
153	180	0.11%	394	0.23%
161	249	0.15%	643	0.38%
169	386	0.23%	1,029	0.60%
177	517	0.30%	1,546	0.90%
185	681	0.40%	2,227	1.30%
193	759	0.44%	2,986	1.75%
201	880	0.51%	3,866	2.26%
209	1,038	0.61%	4,904	2.87%
217	1,076	0.63%	5,980	3.50%
225	1,254	0.73%	7,234	4.23%
229	1,315	0.77%	8,549	5.00%
233	1,483	0.87%	10,032	5.87%
237	1,525	0.89%	11,557	6.76%
241	1,767	1.03%	13,324	7.79%
244	1,826	1.07%	15,150	8.86%
247	1,838	1.07%	16,988	9.93%
251	2,082	1.22%	19,070	11.15%
254	2,188	1.28%	21,258	12.43%
257	2,363	1.38%	23,621	13.81%
260	2,493	1.46%	26,114	15.27%
262	2,660	1.56%	28,774	16.83%
265	2,880	1.68%	31,654	18.51%
268	2,999	1.75%	34,653	20.27%
271	3,276	1.92%	37,929	22.18%
273	3,327	1.95%	41,256	24.13%
276	3,445	2.01%	44,701	26.14%
278	3,735	2.18%	48,436	28.33%
281	3,810	2.23%	52,246	30.55%
283	4,012	2.35%	56,258	32.90%
286	4,108	2.40%	60,366	35.30%
289	4,370	2.56%	64,736	37.86%
291	4,508	2.64%	69,244	40.49%

Table Q15. English Language Arts Grade 5 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
293	4,759	2.78%	74,003	43.28%
296	4,754	2.78%	78,757	46.06%
298	4,822	2.82%	83,579	48.88%
301	5,010	2.93%	88,589	51.81%
303	5,084	2.97%	93,673	54.78%
306	5,256	3.07%	98,929	57.85%
309	5,498	3.22%	104,427	61.07%
311	5,385	3.15%	109,812	64.22%
314	5,290	3.09%	115,102	67.31%
317	5,256	3.07%	120,358	70.39%
320	5,378	3.15%	125,736	73.53%
323	5,307	3.10%	131,043	76.63%
326	5,097	2.98%	136,140	79.61%
329	4,925	2.88%	141,065	82.50%
332	4,805	2.81%	145,870	85.31%
336	4,435	2.59%	150,305	87.90%
339	4,219	2.47%	154,524	90.37%
346	3,735	2.18%	158,259	92.55%
348	3,285	1.92%	161,544	94.47%
352	2,792	1.63%	164,336	96.10%
357	2,202	1.29%	166,538	97.39%
363	1,742	1.02%	168,280	98.41%
370	1,212	0.71%	169,492	99.12%
378	775	0.45%	170,267	99.57%
389	462	0.27%	170,729	99.84%
397	194	0.11%	170,923	99.96%
405	65	0.04%	170,988	99.99%
413	10	0.01%	170,998	100.00%

Table Q16. English Language Arts Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
117	7	0.00%	7	0.00%
125	11	0.01%	18	0.01%
133	21	0.01%	39	0.02%
141	39	0.02%	78	0.05%
149	40	0.02%	118	0.07%
157	69	0.04%	187	0.11%
165	95	0.06%	282	0.16%
173	210	0.12%	492	0.29%
181	305	0.18%	797	0.46%
190	501	0.29%	1,298	0.76%
198	624	0.36%	1,922	1.12%
206	777	0.45%	2,699	1.57%
214	1,016	0.59%	3,715	2.16%
222	1,142	0.66%	4,857	2.83%
227	1,448	0.84%	6,305	3.67%
232	1,554	0.90%	7,859	4.57%
236	1,605	0.93%	9,464	5.51%
240	1,879	1.09%	11,343	6.60%
243	2,007	1.17%	13,350	7.77%
247	2,143	1.25%	15,493	9.01%
250	2,351	1.37%	17,844	10.38%
253	2,422	1.41%	20,266	11.79%
257	2,565	1.49%	22,831	13.28%
259	2,706	1.57%	25,537	14.86%
262	2,846	1.66%	28,383	16.52%
265	2,994	1.74%	31,377	18.26%
268	3,063	1.78%	34,440	20.04%
270	3,239	1.88%	37,679	21.92%
273	3,378	1.97%	41,057	23.89%
276	3,418	1.99%	44,475	25.88%
278	3,591	2.09%	48,066	27.97%
281	3,804	2.21%	51,870	30.18%
283	3,902	2.27%	55,772	32.45%
285	3,890	2.26%	59,662	34.72%
288	4,025	2.34%	63,687	37.06%
290	4,206	2.45%	67,893	39.51%
293	4,316	2.51%	72,209	42.02%
295	4,324	2.52%	76,533	44.53%
297	4,470	2.60%	81,003	47.13%

Table Q16. English Language Arts Grade 6 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
300	4,558	2.65%	85,561	49.79%
302	4,641	2.70%	90,202	52.49%
304	4,635	2.70%	94,837	55.18%
307	4,705	2.74%	99,542	57.92%
309	4,859	2.83%	104,401	60.75%
312	4,707	2.74%	109,108	63.49%
314	4,952	2.88%	114,060	66.37%
317	4,927	2.87%	118,987	69.24%
320	4,916	2.86%	123,903	72.10%
322	4,807	2.80%	128,710	74.89%
325	4,726	2.75%	133,436	77.64%
328	4,605	2.68%	138,041	80.32%
331	4,632	2.70%	142,673	83.02%
334	4,414	2.57%	147,087	85.59%
338	4,066	2.37%	151,153	87.95%
340	3,829	2.23%	154,982	90.18%
344	3,544	2.06%	158,526	92.24%
348	3,073	1.79%	161,599	94.03%
352	2,833	1.65%	164,432	95.68%
357	2,310	1.34%	166,742	97.02%
362	1,852	1.08%	168,594	98.10%
368	1,408	0.82%	170,002	98.92%
376	888	0.52%	170,890	99.44%
385	531	0.31%	171,421	99.75%
397	280	0.16%	171,701	99.91%
405	127	0.07%	171,828	99.98%
413	30	0.02%	171,858	100.00%
421	1	0.00%	171,859	100.00%

Table Q17. English Language Arts Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
98	14	0.01%	14	0.01%
106	24	0.01%	38	0.02%
114	15	0.01%	53	0.03%
122	33	0.02%	86	0.05%
130	35	0.02%	121	0.07%
138	41	0.02%	162	0.10%
146	112	0.07%	274	0.17%
154	173	0.11%	447	0.27%
162	206	0.13%	653	0.40%
170	385	0.23%	1,038	0.63%
178	461	0.28%	1,499	0.91%
186	691	0.42%	2,190	1.33%
194	796	0.48%	2,986	1.81%
202	884	0.54%	3,870	2.35%
210	1,073	0.65%	4,943	3.00%
218	1,233	0.75%	6,176	3.75%
226	1,380	0.84%	7,556	4.59%
230	1,620	0.98%	9,176	5.58%
235	1,769	1.07%	10,945	6.65%
239	1,898	1.15%	12,843	7.80%
242	2,099	1.28%	14,942	9.08%
246	2,294	1.39%	17,236	10.47%
250	2,440	1.48%	19,676	11.96%
253	2,699	1.64%	22,375	13.60%
256	2,846	1.73%	25,221	15.33%
259	2,923	1.78%	28,144	17.10%
262	3,159	1.92%	31,303	19.02%
265	3,339	2.03%	34,642	21.05%
268	3,451	2.10%	38,093	23.15%
271	3,601	2.19%	41,694	25.34%
274	3,747	2.28%	45,441	27.61%
277	3,944	2.40%	49,385	30.01%
279	4,096	2.49%	53,481	32.50%
282	4,250	2.58%	57,731	35.08%
284	4,180	2.54%	61,911	37.62%
287	4,399	2.67%	66,310	40.29%
289	4,423	2.69%	70,733	42.98%
292	4,498	2.73%	75,231	45.72%
294	4,560	2.77%	79,791	48.49%

Table Q17. English Language Arts Grade 7 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
297	4,560	2.77%	84,351	51.26%
299	4,606	2.80%	88,957	54.06%
301	4,589	2.79%	93,546	56.85%
304	4,558	2.77%	98,104	59.61%
306	4,625	2.81%	102,729	62.43%
309	4,545	2.76%	107,274	65.19%
311	4,650	2.83%	111,924	68.01%
313	4,598	2.79%	116,522	70.81%
318	4,481	2.72%	121,003	73.53%
319	4,423	2.69%	125,426	76.22%
321	4,311	2.62%	129,737	78.84%
324	4,270	2.59%	134,007	81.43%
327	4,145	2.52%	138,152	83.95%
330	3,876	2.36%	142,028	86.31%
333	3,765	2.29%	145,793	88.59%
336	3,493	2.12%	149,286	90.72%
340	3,283	1.99%	152,569	92.71%
343	2,916	1.77%	155,485	94.48%
347	2,441	1.48%	157,926	95.97%
352	2,019	1.23%	159,945	97.19%
357	1,716	1.04%	161,661	98.24%
363	1,168	0.71%	162,829	98.95%
370	833	0.51%	163,662	99.45%
379	515	0.31%	164,177	99.77%
390	252	0.15%	164,429	99.92%
398	99	0.06%	164,528	99.98%
406	32	0.02%	164,560	100.00%
414	3	0.00%	164,563	100.00%

Table Q18. English Language Arts Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
100	15	0.01%	15	0.01%
107	20	0.01%	35	0.02%
115	18	0.01%	53	0.03%
123	20	0.01%	73	0.04%
131	53	0.03%	126	0.08%
139	56	0.03%	182	0.11%
147	108	0.07%	290	0.18%
155	159	0.10%	449	0.28%
163	269	0.16%	718	0.44%
171	341	0.21%	1,059	0.65%
179	524	0.32%	1,583	0.97%
187	662	0.41%	2,245	1.38%
194	728	0.45%	2,973	1.82%
202	895	0.55%	3,868	2.37%
210	1,042	0.64%	4,910	3.01%
218	1,169	0.72%	6,079	3.73%
226	1,262	0.77%	7,341	4.50%
231	1,413	0.87%	8,754	5.37%
235	1,505	0.92%	10,259	6.29%
239	1,606	0.98%	11,865	7.27%
243	1,752	1.07%	13,617	8.35%
246	1,917	1.17%	15,534	9.52%
250	1,920	1.18%	17,454	10.70%
253	2,038	1.25%	19,492	11.95%
256	2,124	1.30%	21,616	13.25%
259	2,393	1.47%	24,009	14.71%
262	2,476	1.52%	26,485	16.23%
265	2,543	1.56%	29,028	17.79%
268	2,685	1.65%	31,713	19.44%
270	2,804	1.72%	34,517	21.15%
273	3,016	1.85%	37,533	23.00%
275	3,121	1.91%	40,654	24.92%
278	3,288	2.02%	43,942	26.93%
280	3,352	2.05%	47,294	28.99%
284	3,490	2.14%	50,784	31.12%
285	3,557	2.18%	54,341	33.30%
287	3,890	2.38%	58,231	35.69%
290	3,788	2.32%	62,019	38.01%
292	3,955	2.42%	65,974	40.43%

Table Q18. English Language Arts Grade 8 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
294	4,036	2.47%	70,010	42.91%
296	4,306	2.64%	74,316	45.55%
299	4,366	2.68%	78,682	48.22%
301	4,334	2.66%	83,016	50.88%
303	4,519	2.77%	87,535	53.65%
305	4,524	2.77%	92,059	56.42%
308	4,560	2.79%	96,619	59.21%
310	4,598	2.82%	101,217	62.03%
312	4,812	2.95%	106,029	64.98%
316	4,771	2.92%	110,800	67.91%
317	4,735	2.90%	115,535	70.81%
320	4,670	2.86%	120,205	73.67%
323	4,817	2.95%	125,022	76.62%
325	4,828	2.96%	129,850	79.58%
328	4,754	2.91%	134,604	82.49%
332	4,394	2.69%	138,998	85.19%
335	4,308	2.64%	143,306	87.83%
339	4,040	2.48%	147,346	90.30%
343	3,664	2.25%	151,010	92.55%
348	3,324	2.04%	154,334	94.59%
353	2,746	1.68%	157,080	96.27%
359	2,333	1.43%	159,413	97.70%
367	1,648	1.01%	161,061	98.71%
376	1,134	0.69%	162,195	99.40%
389	605	0.37%	162,800	99.78%
397	273	0.17%	163,073	99.94%
405	87	0.05%	163,160	100.00%
412	7	0.00%	163,167	100.00%

Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
137	8	0.00%	8	0.00%
145	13	0.01%	21	0.01%
153	11	0.01%	32	0.02%
161	23	0.01%	55	0.03%
169	37	0.02%	92	0.05%
177	94	0.05%	186	0.11%
185	184	0.10%	370	0.21%
193	315	0.18%	685	0.39%
201	452	0.26%	1,137	0.64%
209	668	0.38%	1,805	1.02%
217	921	0.52%	2,726	1.54%
225	1,118	0.63%	3,844	2.18%
231	1,358	0.77%	5,202	2.94%
236	1,543	0.87%	6,745	3.82%
240	1,687	0.95%	8,432	4.77%
244	1,850	1.05%	10,282	5.82%
248	2,015	1.14%	12,297	6.96%
252	2,124	1.20%	14,421	8.16%
255	2,165	1.23%	16,586	9.39%
258	2,397	1.36%	18,983	10.74%
261	2,540	1.44%	21,523	12.18%
263	2,587	1.46%	24,110	13.64%
266	2,796	1.58%	26,906	15.23%
269	2,769	1.57%	29,675	16.79%
271	2,857	1.62%	32,532	18.41%
273	3,096	1.75%	35,628	20.16%
276	3,318	1.88%	38,946	22.04%
278	3,307	1.87%	42,253	23.91%
280	3,475	1.97%	45,728	25.88%
282	3,619	2.05%	49,347	27.92%
285	3,586	2.03%	52,933	29.95%
287	3,710	2.10%	56,643	32.05%
289	3,878	2.19%	60,521	34.25%
291	3,932	2.22%	64,453	36.47%
293	4,060	2.30%	68,513	38.77%
295	4,135	2.34%	72,648	41.11%
297	4,062	2.30%	76,710	43.41%
299	4,290	2.43%	81,000	45.84%
302	4,207	2.38%	85,207	48.22%

Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
304	4,345	2.46%	89,552	50.67%
306	4,434	2.51%	93,986	53.18%
308	4,388	2.48%	98,374	55.67%
310	4,536	2.57%	102,910	58.23%
314	4,469	2.53%	107,379	60.76%
315	4,689	2.65%	112,068	63.42%
318	4,585	2.59%	116,653	66.01%
320	4,646	2.63%	121,299	68.64%
323	4,617	2.61%	125,916	71.25%
326	4,704	2.66%	130,620	73.91%
329	4,628	2.62%	135,248	76.53%
332	4,788	2.71%	140,036	79.24%
336	4,718	2.67%	144,754	81.91%
340	4,709	2.66%	149,463	84.58%
344	4,714	2.67%	154,177	87.24%
348	4,562	2.58%	158,739	89.83%
354	4,339	2.46%	163,078	92.28%
360	3,844	2.18%	166,922	94.46%
369	3,580	2.03%	170,502	96.48%
381	2,921	1.65%	173,423	98.13%
389	2,152	1.22%	175,575	99.35%
397	1,145	0.65%	176,720	100.00%

Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
137	5	0.00%	5	0.00%
145	5	0.00%	10	0.01%
153	12	0.01%	22	0.01%
161	23	0.01%	45	0.03%
169	81	0.05%	126	0.07%
177	145	0.08%	271	0.15%
186	353	0.20%	624	0.35%
194	643	0.36%	1,267	0.72%
202	954	0.54%	2,221	1.26%
210	1,451	0.82%	3,672	2.08%
218	1,833	1.04%	5,505	3.11%
226	2,132	1.21%	7,637	4.32%
234	2,462	1.39%	10,099	5.71%
240	2,582	1.46%	12,681	7.17%
246	2,738	1.55%	15,419	8.72%
250	2,724	1.54%	18,143	10.26%
254	2,681	1.52%	20,824	11.78%
258	2,618	1.48%	23,442	13.26%
261	2,639	1.49%	26,081	14.75%
264	2,621	1.48%	28,702	16.23%
267	2,687	1.52%	31,389	17.75%
269	2,706	1.53%	34,095	19.28%
272	2,708	1.53%	36,803	20.82%
274	2,680	1.52%	39,483	22.33%
276	2,662	1.51%	42,145	23.84%
278	2,723	1.54%	44,868	25.38%
280	2,711	1.53%	47,579	26.91%
283	2,813	1.59%	50,392	28.50%
284	2,826	1.60%	53,218	30.10%
286	2,901	1.64%	56,119	31.74%
288	2,776	1.57%	58,895	33.31%
290	2,842	1.61%	61,737	34.92%
291	2,799	1.58%	64,536	36.50%
293	2,846	1.61%	67,382	38.11%
295	2,889	1.63%	70,271	39.74%
296	2,980	1.69%	73,251	41.43%
298	2,921	1.65%	76,172	43.08%
300	3,003	1.70%	79,175	44.78%
301	3,065	1.73%	82,240	46.51%

Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
303	3,013	1.70%	85,253	48.22%
305	3,211	1.82%	88,464	50.03%
306	3,227	1.83%	91,691	51.86%
308	3,237	1.83%	94,928	53.69%
310	3,112	1.76%	98,040	55.45%
311	3,322	1.88%	101,362	57.33%
314	3,247	1.84%	104,609	59.17%
315	3,234	1.83%	107,843	60.99%
317	3,275	1.85%	111,118	62.85%
319	3,412	1.93%	114,530	64.78%
321	3,420	1.93%	117,950	66.71%
323	3,565	2.02%	121,515	68.73%
325	3,552	2.01%	125,067	70.74%
327	3,527	1.99%	128,594	72.73%
329	3,741	2.12%	132,335	74.85%
331	3,726	2.11%	136,061	76.95%
334	3,843	2.17%	139,904	79.13%
337	3,821	2.16%	143,725	81.29%
341	3,921	2.22%	147,646	83.51%
343	3,871	2.19%	151,517	85.70%
346	3,843	2.17%	155,360	87.87%
351	3,908	2.21%	159,268	90.08%
355	3,806	2.15%	163,074	92.23%
361	3,756	2.12%	166,830	94.36%
368	3,420	1.93%	170,250	96.29%
379	3,009	1.70%	173,259	97.99%
397	2,294	1.30%	175,553	99.29%
405	1,254	0.71%	176,807	100.00%

Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
127	3	0.00%	3	0.00%
135	3	0.00%	6	0.00%
143	14	0.01%	20	0.01%
151	14	0.01%	34	0.02%
159	38	0.02%	72	0.04%
167	78	0.05%	150	0.09%
175	161	0.10%	311	0.19%
183	298	0.18%	609	0.36%
191	500	0.30%	1,109	0.66%
199	757	0.45%	1,866	1.11%
207	1,039	0.62%	2,905	1.73%
215	1,297	0.77%	4,202	2.50%
223	1,615	0.96%	5,817	3.47%
232	1,739	1.04%	7,556	4.50%
238	1,833	1.09%	9,389	5.59%
244	1,993	1.19%	11,382	6.78%
249	2,170	1.29%	13,552	8.08%
254	2,263	1.35%	15,815	9.42%
258	2,387	1.42%	18,202	10.85%
261	2,487	1.48%	20,689	12.33%
265	2,570	1.53%	23,259	13.86%
268	2,778	1.66%	26,037	15.51%
272	2,858	1.70%	28,895	17.22%
275	2,970	1.77%	31,865	18.99%
278	3,162	1.88%	35,027	20.87%
280	3,235	1.93%	38,262	22.80%
283	3,375	2.01%	41,637	24.81%
286	3,384	2.02%	45,021	26.83%
288	3,583	2.14%	48,604	28.96%
290	3,628	2.16%	52,232	31.12%
294	3,654	2.18%	55,886	33.30%
295	3,637	2.17%	59,523	35.47%
297	3,700	2.20%	63,223	37.67%
300	3,721	2.22%	66,944	39.89%
302	3,728	2.22%	70,672	42.11%
304	3,715	2.21%	74,387	44.33%
306	3,801	2.26%	78,188	46.59%
308	3,860	2.30%	82,048	48.89%
310	3,725	2.22%	85,773	51.11%

Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
312	3,695	2.20%	89,468	53.31%
314	3,776	2.25%	93,244	55.56%
316	3,671	2.19%	96,915	57.75%
319	3,711	2.21%	100,626	59.96%
320	3,669	2.19%	104,295	62.15%
322	3,774	2.25%	108,069	64.40%
324	3,810	2.27%	111,879	66.67%
326	3,679	2.19%	115,558	68.86%
328	3,695	2.20%	119,253	71.06%
330	3,715	2.21%	122,968	73.27%
333	3,650	2.17%	126,618	75.45%
335	3,678	2.19%	130,296	77.64%
337	3,547	2.11%	133,843	79.75%
340	3,496	2.08%	137,339	81.84%
342	3,466	2.07%	140,805	83.90%
346	3,262	1.94%	144,067	85.85%
348	3,195	1.90%	147,262	87.75%
351	3,062	1.82%	150,324	89.57%
354	2,817	1.68%	153,141	91.25%
358	2,651	1.58%	155,792	92.83%
362	2,418	1.44%	158,210	94.27%
368	2,262	1.35%	160,472	95.62%
373	2,050	1.22%	162,522	96.84%
381	1,675	1.00%	164,197	97.84%
391	1,407	0.84%	165,604	98.68%
399	1,071	0.64%	166,675	99.32%
407	724	0.43%	167,399	99.75%
415	422	0.25%	167,821	100.00%

Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
125	6	0.00%	6	0.00%
133	6	0.00%	12	0.01%
141	12	0.01%	24	0.01%
150	21	0.01%	45	0.03%
158	39	0.02%	84	0.05%
166	88	0.05%	172	0.10%
174	171	0.10%	343	0.21%
182	308	0.18%	651	0.39%
190	526	0.32%	1,177	0.71%
198	892	0.54%	2,069	1.24%
206	1,299	0.78%	3,368	2.02%
214	1,532	0.92%	4,900	2.94%
223	1,974	1.19%	6,874	4.13%
233	2,327	1.40%	9,201	5.53%
241	2,593	1.56%	11,794	7.08%
247	2,617	1.57%	14,411	8.65%
252	2,810	1.69%	17,221	10.34%
256	2,921	1.75%	20,142	12.10%
260	2,969	1.78%	23,111	13.88%
263	3,002	1.80%	26,113	15.68%
267	2,994	1.80%	29,107	17.48%
270	2,990	1.80%	32,097	19.28%
272	3,035	1.82%	35,132	21.10%
275	2,955	1.77%	38,087	22.87%
278	3,016	1.81%	41,103	24.69%
280	3,015	1.81%	44,118	26.50%
282	2,984	1.79%	47,102	28.29%
284	3,013	1.81%	50,115	30.10%
286	3,029	1.82%	53,144	31.92%
289	3,005	1.80%	56,149	33.72%
291	2,984	1.79%	59,133	35.51%
292	2,945	1.77%	62,078	37.28%
294	3,017	1.81%	65,095	39.09%
296	2,985	1.79%	68,080	40.89%
298	2,946	1.77%	71,026	42.66%
300	2,885	1.73%	73,911	44.39%
301	2,855	1.71%	76,766	46.10%
303	2,938	1.76%	79,704	47.87%
305	2,819	1.69%	82,523	49.56%

Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
306	2,888	1.73%	85,411	51.30%
308	2,867	1.72%	88,278	53.02%
310	2,864	1.72%	91,142	54.74%
311	2,750	1.65%	93,892	56.39%
313	2,844	1.71%	96,736	58.10%
315	2,819	1.69%	99,555	59.79%
316	2,785	1.67%	102,340	61.46%
318	2,790	1.68%	105,130	63.14%
319	2,724	1.64%	107,854	64.77%
321	2,765	1.66%	110,619	66.43%
323	2,782	1.67%	113,401	68.11%
324	2,844	1.71%	116,245	69.81%
326	2,685	1.61%	118,930	71.43%
328	2,732	1.64%	121,662	73.07%
330	2,765	1.66%	124,427	74.73%
332	2,823	1.70%	127,250	76.42%
334	2,848	1.71%	130,098	78.13%
336	2,733	1.64%	132,831	79.77%
338	2,713	1.63%	135,544	81.40%
340	2,751	1.65%	138,295	83.06%
342	2,759	1.66%	141,054	84.71%
345	2,797	1.68%	143,851	86.39%
347	2,716	1.63%	146,567	88.02%
350	2,603	1.56%	149,170	89.59%
353	2,641	1.59%	151,811	91.17%
357	2,442	1.47%	154,253	92.64%
360	2,392	1.44%	156,645	94.08%
365	2,259	1.36%	158,904	95.43%
370	2,124	1.28%	161,028	96.71%
376	1,868	1.12%	162,896	97.83%
384	1,610	0.97%	164,506	98.80%
395	1,120	0.67%	165,626	99.47%
403	653	0.39%	166,279	99.86%
411	229	0.14%	166,508	100.00%

Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
124	15	0.01%	15	0.01%
132	11	0.01%	26	0.02%
140	14	0.01%	40	0.03%
148	17	0.01%	57	0.04%
155	35	0.02%	92	0.06%
163	62	0.04%	154	0.10%
171	91	0.06%	245	0.16%
179	169	0.11%	414	0.27%
187	340	0.22%	754	0.48%
194	557	0.36%	1,311	0.84%
202	863	0.55%	2,174	1.39%
210	1,336	0.86%	3,510	2.25%
218	1,769	1.13%	5,279	3.38%
226	2,255	1.44%	7,534	4.83%
233	2,689	1.72%	10,223	6.55%
244	3,079	1.97%	13,302	8.52%
252	3,396	2.18%	16,698	10.70%
258	3,492	2.24%	20,190	12.93%
264	3,708	2.38%	23,898	15.31%
268	3,686	2.36%	27,584	17.67%
272	3,568	2.29%	31,152	19.95%
276	3,679	2.36%	34,831	22.31%
279	3,716	2.38%	38,547	24.69%
282	3,556	2.28%	42,103	26.97%
285	3,433	2.20%	45,536	29.17%
287	3,410	2.18%	48,946	31.35%
290	3,305	2.12%	52,251	33.47%
293	3,285	2.10%	55,536	35.57%
294	3,194	2.05%	58,730	37.62%
296	3,159	2.02%	61,889	39.64%
298	3,055	1.96%	64,944	41.60%
300	2,971	1.90%	67,915	43.50%
302	2,856	1.83%	70,771	45.33%
304	2,973	1.90%	73,744	47.24%
305	2,787	1.79%	76,531	49.02%
307	2,900	1.86%	79,431	50.88%
309	2,761	1.77%	82,192	52.65%
310	2,650	1.70%	84,842	54.35%
312	2,608	1.67%	87,450	56.02%

Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
313	2,516	1.61%	89,966	57.63%
315	2,499	1.60%	92,465	59.23%
316	2,500	1.60%	94,965	60.83%
318	2,447	1.57%	97,412	62.40%
319	2,420	1.55%	99,832	63.95%
321	2,356	1.51%	102,188	65.46%
322	2,438	1.56%	104,626	67.02%
323	2,317	1.48%	106,943	68.50%
325	2,366	1.52%	109,309	70.02%
326	2,323	1.49%	111,632	71.51%
328	2,246	1.44%	113,878	72.95%
329	2,263	1.45%	116,141	74.40%
331	2,306	1.48%	118,447	75.87%
332	2,220	1.42%	120,667	77.29%
334	2,234	1.43%	122,901	78.73%
335	2,222	1.42%	125,123	80.15%
337	2,146	1.37%	127,269	81.52%
339	2,117	1.36%	129,386	82.88%
340	2,148	1.38%	131,534	84.26%
342	2,153	1.38%	133,687	85.63%
344	2,030	1.30%	135,717	86.94%
346	2,130	1.36%	137,847	88.30%
348	2,029	1.30%	139,876	89.60%
350	2,067	1.32%	141,943	90.92%
352	1,917	1.23%	143,860	92.15%
354	1,860	1.19%	145,720	93.34%
357	1,829	1.17%	147,549	94.51%
360	1,710	1.10%	149,259	95.61%
364	1,645	1.05%	150,904	96.66%
368	1,489	0.95%	152,393	97.62%
374	1,336	0.86%	153,729	98.47%
382	1,125	0.72%	154,854	99.19%
390	822	0.53%	155,676	99.72%
398	437	0.28%	156,113	100.00%

Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
124	18	0.01%	18	0.01%
132	11	0.01%	29	0.02%
140	12	0.01%	41	0.03%
148	19	0.02%	60	0.05%
156	35	0.03%	95	0.08%
164	73	0.06%	168	0.13%
172	143	0.11%	311	0.25%
179	230	0.18%	541	0.43%
187	402	0.32%	943	0.76%
195	666	0.53%	1,609	1.29%
203	1,140	0.92%	2,749	2.21%
211	1,457	1.17%	4,206	3.38%
219	1,910	1.53%	6,116	4.91%
227	2,325	1.87%	8,441	6.78%
237	2,714	2.18%	11,155	8.96%
244	2,887	2.32%	14,042	11.28%
250	3,133	2.52%	17,175	13.79%
255	3,159	2.54%	20,334	16.33%
259	3,131	2.51%	23,465	18.85%
263	3,052	2.45%	26,517	21.30%
267	3,011	2.42%	29,528	23.72%
270	2,961	2.38%	32,489	26.09%
273	2,824	2.27%	35,313	28.36%
275	2,799	2.25%	38,112	30.61%
278	2,766	2.22%	40,878	32.83%
280	2,673	2.15%	43,551	34.98%
282	2,575	2.07%	46,126	37.05%
284	2,568	2.06%	48,694	39.11%
287	2,581	2.07%	51,275	41.18%
288	2,553	2.05%	53,828	43.23%
290	2,498	2.01%	56,326	45.24%
292	2,442	1.96%	58,768	47.20%
294	2,373	1.91%	61,141	49.11%
296	2,319	1.86%	63,460	50.97%
297	2,305	1.85%	65,765	52.82%
299	2,241	1.80%	68,006	54.62%
301	2,339	1.88%	70,345	56.50%
302	2,244	1.80%	72,589	58.30%
304	2,226	1.79%	74,815	60.09%

Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution (cont.)

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
306	2,204	1.77%	77,019	61.86%
307	2,098	1.69%	79,117	63.54%
309	2,081	1.67%	81,198	65.22%
310	2,093	1.68%	83,291	66.90%
312	2,053	1.65%	85,344	68.55%
313	2,047	1.64%	87,391	70.19%
315	1,931	1.55%	89,322	71.74%
316	1,961	1.58%	91,283	73.32%
318	1,886	1.51%	93,169	74.83%
319	1,818	1.46%	94,987	76.29%
321	1,829	1.47%	96,816	77.76%
322	1,769	1.42%	98,585	79.18%
324	1,776	1.43%	100,361	80.61%
326	1,671	1.34%	102,032	81.95%
327	1,670	1.34%	103,702	83.29%
329	1,536	1.23%	105,238	84.52%
331	1,534	1.23%	106,772	85.76%
332	1,557	1.25%	108,329	87.01%
334	1,408	1.13%	109,737	88.14%
336	1,318	1.06%	111,055	89.20%
338	1,328	1.07%	112,383	90.26%
340	1,320	1.06%	113,703	91.32%
343	1,264	1.02%	114,967	92.34%
345	1,180	0.95%	116,147	93.29%
349	1,168	0.94%	117,315	94.22%
351	1,037	0.83%	118,352	95.06%
354	980	0.79%	119,332	95.84%
357	986	0.79%	120,318	96.64%
362	869	0.70%	121,187	97.33%
367	890	0.71%	122,077	98.05%
374	842	0.68%	122,919	98.73%
384	670	0.54%	123,589	99.26%
392	561	0.45%	124,150	99.71%
400	356	0.29%	124,506	100.00%